CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 120, No. 4 October 2006

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

Broadband passive synthetic aperture: Experimental results

Edmund J. Sullivan[a)]
EJS Associates, Portsmouth, Rhode Island 02871

Jason D. Holmes[b)] **and William M. Carey**[c)]
Department of Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts, 02215

James F. Lynch[d)]
Department of Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543-1050

Abstract: Bearing estimation using an acoustically short towed array can be enhanced by the incorporation of a realistic signal model. By casting the problem as a joint estimation of bearing and source frequency, the bearing estimation performance, as measured by the variance of the bearing estimate, can exceed that of the conventional array processor. Experimental results based on the radiated noise of a ferry are shown. Bearing estimation results using an array of acoustic length of approximately two wavelengths are compared to results using the same data but with a conventional frequency domain beamformer. A significant improvement in performance is demonstrated.

PACS numbers: 43.60.Gk, 43.30.Wi [James Candy]
Date Received: June 2, 2006 **Date Accepted:** June 28, 2006

1. BACKGROUND

Passive synthetic aperture is a means of enhancing the performance of a moving array by exploiting the information in the Doppler. For example, in the case of bearing estimation, the forward motion of a towed array causes a difference in the spectrum of the received signal as compared to that of the signal at the source.[1] A simple example of this can be seen from the Doppler equation as follows. Consider a narrow-band plane wave signal of radian frequency ω_0 arriving at a receiver that is moving with speed v, where the direction of propagation of the signal is at angle θ with respect to the normal to the direction of motion of the receiver. The frequency of the received signal will be Doppler shifted to frequency ω, and the sign of the product $v \sin \theta$ determines the sign of the Doppler, i.e., positive implies up Doppler. The relation between ω and ω_0 is given by

$$\omega = \omega_0[1 + (v/c)\sin \theta]. \quad (1)$$

Here, c is the speed of sound in the water. Thus, if one has knowledge of the source frequency, the bearing can be found. Passive synthetic aperture bearing estimation exploits this idea by casting the problem as a joint estimation of the source frequency and the bearing angle. In this paper, the problem is cast in the form of an extended Kalman filter.[2] More information on passive synthetic aperture and its history can be found in Ref. 3 and references therein.

[a)]Electronic mail: paddypriest@aol.com
[b)]Electronic mail: jholmes@bu.edu
[c)]Electronic mail: wcarey@bu.edu
[d)]Electronic mail: jlynch@whoi.edu

2. THEORY

Consider first a narrow-band version of the problem. Let a line array of N receiver elements be moving with speed v in the $+x$ direction of a $x-y$ coordinate system. Let the plane-wave signal be arriving at angle θ with respect to the y axis, measured to be positive for clockwise rotation. The signal at the nth receiver can then represented in complex form as

$$s_n(t) = a_n e^{i(\omega_0/c)(x_n+vt)\sin\theta + i\omega_0 t}. \tag{2}$$

As before, ω_0 is the radian frequency of the signal at the source, x_n is the coordinate of the nth receiver element, θ is the bearing angle measured from broadside, a_n is the signal amplitude, and t is time. We choose to work in the phase domain, since this avoids the need to include the signal amplitude as a nuisance parameter. This leaves the two parameters ω_0 and θ to be estimated. The state equation of the Kalman filter is given by

$$\begin{bmatrix} \theta(t|t-1) \\ \omega_0(t|t-1) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \theta(t-1|t-1) \\ \omega_0(t-1|t-1) \end{bmatrix}. \tag{3}$$

Note that there are no dynamics in the state equation. This is based on the assumption that the parameters are changing slowly in time.

Since there is a non-negligible bearing rate in parts of the data record, a version of the algorithm was constructed where a bearing rate was augmented into the state equations. For this case, Eq. (3) becomes

$$\begin{bmatrix} \theta(t|t-1) \\ \alpha(t|t-1) \\ \omega_0(t|t-1) \end{bmatrix} = \begin{bmatrix} 1 & \Delta t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta(t-1|t-1) \\ \alpha(t-1|t-1) \\ \omega_0(t-1|t-1) \end{bmatrix}, \tag{4}$$

where α is an estimate of $\partial\theta/\partial t$, and Δt is the update time increment. In both Eqs. (3) and (4), the notation is designed to indicate the update process. For example, $\theta(t|t-1)$ is the predicted value of θ at time t based on the data up to time $t-1$.

Now consider the broadband case. Since we choose to work in the phase domain, the measurement equation is based on the exponent in Eq. (2). Consider a discrete spectrum of the broadband signal. This will have a set of phases associated with it. Since these phases are related by the frequency index, an average phase can be computed as

$$\psi_n = \frac{1}{\Delta m}\sum_{m=M_{\mathrm{Lo}}}^{M_{\mathrm{Hi}}} \{m(\omega_0/c)(nd+vt)\sin\theta + m\omega_0 t + \phi_{mn}\}/m = \frac{1}{\Delta m}\sum_{m=M_{\mathrm{Lo}}}^{M_{\mathrm{Hi}}} \{m(\omega_0/c)nd\sin\theta + \phi_{mn}\}/m + i\omega t. \tag{5}$$

Here, $\omega = \omega_0[1+(v/c)\sin\theta]$ and is the lowest frequency of a discrete Fourier transform (DFT) of the data at the receiver and M_{Lo} and M_{Hi} are the respective low frequency and high frequency indices of this DFT. $\Delta m = M_{\mathrm{Hi}} - M_{\mathrm{Lo}} + 1$ is the number of frequency components of the DFT, $d = x_n - x_{n-1}$ is the interelement spacing of the line array receiver elements, ϕ_{mn} is an arbitrary phase, and m is the frequency index. Note that this refers all of the phases to that of the lowest frequency line of a virtual DFT at the source.

Since only the phase *differences* are relevant to the problem, there are only $N-1$ receiver-based measurements for the N receivers. Assuming that the arbitrary phase terms ϕ_{mn} average to a negligibly small value, the $N-1$ measurements, based on the model of Eq. (5), are given by

$$y_n = \psi_{n+1} - \psi_n, \quad n = 1,2,\ldots,N-1, \tag{6}$$

where the ωt term has canceled out. There is an auxiliary measurement equation that is based on the observed frequency. This is basically the Doppler relation of Eq. (1) and is given by

$$y_N = \omega = \omega_0[1 + (v/c)\sin\theta]. \tag{7}$$

In this equation, ω_0 can be thought of as the lowest frequency of a virtual DFT of the signal at the source. The resulting measurement system has the (nonlinear) form

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N-1} \\ y_N \end{bmatrix} = \begin{bmatrix} (d/c)\omega_0 \sin\theta \\ (d/c)\omega_0 \sin\theta \\ \vdots \\ (d/c)\omega_0 \sin\theta \\ \omega_0 + (v/c)\omega_0 \sin\theta \end{bmatrix}. \tag{8}$$

3. ALGORITHM

Since the measurements are based on the relative phases of the receiver signals, some preprocessing must be done. The first step is to perform a discrete Fourier transform (DFT) over a selected time window. This time window must be no longer than the time of any significant change in the bearing. The phases of the complex amplitudes of the DFTs are then computed. Next, these phases are averaged over frequency and the pairwise elemental differences are computed. That is, for the phase difference, we must evaluate the following preprocessed measurement for each DFT.

$$y_n = \frac{1}{\Delta m} \sum_{m=M_{\mathrm{Lo}}}^{M_{\mathrm{Hi}}} \{\psi_{n+1}^m - \psi_n^m\}/m, \quad n = 1,2, \ldots, N-1 \tag{9}$$

In the above equation ψ_n^m is the measured phase of the mth frequency component for the nth receiver, and is to be compared with the bracketed term in Eq. (5). The frequency measurement for Eq. (7) is taken to be the lowest frequency of the DFT of the received signal.

4. DESCRIPTION OF THE EXPERIMENT

The experimental data were obtained as a data set of opportunity. During an experiment carried out jointly by Boston University and Woods Hole Oceanographic Institution, using the autonomous Undersea vehicle REMUS, a short (six element) array was towed. During the experiment, a ferry from the mainland of Cape Cod on its way to the island of Nantucket passed through the area. The resulting data provide the basis for this work.[4,5]

The six element array, which had an element spacing of 0.75 m, was towed at a speed of 1.5 m/sec. The ferry appeared by emerging from a shallow region, known as Tuckernuck Shoal, at an angle very close to broadside (0^o) to the towed array, and the closest point of approach occurred at approximately 20^o. The array was moving in a straight line toward the course of the ferry, which was moving at approximately 20 kts, on a straight course from left to right with respect to forward endfire of the array. This configuration is depicted in Fig. 1. The points A and B are the ferry positions for the respective beginning and closet point approach (CPA) of the data used in this work. The distance between these two points is approximately 2 km. Although the radiated sound from the ferry was quite broadband, extending over a band from about 100 to 1000 Hz, there was a particularly strong band of energy occurring between 890 and 920 Hz. This energy band was selected for the data in this paper. At this band of frequencies, the array has an acoustic length of approximately 2.3 λ.

A sequence of 8000 0.1 sec DFTs was generated in order to obtain the phase averages over the full 800 sec of the data. The band was then constrained to the lines between 890 and 920 Hz. These phase averages constituted the basis of the measurements used in Eq. (9). The frequency measurement was taken directly as the lowest frequency of the DFT of the data. An extended Kalman filter (EKF), based on Eqs. (3), (4), and (8) was then used to process the data.[6]

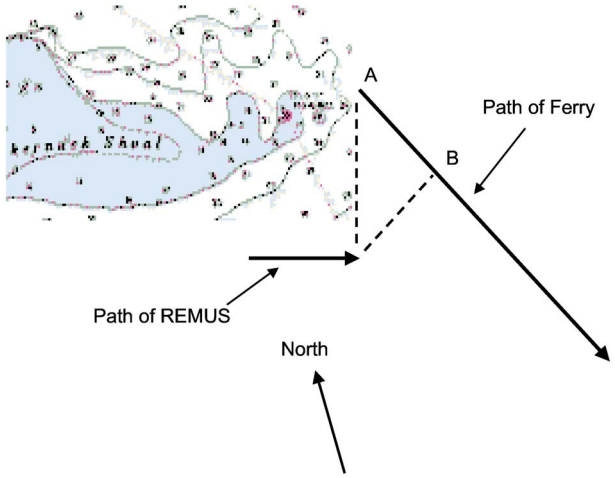


Fig. 1. (Color online) Configuration of the experiment.

5. RESULTS AND DISCUSSION

The results are shown in Figs. 2 and 3. In both figures the vertical axis is time in seconds. The left panel of Fig. 2 is the result of beam forming the data with a conventional frequency-domain beam former, and is computed from

$$p(t_j,\theta_k)=\sum_{n=1}^{N}\sum_{m=M_{\mathrm{Lo}}}^{M_{\mathrm{Hi}}}\{e^{-i2\pi f_m n(d/c)\sin\theta_k}\}F^*_{t_j}(f_m,n). \tag{10}$$

The bracketed term is the steering vector and $F_{t_j}(f_m,n)$ is the frequency domain signal at the nth hydrophone associated with the time t_j. The left panel of Fig. 2 depicts the squared magnitude of $p(t_j,\theta_k)$, normalized for each t_j.

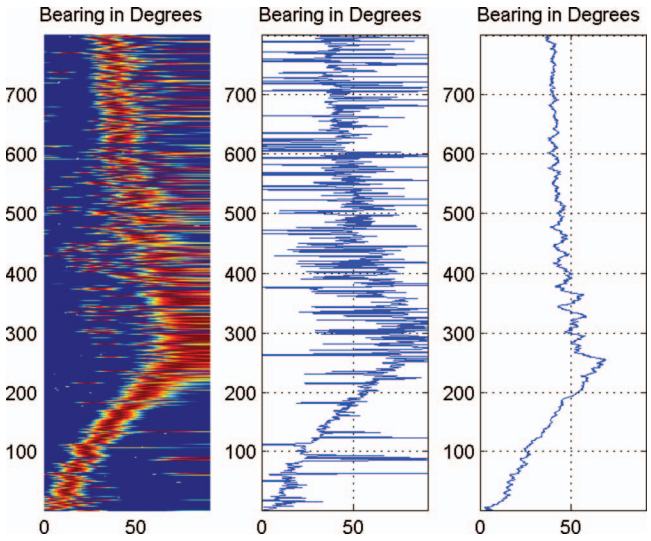


Fig. 2. Results for the random walk case.

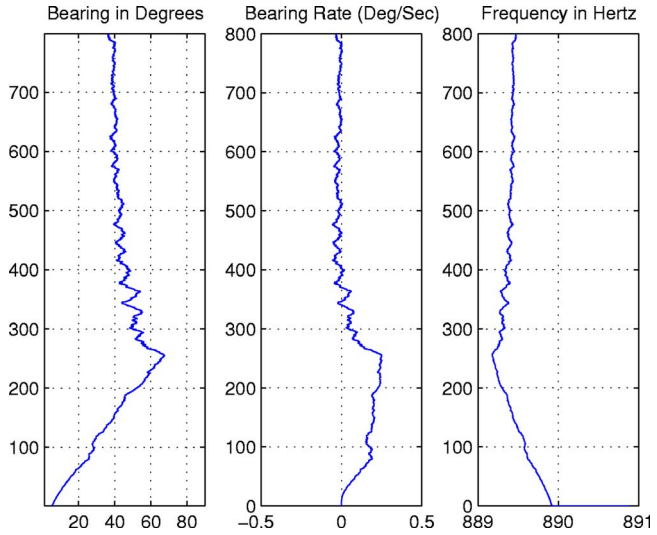


Fig. 3. (Color online) Results for the bearing-rate augmented case.

The center panel shows the maxima of the plot in the left panel, and the right panel shows the synthetic aperture result. As expected, both estimators fail to resolve the bearing in the neighborhood of endfire. After endfire, beginning at about 400 sec, the synthetic aperture clearly shows the cumulative effect expected of such a processor. This occurs since the bearing rate is small, and therefore the random walk state model of Eq. (3) is valid.

Figure 3 depicts the results for the case where the bearing rate is augmented into the processor. The left panel shows the bearing estimate, the center panel shows the estimate of the bearing rate, and the right-hand panel shows the estimate of the source fundamental frequency. Note that this frequency is not constant, since the source itself is undergoing nonzero accelerations. Thus, before endfire, it has an up Doppler and after endfire, a down Doppler. Thus, the apparent fundamental frequency of the virtual DFT at the source must adapt to these speed changes.

The fact that the bearing estimate in Fig. 3 shows some improvement over that of Fig. 2 bears some explanation. The Kalman filter requires that the user specify a trial value for the state error covariance. The value chosen constitutes a lower bound on the eventual state error covariance. This provides a means for the user to control the convergence rate of the process. That is, the larger this covariance is chosen to be, the faster the convergence of the processor; but at the price of a noisier estimate. The estimate in Fig. 3 allowed a smaller value for this covariance to be used, since the convergence requirements for the case of a nonzero bearing rate are eased by the inclusion of the bearing rate directly into the dynamics via Eq. (4). Thus, the limiting state estimation error is smaller in Fig. 3 (Left panel), than that in Fig. 2 (Right panel). This adjustment of the covariance input to the Kalman filter is referred to as "tuning," and is discussed in depth in Ref. 2.

The performance of the synthetic aperture processor presented here is a consequence of proper modeling. There are three elements to the model structure. First, the proper inclusion of the Doppler provides additional bearing angle information, second, the modeling of the state as a Gauss-Markov process exploits the memory implicit in such a recursive model, and third, explicitly including the bearing rate in the model further decreases the bearing error.

ACKNOWLEDGMENTS

The authors would like to thank Amy Kukulya, Ben Allen, Greg Packard, and Art Newhall of WHOI for their valuable assistance. This work was partly funded by WHOI, Boston University,

ONR codes (321,322) and the NDSEG fellowship.

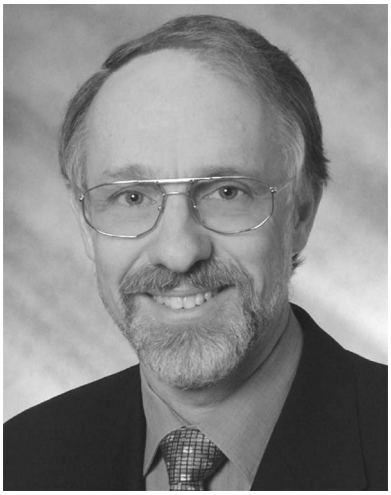
[1]E. J. Sullivan and J. V. Candy, "Space-time array processing: The model-based approach," J. Acoust. Soc. Am. **102**(5), 2809–2820 (1997).
[2]J. V. Candy, *Model-Based Signal Processing* (Wiley, New York, 2006).
[3]E. J. Sullivan, "Passive Acoustic Synthetic Aperture Processing," IEEE OES Newsletter **38**(1), 21–24 (2003).
[4]J. D. Holmes, W. M. Carey, J. F. Lynch, A. E. Newhall, and A. Kukaly, "An autonomous underwater vehicle towed array for ocean acoustic measurements and inversions," *Proceedings of IEEE Oceans 2005—Europe*, June 20–23, Vol. **2** 2005), pp. 1061–1068.
[5]J. D. Holmes, W. M. Carey, and J. F. Lynch, "Results of an autonomous underwater vehicle towed hydrophone array experiment in Nantucket Sound," J. Acoust. Soc. Am. **120** (2), EL15 (2006).
[6]The use of the EKF was necessitated by the fact that the measurement equations are nonlinear. Details of this can be found in Ref. 2.

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.*

New Fellows of the Acoustical Society of America



Anthony W. Gummer
For contributions to the mechanics of mammalian and avian cochlea

Hiroshi Riquimaroux
For contributions to bat sonar

Joos Vos
For contributions to community response to impulsive noise

Ben Zinn
For contributions to combustion acoustics

The 151st meeting of the Acoustical Society of America held in Providence, Rhode Island

The 151st meeting of the Acoustical Society of America was held 6-9 June at the Rhode Island Convention Center in Providence, Rhode Island. Some meeting events were also held at the Westin Providence Hotel, which is adjacent to the convention center. This is the second time that the Society has met in this city, the previous meeting being held in 1974.

The meeting drew a total of 1353 registrants, including 248 nonmembers and 362 students. Attesting to the international ties of our organization, 113 of the registrants were from outside North America. There were 29 registrants from the U.K., 16 from Japan, 9 each from France and Taiwan, 7 from Spain, 5 each from the Netherlands and South Korea, 4 from Belgium, 3 each from Australia, Denmark, Germany and India, 2 each from Austria,

 0001-4966/2006/120(4)/1743/9/$22.50

Ireland, Italy and New Zealand, and 1 each from Brazil, China, Greece, Hungary, Kuwait, Malaysia, Russia, Switzerland and the United Arab Emirates. North American countries, Canada, Mexico, and the United States, accounted for 43, 2, and 1195, respectively.

A total of 1046 papers, organized into 102 sessions, covered the areas of interest of all 13 Technical Committees and the Committee on Education in Acoustics. There were also 23 meetings dealing with standards. The Monday evening tutorial lecture series was continued by Emil Okal, Northwestern University. His tutorial, "The 2004 Sumatra earthquake and tsunami: Multidisciplinary lessons from an oceanic monster," was presented to an audience of about 80. A short course on "Underwater Acoustic Communications" was presented by Pierre-Philippe Beaujean, Florida Atlantic University, to a class of 21 participants. Nikolai Andreevich Dubrovskiy, Director of the N. N. Andreyev Acoustics Institute, Russian Academy of Sciences, presented a Distinguished Lecture titled "Status of Acoustics in Russia."

The Society's 13 Technical Committees held open meetings during the Providence meeting where they made plans for special sessions at upcoming ASA meetings, discussed topics of interest to the attendees and held informal socials after the end of the official business. These meetings are working, collegial meetings and all people attending Society meetings are encouraged to attend and to participate in the discussions. More information about Technical Committees, including minutes of meetings, can be found on the ASA Website ⟨http://asa.aip.org/committees.html⟩ and in the Acoustical News USA section of the September issue of JASA.

The Technical Committee on Signal Processing in Acoustics sponsored its eighth Gallery of Acoustics at the meeting. The winning entry titled "Kemplen's Speaking Machine" was submitted by Tamás Böhm, Massachusetts Institute of Technology, who received the $350 prize.

The Technical Committee on Architectural Acoustics sponsored a Student Design Competition that involved the design of a city municipal building, including a council chambers and courtroom. The entries were judged by a panel of architects and acoustical consultants. The first prize winner will receive a cash award of $1000 and entries selection for "commendation" awards will receive $500 each. Announcement of the award winners had not yet been made at the time this report was written.

There were several other special events presented. A "Composed Spaces Loudspeaker Concert," sponsored by the Technical Committee on Architectural Acoustics, was a session where works of music and sound art were presented by their composers. Sound artist China Blue presented two of her works, including "4 Ball Corner Pocket," which uses spatial recording to capture and manipulate the acoustic element of a billiard game and "Mikey vs. Fabio," a study of the acoustics of a ping-pong game. She also displayed paintings of her visual interpretation of acoustic flow in different environments.

An exhibit was held that featured displays with instruments, materials, and services for the acoustical and vibration community, including sound level meters, sound intensity systems, and signal processing systems. Exhibitors were as follows: American Acoustical Products ⟨www.aap_usa.com⟩, bkm ⟨www.bkmtech.com⟩, Bruel and Kjaer ⟨www.bkhome.com⟩, Eckel Industries ⟨www.eckelusa.com⟩, G.R.A.S. Sound & Vibration ⟨www.gras.us⟩, MBI Products Company, Inc. ⟨www.mbiproducts.com⟩, PCB Piezotronics/Larson Davis ⟨www.pcb.com⟩, RESON Inc. ⟨www.reson.com⟩, Seneca Lake Test Facility ⟨www.npt.nuwc.navy.mil/seneca⟩, Soundown Corp. ⟨www.soundown.com⟩, Tucker-Davis Technologies ⟨www.tdt.com⟩.

Over 300 meeting attendees attended the showing of the film *Touch the Sound* at a theater near the Convention Center. The movie featured Evelyn Glennie, a Grammy winning classical percussionist who is blind.

The Student Council sponsored a Grant Writing Workshop for students and post-docs that focused on the mechanics of grant writing, including white papers and letter proposals, full proposals, essential components, and budget writing. This was followed by the Student Reception, which provided an opportunity for students to meet informally with fellows students and other members of the Acoustical Society. The Student Council presented the Student Council Mentoring Award to Lawrence A. Crum, University of Washington, at the social.

The local committee arranged two technical tours. The first was held on Monday, 5 June, and included a visit to the acoustic test facilities at the Naval Underwater Warfare Center (NUWC) in Newport, RI. The tour included the acoustic tank, pressure tank, antenna test chamber, and anechoic chamber. The second tour was to the manufacturing facilities of Eastern Acoustic Works (EAW) on Thursday, 8 June. EAW has been one of the industry leaders in the design and manufacture of loudspeakers, electronics, and acoustic test and measurement equipment. The tour included visits to all major design and manufacturing areas, a presentation of some selected products and technologies by the Engineering group.

FIG. 1. Lawrence A. Crum, recipient of the 2006 Student Council Mentoring Award.

Social events included the two social hours held on Tuesday and Thursday, an opening reception for the Exhibit, an ice-breaker and a reception for students, the Fellows Luncheon, and the morning coffee breaks. A special program for students to meet one-on-one with members of the ASA over lunch, which is held at each meeting, was organized by the Committee on Education in Acoustics. The luncheon sponsored by the Committee on Women in Acoustics drew over 100 participants. A program was also arranged for the 40 accompanying persons who attended the meeting.

Amar Bose, Bose Corporation, was the speaker at the Fellows luncheon, which was attended by over 150 people. The Fellows luncheon is now open to all meeting attendees.

These social events provided the settings for participants to meet in relaxed settings to encourage social exchange and informal discussions.

The plenary session included a business meeting of the Society, announcements, acknowledgment of the members and other volunteers who organized the meeting and the presentation of awards and certificates to newly elected Fellows.

An announcement was made of the 2006 Student Council Mentoring Award to Lawrence A. Crum, University of Washington (see Fig. 1). Rajka Smiljanic, recipient of the American Speech-Language-Hearing Foundation Research Grant in Speech Science was acknowledged (see Fig. 2).

FIG. 2. Rajka Smiljanic, recipient of the 2006 Research Grant in Speech Science (l) is congratulated by ASA President William Yost (r).

FIG. 3. John K. Horne, recipient of the 2006 Medwin Prize in Acoustical Oceanography (l) is congratulated by ASA President William Yost (r).

The 2006 Medwin Prize in Acoustical Oceanography was presented to John K. Horne, the University of Washington School of Aquatic and Fishery Science (see Fig. 3). Dr. Horne presented the Acoustical Oceanography Prize Lecture titled "Acoustic species identification: When biology collides with physics" earlier in the meeting.

The R. Bruce Lindsay Award was presented to Purnima Ratilal, Northeastern University, "for contributions to the theory of wave propagation and scattering through a waveguide, and to the acoustic remote sensing of marine life" (see Fig. 4). The Helmholtz-Rayleigh Interdisciplinary Silver Medal in Biomedical Ultrasound/Bioresponse to Vibration and Acoustical Oceanography was presented to Mathias Fink, Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI) "for contributions to the understanding of time reversal acoustics." (see Fig. 5). The Gold

FIG. 4. ASA President William Yost (r) congratulates Purnima Ratilal (l) the 2006 R. Bruce Lindsay Award recipient (r).

FIG. 5. ASA President William Yost (r) presents the 2006 Helmholtz-Rayleigh Interdisciplinary Silver Medal to Mathias Fink (l).

Medal was presented to James E. West, Johns Hopkins University "for development of polymer electret transducers, and for leadership in acoustics and the Society." (see Fig. 6).

The election of 14 members to Fellow grade was announced and fellowship certificates and pins were presented. New fellows are as follows: Anders Askenfelt, Alexander U. Case, Torsten Dau, Carol Espy-Wilson, J. Gregory McDaniel, Sheryl Gracewski, Lee A. Miller, Bertel Møhl, Shrikanth Narayanan, Simon D. Richards, Charles M. Salter, Ralph A. Stephen, Dajun Tang, and Gail ter Haar (see Fig. 7).

ASA President William Yost expressed the Society's thanks to the Local Committee for the excellent execution of the meeting, which clearly evidenced meticulous planning. He introduced James H. Miller, University of Rhode Island, Chair of the Providence meeting (see Fig. 8), who ac-

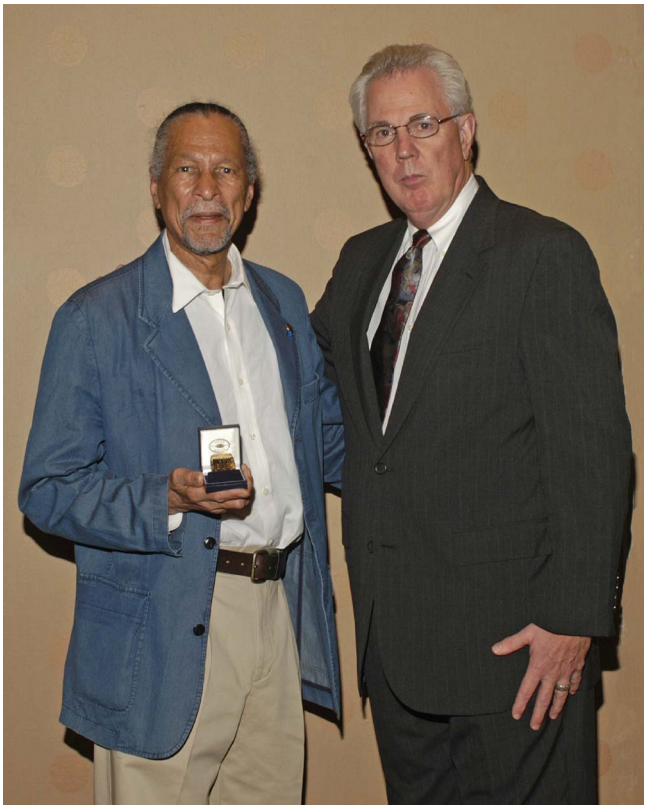


FIG. 6. ASA President William Yost (r) presents the Gold Medal to James E. West (l).

FIG. 7. New Fellows of the ASA.

knowledged the contributions of the members of his committee, including James F. Lynch, Technical Program Chair, Gail Paolino, Meeting Administrator, Peter M. Scheifele, Audio-Visual, John R. Buck, Signs/Publicity, Gopu Potty, Meeting Room Coordinator, Ian Wilson and Jeremy Perkins, Student Issues, Nahal Namadaran, Poster Sessions, James A. Simmons and Andrea Simmons, Cultural Attaché, and David Moretti, Technical Tour. He also expressed thanks to the members of the Technical Program Organizing Committee: James F. Lynch, Technical Program Chair, Andone C. Lavery, Mohsen Badiey, Acoustical Oceanography; Andrea M. Simmons, John R. Buck, Animal Bioacoustics; Damian J. Doria, Alexander U. Case, Architectural Acoustics; Robin O. Cleveland, R. Glynn Holt, Biomedical Ultrasound/Bioresponse to Vibration; Courtney B. Burroughs, Education in Acoustics, Musical Acoustics and Structural Acoustics and Vibration; Thomas R. Howarth, Jeffrey E. Boisvert, David A. Brown, Engineering Acoustics; Nancy S. Timmerman, Noise; Ronald A. Roy, Joseph A. Turner, Charles R. Thomas, Physical Acoustics; Laurie M. Heller, Psychological and Physiological Acoustics; Ning Xiang, David J. Moretti, William M. Carey, Signal Processing in Acoustics; Doug H. Whalen, Harriet S. Magen, Speech Communication; Gopu R. Potty, Kathleen E. Wage, Underwater Acoustics; Edmund R. Gerstein, Thomas G. Muir, Joe Blue Memorial Sessions.

FIG. 8. James H. Miller, Providence meeting Chair.

FIG. 9. Whitlow Au, ASA Vice President-Elect (r) presents gavel to Donna Neff, Vice President (l).

Whitlow Au, Vice President-Elect, presented the Vice President's gavel to Donna Neff, outgoing Vice President and Anthony Atchley, President-Elect, presented the President's Tuning Fork to William Yost, outgoing President, in recognition of their service to the Society during the past year (see Figs. 9 and 10).

The full technical program and award encomiums can be found in the printed meeting program or online at ⟨scitation.aip.org/JASA⟩ (select Vol. 119, No. 5) for readers who wish to obtain further information about the

FIG. 10. President-Elect Anthony Atchley (r) presents President's Tuning Fork to William Yost (l).

Providence meeting. We hope that you will consider attending a future meeting of the Society to participate in the many interesting technical events and to meet with colleagues in both technical and social settings. Information about future meetings can be found in the *Journal* and on the ASA Home Page at ⟨http://asa.aip.org⟩.

WILLIAM A. YOST
President 2005-2006

USA Meetings Calendar

Listed below is a summary of meetings related to acoustics to be held in the U.S. in the near future.

	2006
28 Nov.–2 Dec.	152nd Meeting of the Acoustical Society of America joint with the Acoustical Society of Japan, Honolulu, Hawaii [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
	2007
4–8 June	153rd Meeting of the Acoustical Society of America, Salt Lake City, Utah [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
27 Nov.–2 Dec.	154th Meeting of the Acoustical Society of America, New Orleans, Louisiana (note Tuesday through Saturday) [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
	2008
28 July–1 August	9th International Congress on Noise as a Public Health Problem (Quintennial meeting of ICBEN, the International Commission on Biological Effects of Noise). Foxwoods Resort, Mashantucket, CT [Jerry V. Tobias, ICBEN 9, Post Office Box 1609, Groton CT 06340-1609, Tel.860-572-0680; Web: www.icben.org. E-mail: icben2008@att.net].

Cumulative Indexes to the Journal of the Acoustical Society of America

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

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

REVISION LIST

New Associates

Abercrombie, Clemeth L., 1204 NW 20th Ave., #305, Portland, OR 97209
Adcock, James L., 5005 155th Place, SE, Bellevue, WA 98006
Bard, Seth E., IBM, M/S P226, 2455 South Rd., Bldg. 704, Boardman Road Site, Poughkeepsie, NY 12601
Beard, James K., 122 Himmelein Rd., Medford, NJ 08055
Beard, Paul C., Medical Physics, Univ. College London, Gower St., London WC1E 6 BT UK
Chang, Enson, Applied Signal Technology, Inc., Ocean Systems Div., 21311 Hawthorne Blvd., Ste. 300, Torrance, CA 90503
Cichock, Joseph A., Northrop Grumman, SD&T, 2000 West NASA Blvd., Melbourne, FL 32902
Clark, Cathy Ann, NUWCDIVNPT, Sensors & Sonar Systems Dept., 1176 Howell St., Newport, RI 02841
Cochran, Sandy, Univ. of Paisley, Microscale Sensors, School of Engineering and Science, High St., Paisley PA1 2BE, Scotland
Cross, Emily L., Cavanaugh Tocci Associates, Inc., 327F Boston Post Rd., Sudbury, MA 01776
Curra, Francesco P., Ctr. for Industrial and Medical Ultrasound, Applied Physics Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105
Damljanovic, Vesna, Biomedical Engineering, Boston Univ., 44 Cummington St., Boston, MA 02215
de Groot-Hedlin, Catherine D., Scripps Inst. of Oceanography, Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0225
Derveaux, Gregoire P., INRIA, Domaine de Voluce Av., BP 105, Le Chesnay 78153 France
Fisher, David A., 187 Highland Ave., Floor 1, Somerville, MA 02143
Fleischman, Robert, Soning Praha, A.S., Plzenska 66, Prague 151-24, Czech Republic
Fristrup, Kurt M., Natural Sounds Program, National Park Service, 1201 Oakrdige Dr., Ste. 100, Fort Collins, CO 80525
Golay, Francis, Lab. Regionale des Strasbourg, G5 (Acoustique), 11 rue Mentelin, Strasbourg 67035, France
Greenwood, William A., Pacific Lutheran Univ., 121st and Park Ave., Tacoma, WA 98447
Han, Jin-Ha, Prosonic Co. Ltd., 71-6 Sinpeung-Ri, Geunchen-Eub, Gyongju-Si, Gyongbuk 780-900, South Korea
Harrison, Patrick W., Sound Fighter Systems, LLC, 6135 Linwood Ave., Shreveport, LA 71106
Hofbeck, Mary, 5622 180th St., SW, Lynnwood, WA 98037
Huang, Lianjie, Los Alamos National Lab., MS D443, Los Alamos, NM 87545
Jacobs, Jodi, Lencore Acoustics Corp., 1 Crossways Park Dr., West, Woodbury, NY 11797
Jenkins, Adam C., The Greenbusch Group, 1900 West Nickerson St., Ste. 201, Seattle, WA 98119
Kaufman, Jay R., Kaufman and Associates, 5832 Burnet Ave., Van Nuys, CA 91411
Kautzman, Craig S., The Tennant Company, 701 North Lilac Dr., Minneapolis, MN 55422
Ko, Wing P., Transportation Business Group, CH2M Hill, 555 South Flower St., Ste. 3550, Los Angeles, CA 90071

Kollevoll, Kristan G., BRD Noise and Vibration Control, Inc., 112 Fairview Ave., P.O. Box 127 Wind Gap, PA 18091-0127
Lauback, Scott, Cambridge Sound Management, 33 Moulton St., Cambridge, MA 02138
Leonard, Jonathan S., Lencore Acoustics Corp., 1 Crossways Park Dr., West, Woodbury, NY 11797
Li, Zhaung, 900 Pump Rd., Apt. 76, Richmond, VA 23238
Luo, Xin, Dept. of Auditory Implants and Perception, House Ear Inst., 2100 West Third St., Los Angeles, CA 90057
Mukdadi, Sam M., Mechanical and Aerospace Eng., West Virginia Univ., P.O. Box 6106, Morgantown, WV 26506-6106
Nusbaum, Howard C., Univ. of Chicago, Dept. of Psychology, 5848 South University Ave., Chicago, IL 60637
Otani, Makoto, Toyama Prefectural Univ., Dept. Intelligent Systems Design Eng., Kurokawa 5180 Imizo, Toyama 939-0398, Japan
Prada, Claire, CNRS-RSPCI, Lab. Ondes et acoustique, 10 Rue Vauquelin, Paris 75005, France
Qin, Shengping, Biomedical Engineering, Univ. of California, Davis, One Shields Ave., Davis, CA 95616
Saucier, Scott G., Tibbetts Industries, Inc., Colcord Ave., P.O. Box 1096, Camden, ME 04843
Smith, David R. R., Physiology, Development & Neuroscience, Univ. of Cambridge, Downing St., Cambridge CB2 3EG, UK
Southall, Brandon L., US Dept. of Commerce, Natl. Oceanic and Atmospheric Admin., Natl. Marine Fisheries Service, 1315 East-West Highway, SSMC III #12539, Silver Spring, MD 20910
Tardelli, John D., ARCON Corp., Digital Speech Processing, 150K New Boston St., Wobun, MA 01801
Tyler, Michael D., MARCS Auditory Laboratories, Univ. of Western Sydney, Bldg. 5, Bankstown Campus, Locked Bag 1797, Penrtih South NSW 1797, Australia
Vondrasek, Martin, Soning Praha, A. S., Plzenska 66, Prague 151-24, Czech Republic
Walters, David T., 1344 South 7th St., Lincoln, NE 68502
Xiao, Jianqiang, Hunter College, Psychology, 695 Park Ave., New York, NY 10021
Yasutaka, Ueda, Hazama Corp., TRI, 515-1 Karima, Tsukuba, Ibaraki, 305-0822, Japan
Zarnetske, Michael R., 115 Bay State Dr., Braintree, MA 02184

New Students

Abada, Shani H., 3601 Sainte Famille, Apt. 803, Montreal QC H2X 2L6, Canada
Anguah, Kofi A., Swarthmore College, 500 College Ave., Swarthmore, PA 19081
Arz, Jean-Pierre, 5011 la Fontaine, Montreal QC H1V 1R6, Canada
Azzara, Alyson J., 401 Anderson St., #8A, College Station, TX 77840
Babel, Molly E., Linguistics, Univ. of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650
Barroso, Celia, 3556 Maplewood Ave., Los Angeles, CA 90066
Baumgart, Johannes, Inst. for Aerospace Eng., Technische Univ. Dresden, Dresden 01062, Germany
Beckmann, Daniel F., 5507 Rosewood St., Roeland Park, KS 66205
Benson, David H., 1610 Sherbrooke St., W., Apt. 112, Montreal QC H3H 1E1, Canada
Berkowitz, Michael J., Psychology Dept., Vanderbilt Univ., 301 Wilson Hall, 111 21st Ave., South, Nashville, TN 37203
Blumenrath, Sandra H., Dept. of Biology and Psychology, Univ. of Maryland, College Park, MD 20742
Bodson, Anais, Universit Bochum, Universitatsstrasse ND 6/33, Bochum D-44780, Germany
Bradonikic, Kaca, 311 Allston St., Apt. 12, Brighton, MA 02135
Camacho, Arturo, 700 SW 16th Ave., #108, Gainesville, FL 32601
Chang, Chiung-Yun, Speech and Hearing Science, Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210
Cheng, Rui, Aerospace Engineering, Penn State Univ., 225A Hammond Bldg., University Park, PA 16802
Choi, James J., 379 Windsor Rd., Englewood, NJ 07631
Cholewiak, Danielle M., Cornell Univ., Bioacoustics Research Program, 159 Sapsucker Woods Rd., Ithaca, NY 14850
Clark, Mehgan, 3-212 Lamoine Village, Macomb, IL 61455
Collin, Jamie R., Magdalen College, High St., Oxford OX1 4AU, U.K.
DeAngelis, Chris, Box 20584, Cranston, RI 02920
Derezinsk, Steve J., 1008 Mass Ave., Apt. 702, Cambridge, MA 02138
DiZinno, Nicholas, 1094 Reilly St., Bay Shore, NY 11706
Dossot, Georges A., Ocean Engineering, Univ. of Rhode Island, Sheets Bldg., Narragansett Bay Campus, Narragansett, RI 02882
Duke, Jessica E., Emory Univ, Psychology Dept., 532 Kilgo Circle, Atlanta, GA 30322
Feizollahi, Zhaleh, Linguistics, Georgetown Univ., 37th & '0' St., Washington, DC 20057
Foale, Cameron B., 808 Tress St., Ballarat VIC 3350, Australia
Gaston, Jeremy R., Psychology Dept., Binghamton Univ., Binghamtom, NY 13901
Gauthier, Bruno, 2831 Jenne d'Arc, Apt. 202, Montreal QC H1W3V8, Canada
Giannos, Evangelia, 217 South St., Jamaica Plain, MA 01230
Gomes, Maria L., FACINTER Faculdade Internacional, Avenida Luiz Xavier 103, Curitiba, Parana, 80021-980, Brazil
Gregg, Mellisa K., 333 Candee Ave., Apt. G4, Sayville, NY 11782
Harvey, Ryan B., BAE Systems, 1250 24th St., NW, Ste. 850, Washington, D.C. 20037
Hearst, Jason, 30 St. Stephen St., Boston, MA 02115
Hon, Elisabeth, MIT, 60 Wadsworth St., Apt. 19E, Cambridge, MA 02142
Hsieh, I-Hui, 5513 Verano Pl., Irvine, CA 92617
Joshi, Aditya, 17C Bayberry Rd., New Bedford, MA 02740
Kenny, Ryan J., Boston College, Psychology Dept., 301 McGuinn Hall, 140 Commonwealth Ave., Chestnut Hill, MA 02467
Kracht, Jonathan M., 295 East Church St., Sellersville, PA 18960
Kumpulanian, Danielle, 7 Country Lane, Palmer, MA 01069
Lai, Puxiang, Aerospace and Mechanical Eng., Boston Univ., 110 Cummington St., Boston, MA 02215
Law, II, Franzo F., 1337 Jefferson Ave., Brooklyn, NY 11221
Lee, Ji-Yeoun, R405, 103-6, Munji-dong, Yuseong-gu, Daejeon 305-732, Republic of Korea
Leger, Mark L., 8 Charles Plaza, Apt. 209, Baltimore, MD 21201
Lezamiz, Lucas, Applied Physics Lab., Univ. of Washington, 1013 NE 40th St., Box 355640, Seattle, WA 98105-6698
Li, Naihsin, National Taiwan Univ., Graduate Inst. of Linguistics, No. 1, Sec. 4 Roosevelt Rd., Taipei 10617, Taiwan
Lin, Wan, 337795 Gatech Station, Atlanta, GA 30332
Lin, Yao-ju, National Taiwan Normal Univ., English Dept., Linguistic Div., 162 HePing East Rd., Sec. 1 Taipei 106, Taiwan
Liu, Fei, 282 Corry Village, Apt. 7, Gainesville, FL 32603
Liu, Sheng, 865 Tucker Rd., Apt. 4, North Dartmouth, MA 02747
Lopez, Karece, 90-79 180th St., Jamaica, NY 11432
Love, Katharine L., 4015 Hudson Dr., Hoffman Estates, IL 60195
Luo, Haibaio, Aerospace and Mechanical Eng., Boston Univ., 110 Cummington St., Boston, MA 02134
MacLeod, Mark H., Mechanical Engineering, Johns Hopkins Univ., 3400 North Charles St., Baltimore, MD 21218
Maleke, Caroline, Biomedical Engineering, Columbia Univ., New York, NY 10027
Mancini, Jolene A., Hearing, Speech, and Language Science, Gallaudet Univ., 800 Florida Ave., NE, Washington, DC 20002
Maruska, Karen P., Hawaii Inst. of Marine Biology, 46-007 Lilipuna Rd., Kaneohe, HI 96744
McCarty, Candice Q., 3610 Driftwood Dr., Lafayette, IN 47905
Meyer, Matthias, Gartenstrasse 13, Wiefelstede 26215, Germany
Mihalcik, Ladislav, National Healthy Organization, NRC Noise and Vibration, RUZ Bratislava, Ruzinovska 8, Bratislava, Slovakia 829 09, Slovak Republik
Miller, Denise M., 1013 Old Boalsburg Rd., Apt. 6, State College, PA 16801
Moallem, Theodore M., RLE Sensory Communication Group, Massachusetts Inst. of Tech., 77 Massachusetts Ave., Rm. 36-737, Cambridge, MA 02139
Navaladi, Akshay, Biomedical Engineering, Boston Univ., 44 Cummington St., Boston, MA 02215
Newman, Kelly A., School of Fisheries and Ocean Sciences, Univ. of Alaska Fairbanks, 245 Oneill Bldg., P.O. Box 757220, Fairbanks, AK 99775-7220

Noirot, Isabelle C., Charles Plaza 8, North Tower, Apt. 1303, Baltimore, MD 21201
Noisternig, Markus, Inst. of Electronic Music & Acoustics, Univ. of Music and Dramatic Arts, Inffeldgasse 10/3, Graz 8010, Austria
Olmstead, Anne J., 69 Varga Rd., #112, Ashford, CT 06278
Oorellana, Douglas W., 104 West University Parkway, Apt. B1, Baltimore, MD 21210
Pogal-Sussman, Tracy, Boston Univ., BME Dept., 44 Cummington St., Boston, MA 02215
Pruthi, Tarun, 9314 Cherry Hill Rd., 917, College Park, MD 20740
Quijano, Jorge E., 2139 West Burnside St., Apt. 203, Portland, OR 97210
Roberts, Paul L. D., Electrical and Computer Engineering, Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0238
Rosenbaum, Joyce E., 2 Horizon Rd., PH1, Fort Lee, NJ 07024
Rossi-Katz, Jessica, 112 Kolar Court, Erie, CO 80516
Rousounelos, Andreas, Flayt 147, Matthias Court, Silk St., Salford M3 6JF, U.K.
Rowland, Sarah A., Univ. of Connecticut, Psychology Dept., 406 Babbidge Rd., Unit 1020, Storrs, CT 06269-1020
Sadaka, Janine, 9 Jeffrey Lane, Great Neck, NY 11020
Sathyendra, Harsha M., Electrical Engineering, Univ. of Florida, 216 Larsen Hall, Gainesville, FL 32611
Savitala, Hari V., 405 3rd St., #2, Troy, NY 12180
Schneider, Jennifer N., SUNY Buffalo, Psychology Dept., Park Hall, Room 206, Buffalo, NY 14260-4110
Son, Minjung, Haskins Labs., 300 George St., #900, New Haven, CT 06511
Steen, Thomas L., Boston Univ., Mechanical Eng., 110 Cummington St., Boston, MA 02215
Threw, Barry J., 2215 Ward St., Berkeley, CA 94705
Tibrea, Roxana D., 20 Fairway Dr., East Hampton, NY 11937
Trout, Justin N., 196-1 Allston St., Allston, MA 02134
Uysal, Ismail, Electrical and Computer Eng., Univ. of Florida, 216 Larsen Hall, Ctr. Dr., Gainesville, FL 32611
Van Ark, Emily M., MIT-WHOI, Geology and Geophysics, 77 Massachusetts Ave., 54-517A, Cambridge, MA 02139
Vasques, Cesar M. A., DEMEGI/FEUP, Rua Dr. Roberto Frias, s/n Edificio M, Sala 206, Porto 4200-465, Portugal
Viswanathan, Navin, 90a Birch St., Willimantic, CT 06226
Wang, Weixiong, Dept. of Engineering Mechanics, Fluid Acoustics Lab., Tsinghua Univ., Beijing 100084, China
Wiessner, Nicole L., 2010 Misty Hollow Court, Forney, TX 75126
Wilkason, Colby, 120 Fairways Dr., Warner Robins, GA 31088
Yarbrough, Ray A., Applied Research Labs., Univ. of Texas, 10000 Burnet Rd., Austin, TX 78758

New Electronic Associates

Aura, Matias, Wartsila Finland Oy, Calculations and Simulations, Jarvikatu 2-4, Vaasa FI-65101, Finland
Boehme, Hollis, 320 King Arthur Court, Austin, TX 78746
Ess, Robert H., 1107 Pine Hollow Dr., Friendswood, TX 77546
Garrido Lopez, David, Musicos 26, Tres Cantos, Madrid 28760, Spain
Gazagnaire, Julia, Naval Surface Warfare Ctr., Dept. of Defence, 110 Vernon Ave., Panama City Beach, FL 32407
Greenlee, Doug, 995 South Garfield St., Denver, CO 80209
Griffin, Sararh J., MRC Toxicology Unit, Univ. of Leicester, P.O. Box 138, Hodgekin Bldg., Lancaster Rd., Leicester LE1 9HN, U.K.
Hain, Tim C., Northwestern Univ., 645N. Michigan, Chicago, IL 60611
Heintze, Olaf, Inst. of Composite Structures and Adaptive Systems, German Aerospace Ctr., Lilienthalpaltz 7, Braunschweig 38108, Germany
Hetzer, Claus, National Ctr. for Physical Acoustics, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677
Huang, Caroline B., 39 Howells Rd., Belmont, MA 02478
Lehman, Mark E., Communication Disorders, Central Michigan Univ., 2173 Health Professional Bldg., Mt. Pleasant, MI 48859
Maetani, Toshiki, Otolaryngology/Head and Neck Surgery, Ehime Univ. School of Medicine, Shitsukawa, Toon, Ehime 791-0295, Japan
Mahajan, Sanjay K., 28834 W. King William Dr., Farmington Hills, MI 48331
Mansky, James M., Earth Tech, One World Financial Ctr., 200 Liberty St., 25th Floor, New York, NY 10281
Marquez, Ronald, IDS, 353 James Record Rd., Huntsville, AL 35801
Nielsen, Johan L., Tandberg, Philip Redersens vei 22, Lysaker 1355, Norway
Ono, Nobutaka, The Univ. of Tokyo, Dept. Information Physics and Computing, Grad. Sch. of Info. Sci. and Tech., 7-3-1 Hongo, Tokyo 113-8656, Japan
Pastuszek-Lipinska, Barbar E., ul. Iglasta 8, Grotniki-Ustronie, Poland 95-073
Pfaffli, Gregor Grischa, Gerberngasse 21A, Bern 3011, Switzerland
Quinn, Sandra, Univ. of Stirling, Psychology Dept., Stirling FK9 4LA, Scotland, UK
Ramprashad, Sean A., DoCoMo USA Labs, Media Lab., 181 Metro Dr., San Jose, CA 95110
Serkhane, Jihene E., Cold Spring Harbor Lab., Freeman Bldg., 1 Bungtown Rd., Cold Spring Harbor, NY 11724
Shah, Gaurav S., Eliza Corp., 100 Cummings Ctr., Ste. 348G, Beverly, MA 01915
Slaughter, Julie C., Etrema Products, Inc., 2500 North Loop Dr., Ames, IA 50010
Tregenza, Nicholas J., 5 Beach Ter., Long Rock, Cornwall, TR20 8JE, UK
Vitchev, Nikolay V., 4363 Cherry Ave., San Jose, CA 95118
Whittum, David H., Varian Medical Systems, Microwave and Physics R&E, 911 Hansen Way, C077, Palo Alto, CA 94304
Wilson, M. Lee, Shimoda Environmental, Inc., 7602 Stoneywood Dr., Austin, TX 78731
Yokotani, Yoshikazu, Westliche Stadtmauerstr. 38, Zimmer #9, Erlangen, Bayern 91054, Germany

New Corresponding Electronic Associates

Bistafa, Sylvio R., Dept. of Mechanical Eng., Univ. of Sao Paulo, Av. Prof. Melo Moraes 2231, Sao Paulo SP 05508-900, Brazil
Daunys, Gintautas, Siauliai Univ., Electronics, Vilniaus 141, Siauliai 76353, Lithuania
Meesawat, Kittiphong, Khon Kaen Univ., Dept. of Electrical Eng., 123 Mitraphab Rd., Mueng, Khon Kaen, 40002, Thailand
Pal, Amita, Bayesian & Interdisciplinary Research Unit., Indian Statistical Inst., 203 Barrackport Trunk Rd., Kolkata, West Bengal, 700108, India
Srinivasan, K., Apt. 009, Admiralty Manor 83, 6th Main, Indira Nagar II Stage, Bangalore, Karnataka, 560008, India
Szigetvari, Andrea, Berzsenyi u. 14/A, Dunakeszi, Pest, 2120, Hungary

Members Elected Fellows

P. Blanc-Benon, D. A. Conant, A. W. Gummer, C. W. Holland, J. E. Kreiman, K. D. LePage, J. A. McAteer, D. R. Palmer, M. G. Prasad, P. A. Rona, M. Vorländer, J. Vos

Associates Elected Fellows

S. Beristain, A. C. Gade, H. Riquimaroux, M. V. Trevorrow, B. T. Zinn

Associates Elected Members

J. Ahlstrom, S. B. Blaeser, C. E. Hughes, G. E. Jacobs, X. Jiang, R. E. Kumon, B. R. Munson, S. M. Perron, T. S. Talayman

Students to Associates

W. C. K. Alberts, II, R. B. Astrom, J. G. Bernstein, L. K. Bornstein, J. L. Hiatt, A. Roginska

Students to Electronic Associates

M. S. Allen, L. B. Berry, P. A. Cariani, S. A. Cheyne, S. F. Disner, B. L. Engdahl, M. J. Epstein, W. T. Fitch, C-F. Huang, A. J. Morgan, I. Paek, A. J. Subkey, S. J. van Wijingaarden

Resigned

J. R. Ison, M. R. Noble—Members
R. Eklund—Associate
C. Lin—Student
Y. Igarashi—Electronic Associate

Deceased

S. Buus, L. Lisker, T. Litovitz, W. A. Watkins—Fellows
J. G. Harris, B. H. Pasewark—Members

Dropped

Fellows

Bender, Erich K., Cantrell, John H., Fuller, Christopher R., Guernsey, Richard M., Koyasu, Masaru, Pickles, James O., Salvi, Richard J., Schusterman, Ronald J.

Members

Abu-Hassan, Rachid K., Barber, D. Christopher, Bell, Alan E., Bennett, Mary Beth S., Berliner, Marilyn J., Bohme, Johann F., Bolia, Robert S., Brown, R. William, Bullock, Gary L., Castagnede, Bernard R., Chamuel, Jacques R., Cheng, Raymond, Clifton, Mark A., Cook, Reginald O., Cunitz, Robert, Du, Gonghuan, El-Raheb, Michael, Gawtry, Randall R., Geddes, Earl R., Ghen, David C., Hansel, Celeste Z., Hansen, Robert A., Haverstick, Gavin A., Henry, Belinda A., Hosseini, Seyed H.R., Hsu, David Kuei-Yu, Huang, Frank, Jackson, Oliver H., Koch, John E., Lagier, Michel R., Lang, Mark A., Larson, Vernon D., Lashkari, Khosrow, Lee, Jaiyong, Lee, Sook-Hyang, Lin, Qiguang, Madanshetty, Sameer I., Miller, Roger L., Murray, Rachel V., Murray, Todd W., Noffsinger, Paul D., Osborn, Jay K., Pritchard, Robert S., Rajagopalan, Subramonium, Read, Robert D., Seo, Jong-Soo,
Shallcross, William D., Shankar, P. Mohana, Sharma, Anu, Staal, Philip R., Stanke, Fred E., Swanson, Cal T., Tabrikian, Joseph, Takagi, Naoki, Thys, Willy J.V., Trainor, Laurel J., Tran Van Nhieu, Michel, Van Hoof, H.A.J.M., Walton, Dennis C., Wang, Chong, Wang, Shuo, Wilson, Gary R., Wu, Lei, Zoccola, Paul J.

Associates

Abdul Hamid, Siti Zaleha, Agnew, Nancy E., Andersen, Bjorn K., Arata, Jonathan J., Arens, Egidius, Avsic, Tom, Bambery, John E., Bassrei, Amin, Beurskens, Kees L., Bielby, Gregory J., Blonigen, Florian John, Bogdanowicz, Kenneth J., Brienzo, Richard K., Brown, Gordon, Buckingham, Christian E., Cariani, Peter A., Carter, Benjamin A., Catanzariti, Scott P., Chandler, Ray L., Charles, Terri, Chaudhary, R.S., Chen, Alex C., Chen, C. Julian, Cheng, Jason Yu-Lin, Cherry, Peter V., Cheyne, Harold A., Chiang, Lilian C.W., Chu, Wai C., Chulani, Haresh M., Coleman, K.A., Colombo, Joseph, Compton, Cynthia L., Conroy, James P., Cosgrove, Michael A., Daggett, John E., Day, Joseph L., de Heer, Raymond C., Dennis, John A., Dille, Marilyn F., Dolan, David F., Dooley, Benjamin D., Drake, Carolyn, Duffy, William C., Duncan, Mike E., Eidens, Richard S., Elchik, Michael E., Estill, Jo, Fabian, Martin J., Fagelson, Marc A., Fisher, Craig A., Fortune, Todd W., Fraas, Michael R., Fujioka, Chieko N., Galvez, Carlos E., Ghandi, Ahmad, Girolami, Gerard M., Gooding, Frank G., Goold, John C., Gregg, Helen L., Grove, Deborah M., Haines, Garrett, Halberstam, Benjamin, Hales, L. Paul, Hall, Carl M., Hall, David S., Hamrin, John E., Hastings, Aaron L., Heerwagen, Dean R., Heise, Ulrich, Hicks-Postar, Lori A., Honda, Kiyoshi, Horne, Peter R., Hou, Zezhang, Huettel, Lisa C., Hughes, Michael S., Jeng, Jing-Yi, Jeong, Hong, Johnstone, Tom, Jouppi, Norman P., Kane, M.R., Kang, Gye Nam, Kanta, Ravi, Katsnelson, Boris G., Kemp, Robert H., Kimmel, Eitan, King, Charles B., King, Wayne M., Klein, Alan J., Knezek, Kathleen M., Kochanski, Greg P., Konofagou, Elisa E., Korber, Dennis J., Kresge, James K., Kumaresan, Jeevith I., Lage, Maria Oti'Lia P., Leary, Del M., Legleiter, Kurt A., Li, Haiying, Liehr, April M., Littman, Thomas A., Livengood, Kim J., Lloyd, Daniel J., Low, Robert, Luce, Paul A., Lyberg, Bertil, Madden, John Patrick, Man, Xiuting Kaleen C., Maniscalco, Albert M., Mansy, Hussein A., Mariano, Marcos, Mazur, Martin A., McArthur, Rod L., Medlin, Kathleen, Mees, Paul, Meghezzi, Fatiha, Meyers, James A., Modarresi, Golnaz, Mody, Maria B., Mullennix, John., Nagy, Attila B., Neuhoff, John G., Noble, Kevin M., Norris, Thomas, Norrix, Linda W., O'Donovan, Jonathan J., O'Ulainn, Seamus P., Odgaard, Eric C., Pastore, Robert A., Pathak, Ardhendu G., Peterson, Richard J., Philhong, Lee, Phillips, Daniel B., Pinyard, Scott, Pouliquen, Eric, Prentice, Scott C., Prince, David J., Quintos, Dario J., Radentz, Michael G., Ramani, Deepak V., Restrepo, Juan M., Ritchie, James A., Rodriguez, Joyce M., Rosati, Robert A., Ross, Annie, Rossi-Tison, Lucile S., Rutherford, Peter, Saint-Vincent, Stephen, Savard, Jacques, Sawyer, Eric, Scofield, Glenn A., Scott, Brian L., Seifert, Eric Heinz, Sharp, Stephen J., Shaw, Simon A., Shiau, Wendy, Shilling, Russell D., Singh, Dhiraj, Singh, Moninderjit, Skorski, Edwin S., Skurka, John C., Smith, Gordon P., Sowizal, John, Speaker, William H., Stewart, Marc C., Tabei, Makoto, Tallapragada, Bhanuprakash, Tidd, Richard A., Tincoff, Ruth J., Torres, Rendell R., Tse, John Kwock-Ping, Tutton, Robert L., Uhlman, James S., Van Moorhem, William K., Varady, Mark J., Vaydik, Frank W., Verbsky, Babette L., Verdonk, Edward D., Vertegaal, Han, Visintini, Lucio, White, Richard M., Whitfield, James, Willems, Stefan M.J., Winter, Joseph F., Witton, Caroline, Wolf, Len A., Yoo, Young-Joo, Young, Thomas M., Zhang, Weiguo

Electronic Associates

Ackley, Robert S., Alonso, Adriana Molero, Bachenko, Joan C., Barrable, Ross, Bello, Ryan M., Bozeman, John K., Brown, Richard O., Candiver, Amy, Daves, Brian W., Davidson, Robert J., Dreini, Marco, Dunbar, John A., Graham, Joelle G., Gwinn, William R., Gyurko, Harrison D., Harwell, Ross M., Ingram, Ian L.H., Krysac, Lorraine Cindy, Manning, Mary D., McBride, Dennis K., McCabe, Terence, Michaels, William L., Nedwell, Jeremy, Noel, Claire, Peters, Dennis J., Peterson, D. Kent, Pietrzyk, John, Randolph, Patricia, Rocaboy, Francoise M.J., Schmerr, Lester W., Smith, Steven J., Strong, Henry, Suzuki, Mikio, Taylor, Lawrence S., Torio, Guy A., Vallabha, Gautam K., Yapura, Carlos L., Yin, Chuan

Corresponding Electronic Associate

Barros, Alessio T., Hun, Ryang Woo, Kim, Jeehyun, Oh, Suntaek, Pimentel, Jamie A.

Students

Adams, Jason B., Aguilar, Rolando N., Al-Khairy, Mohamed A., Alam, Iftekhar, Altinsoy, Mehmet E., Alvarenga, Andre V., Anderson, Kate T., Angert, Phillip E., Atwood, Steven G., Austin, Kimberly, Badertscher, Jeff W., Baird, David, Barger, Mari M., Bartsch, Mark A., Bemis, Jeremy P., Bennison, Corrie J., Bharitkar, Sunil G., Bishop, Joseph R., Block, Gareth I., Boley, Jonathan, Bolin, Michael J., Bonati, Joshua T., Boucher, Matthew A., Bracken, Jeffrey A., Brill, Stefan M., Brown, Meredith A., Bubnash, Brian, Budd, Sarah R., Budhlakoti, Suvrat, Bush, Adam R., Caclin, Anne, Callaway, Jason E., Campbell, Fiona M., Cao, Ji, Cardillo, Dominic J., Carmichael, Lesley M., Chan, Kwan, Chandrasekhar, B., Charmes, Emmanuel, Chen, Larissa, Chen, Quan, Chen, Xi, Cherry, Sean D., Chester, Scott M., Chinchilla-Rodriguez, Sherol S., Choy, Bill, Clarke, Clyde C., Coffin, Allison, Coleman, Mark N., Connaghan, Kathryn P., Cook, Ian M., Corcoran, David E., Coren, Amy E., Cox, Ethan A., Crooks, Gary E Cutliffe, Steven P., de Cardoso, Guilherme C., Deshmukh, Om D., Diankha, Ousmane, Dohen, Marion S., Doolittle, Cory S., Doolittle, Daniel F., Duffy, Chris M., Eaton, Cortney A., Eichfeld, Jahn D., Einarsson, Thorvaldur, Epstein, Michael J., Espinoza, German A., Everhard, Ian L., Faulkner, Katie F., Federici, Jovi P., Fernandes, Luis L., Ferrini, Vicki L., Finney, Nathaniel R., Flanagan, Sheila A., Flurie, Alexander D., Freeman, Smith A.M., Fujimoto, Antonio K., Gabor, Nathaniel M., Gant, Valdez L., Giannini, Roberto, Giessler, Kurt E., Gilcrist, Laura E., Gokhale, Nachiket H., Goor, Mina, Grace, Angela, Gratke, Jesse T., Grow, David I., Guerra, Melania, Gunel, Banu, Hacihabiboglu, Huseyin, Hall, Jessica M.., Hallenbeck, Stephen A., Hanna, Emily J., Harb, Hadi, Harris, Michael C., Hawbaker, Joel, Heiman, Catherine P., Herrmann, Kristen G., Holdhusen, Mark H., Hong, Robert S., Horak, Jennifer R., Horton, Gregory P., Hoyle, Matthew W., Huber, Terese, Huss, Martina M.E.F., Hyeong-Seok, Kim, Jing, Yuan, Johnson, Christopher T., Johnson, Sean P., Jong-Yeon, Kim, Kalyanapasupathy, Vijayaraghavan, Kang, Soyoung, Karbeyaz, Basak Ulker, Kelly, John K., Kelly, Thomas P., Khioe, Fung Wah, Khodayari-Rostamabad, Ahmad, Kim, Sahyang, Kolodziej, Martin P., Kondash, James W., Kottke, Nelson J., Kovacyk, Kristie J., Kroesen, Maarten, Kurz, Anja, Larson, Julie C., Lee, Judy, Lee, Linda-Eling, Lee, Matthew E., Lefkowitz, Kimberly A., Leider, Martha D.K., Lewis, Johnathan C., Li, Anxiang, Lie, Ki, Light, Jason R., Lilly, Christopher F., Lin, Hejie, Lin, Pei-Yu, Liu, Siyun, Liu, YaoJan, Livingston, Brian E., Londono, Jairo, Lubich, Jeffrey N., Lufi, Samuel J., Lutz, Steven D., Maguluri, Gopi N., Manis, Matthew K., Mantha, Sravan P., Martin, Eric J., Martinson, Eric B., Massot, Olivier,

Matthis, Kyle J., McKowen, Mark D., Megerson, Susan C., Mondini, Michele, Morgan, Andrew J., Morgans, Richard C., Morris, Gary L., Mou, Xiaomin, Mueller, Jacob L., Mulder, Theresa D., Namasivayam, Aravind K., Narayan, Chandan R., Nichols, Matt J., Niehus, Rebecca, Noriega, Lauren E., Novak, Barbra J., Ntanos, Christos, O'Donnell, Matthew L., Otero, Sebastian, Ovalle, Arlene, Paliatsos, Demetrios N., Perimeter, Mike R., Phan, Ha T., Ping, Tan Tien, Plichta, Bartlomiej, Poling, Jeremy R., Posdamer, Stephanie H., Pradhan, Rajdeep S., Prasad, Kunal, Pugh, Adam G., Qin, Michael K., Rabbitt, Alicia A., Rampersad, Joanne V., Richey, Scott M., Riggs, Daylen B., Ritchie, Peter J., Rodda, Judith L., Rodriguez, Benjamin, Root, Benjamin J., Rosales, Paola, Rosengaard, Peninah S., Rowland, Daniel C., Ryan, Robert H., Saenz, Michael K., Sagastegui, Maria M., Saikachi, Yoko, Salameh, Amjad, Sarangapani, Sairajan, Sarpun, Ismail, Sato, Momoko, Saweikis, Meghan G., Schmidt, Benjamin A., Seal, Christopher R., Semenova, Tatiana V., Seo, Jongbum, Shah, Jashmin K., Skovenborg, Esben, Slaney, Jacob H., Slowey, Alison K., Smith, Heather M., Somerville, Andrew L., Spahr, Erik J., Stanley, Cheung M.L., Steeve, Roger W., Stumpf, Kelley R., Swick, Andrew H., Takamaru, Keiichi, Taylor, Michael J., Theis, Kevin R., Thirwani, Kapil, Thompson, Lauren M., Trenton, John, Turk, Oytun, Turner, Andrew D., Ulrich, T. J., Utami, Sentagi S., Utley, Daniel L., van Etten, Chris P., van Wassenhove, Virginie Velikic, Gordana, Vidal, Miguel A., Vijayakumar, V., Wagner, James, Walsh, Bridget M., Walter, Michael J., Wang, Hai, Wang, Yong, Warren, Laura J., Watkins, Emily S., Watkinson, Rebecca K., Watts, Matthew K., Wei, Wei, Whitehouse, Andrew M., Wieberg, Kimberly M., Wilbur, Jed C., Williamson, Rene G., Wise, Jason A., Woolley, Jonathan A., Wysocki, Tamra M., Xu, Ching X., Yang, Dan, Yerikalapudi, Aparna V., Yu, Linxiao, Zanetti, Paolo, Zhi, Ni, Zollinger, Sue Anne

Fellows	903
Members	2243
Associates	2563
Students	981
Electronic Associates	628
	7318

ACOUSTICAL NEWS—INTERNATIONAL

Walter G. Mayer
Physics Department, Georgetown University, Washington, DC 20057

International Meetings Calendar

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

October 2006

3–6 **IEEE International Ultrasonics Symposium,** Vancouver, BC, Canada (Web: www.ieee-ultrasonics2006.org).

4–6 **33rd International Acoustical Conference "Acoustics High Tatras 06"–EAA symposium**,Štrbské Pleso, Slovakia (E-mail: 33iac@skas.sk; Web: www.skas.sk/acoustics/2006).

11–13 **Annual Conference of the Canadian Acoustical Association**, Halifax, Nova Scotia, Canada (Web: www.caa-aca.ca/halifax-2006.html).

16–17 **Institute of Acoustics Autumn Conference**, Oxford, UK (Web: www.ioa.org.uk/viewupcoming.asp).

18–20 **37th Spanish Congress on Acoustics–EAA Symposium of Hydroacoustics–Iberian Meeting on Acoustics**,Gandia-Valencia, Spain (Web: www.ia.csic.es/sea/index.html).

25–28 **Fifth Iberoamerican Congress on Acoustics**, Santiago Chile (Web: www.fia2006.cl.).

November 2006

2–3 **Swiss Acoustical Society Fall Meeting**, Luzern, Switzerland (Web: www.sga-ssa.ch).

3–4 **Reproduced Sound 22**, Oxford, UK (Web: ioa.org.uk/viewupcoming.asp).

9–10 *__Baltic -Nordic Acoustics Meeting__, Göteborg, Sweden (Web: www.ingemansson.com.).

20–22 **1st Joint Australian and New Zealand Acoustical Societies Conference**, Christchurch, New Zealand (Web: www.acoustics.org.nz).

March 2007

13–17 *__Spring Meeting of the Acoustical Society of Japan__, Tokyo, Japan (Acoustical Society of Japan, Nakaura5th-Bldg., 2-18-20 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan; fax: +81 3 5256 1022; web:www.asj.gr.jp/index-en.html).

15–17 *__AES 30th International Conference on Intelligent Audio Environment__, Saariselkä, Finland (Web:www.aes.fi/aes30/).

19–22 *__German Acoustical Society Meeting (DAGA2007)__, Stuttgart, Germany (Web: www.daga2007.de).

April 2007

10–12 **4th International Conference on Bio-Acoustics**, Loughboro, UK (Web: www.ioa.org.uk/viewupcoming).

16–18 **29th International Symposium on Acoustical Imaging**, Shonan Village Center, Kanagawa Pref., Japan (Web:publicweb.shonan-it.ac.jp/ai29/AI29.html).

June 2007

1–3 *__Second International Symposium on Advanced Technology of Vibration and Sound__, Lanzhou, China (Web: www.jsme.or.jp/dmc/Meeting/VSTech2007.pdf).

3–7 **11th International Conference on Hand-Arm Vibration**, Bologna, Italy (Web:associazioneitalianadiacustica.it/HAV2007/index.htm).

18–21 **Oceans07 Conference**, Aberdeen, Scotland, UK (Web: www.oceans07ieeeaberdeen.org).

25–29 *__2nd International Conference on Underwater Acousitc Measurements: Technologies and Results__, Heraklion, Crete, Greece (Web: www.uam2007.gr).

July 2007

2–6 **8th International Conference on Theoretical and Computational Acoustics**, Heraklion, Crete, Greece (Web:www.iacm.forth.gr/~ictca07).

4–7 *__International Clarinet Association Clarinetfest__, Vancouver, British Columbia, Canada (E-mail:john.cipolla@wku.edu; phone: 1 270 745 7093).

9–12 *__14th International Congress on Sound and Vibration (ICSV14)__, Cairns, Australia (Web: www.icsv14.com).

16–21 *__12th International Conference on Phonon Scattering in Condensed Matter__, Paris, France (Web: www.isen.fr/phonons2007).

August 2007

6–10 **16th International Congress of Phonetic Sciences (ICPh2007)**, Saarbrücken, Germany (Web: www.icphs2007.de).

26–29 **Inter-noise 2007**, İstanbul, Turkey (Web: www.internoise2007.org.tr).

 0001-4966/2006/120(4)/1752/2/$22.50

27–31 **Interspeech 2007**, Antwerp, Belgium (Web: www.interspeech2007.org).

September 2007

2–7 **19th International Congress on Acoustic (ICA2007)**, Madrid, Spain (SEA, Serrano 144, 28006 Madrid,Spain (Web: www.ica2007madrid.org).

9–12 **ICA Satellite Symposium on Musical Acoustics (ISMA2007)**, Barcelona, Spain (SEA, Serano 144,28006 Madrid, Spain (Web: www.ica2007madrid.org).

9–12 **ICA Satellite Symposium on Room Acoustics (ISRA2007)**, Sevilla, Spain (Web: www.ica2007madrid.org).

17–19 **3rd International Symposium on Fan Noise**, Lyon, France (Web: www.fannoise.org).

19–21 *__Autumn Meeting of the Acoustical Society of Japan__, Kofu, Japan (Acoustical Society of Japan, Nakaura5th-Bldg., 2-18-20 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan; fax: +81 3 5256 1022; web:www.asj.gr.jp/index-en.html).

24–28 *__XIX Session of the Russian Acoustical Society__, Nizhny Novgorod, Russia (Web: www.akin.ru).

June 2008

30–4 **Acoustics 08 Paris: 155th ASA Meeting +5th** Forum Acousticum **(EAA)+9th Congrès Françaisd'Acoustique (SFA)**, Paris, France (Web: www.acoustics08-paris.org).

July 2008

7–10 **18th International Symposium on Nonlinear Acoustics (ISNA18)**, Stockholm, Sweden (Temporary e-mail: ⟨Bengt Enflo⟩ benflo@mech.kth.se).

28–1 **9th International Congress on Noise as a Public Health Problem**, Mashantucket, Pequot Tribal Nation(ICBEN 9, P.O. Box 1609, Groton, CT 06340-1609, USA; web: www.icben.org).

September 2008

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

November 2008

1–5 **IEEE International Ultrasonics Symposium**, Beijing, China (Web: www.ieee-uffc.org/ulmain.asp?page=symposia).

August 2010

23–27 **20th International Congress on Acoustics (ICA2010)**, Sydney, Australia (Web: www.ica2010sydney.org).

Belgian Acoustical Association—40 Years Old

The Belgian Acoustical Association, ABAV, celebrates its 40th birthday on 15 September 2006 with a special meeting in Liège (or Luik, in Flemish.) ABA is the abbreviation of the French name of the Association while BAV stands for the Flemish name. ABAV is both.

The program for the day also reflects the cooperation of different nationalities. There will be four keynote lectures as well as poster papers. The first keynote lecture is on room acoustics and it will be given by the President of the EAA (Vorländer, Germany), followed by a lecture on auralisation et acoustique virtuelle by the Vice President of EAA (Polack, France). The afternoon session features lectures on geluidwering in gebouwen (noise in buildings) (Gerretsen, The Netherlands) and on the interaction of sound with the ground (Attenborough, UK.) The lectures are given in the original languages; only the support information (slides, videos, etc.) are all in English.

The birthday celebration comes to an end with a 90-minute cocktail hour and a gala dinner. Happy Birthday, ABAV!

ADVANCED-DEGREE DISSERTATIONS IN ACOUSTICS

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

Analysis of parameter effects on sound energy decay in coupled volume systems [43.55.Ka, 43.55.Gx, 43.55.Hy] —David Timothy Bradley, *Architectural Engineering Program, University of Nebraska–Lincoln, Lincoln, NE 68182, May 2006 (Ph.D.).* This study characterized the effects of modifying architectural parameters on coupled volume (CV) sound fields. The parameters included the ratio of main volume (MV) and CV size, the ratio of absorption, and aperture size. The level of double slope effect (DSE) was determined for various configurations of the CV systems. Additionally, the human subjective response to the sound fields in these systems was investigated. A simple CV system was studied using computational modeling, where DSE decreased with increasing ratio of CV absorption over MV absorption. Also, DSE increased with increasing ratio of CV size over MV size, referred to as volume ratio. DSE values were maximized at relatively small aperture sizes. Subjective response testing showed an increase in perceived reverberance with increasing volume ratio and aperture size, while perceived clarity did not change by a statistically significant amount across any of the architectural parameters. The effects of varying absorption ratio and aperture size on DSE in a complex CV system, typifying an average CV concert hall, was also investigated. DSE trends generally corresponded with those from the simple system and previous research. Subjective preference testing of this virtual hall showed that listeners preferred low and medium levels of DSE.

Advisor: Lily M. Wang

Investigation of ocean acoustics using autonomous instrumentation to quantify the water-sediment boundary properties [43.30.Ma, 43.30.Pc, 43.30.Re, 43.60.Fg] —Jason David Holmes, *College of Engineering, Boston University, Boston, MA 02215, June 2006 (Ph.D.).* Sound propagation in shallow water is characterized by interaction with the ocean's surface, volume, and bottom. In many coastal margin regions, including the Eastern U.S. continental shelf and the coastal seas of China, the bottom is composed of a depositional sandy-silty top layer. Previous measurements of narrow and broadband sound transmission at frequencies from 100 Hz to 1 kHz in these regions are consistent with waveguide calculations based on depth and frequency dependent sound speed, attenuation and density profiles. Theoretical predictions for the frequency dependence of attenuation vary from quadratic for the porous media model of Biot to linear for various competing models. Results from experiments performed under known conditions with sandy bottoms, however, have agreed with attenuation proportional to frequency raised to the 1.84 power, which is slightly less than the theoretical value of frequency squared [J. Zhou and X. Zhang, J. Acoust. Soc. Am. 117, 2494]. This dissertation presents a reexamination of the fundamental considerations in the Biot derivation and leads to a simplification of the theory that can be coupled with site-specific, depth dependent attenuation and sound speed profiles to explain the observed frequency dependence. Long-range sound transmission measurements in a known waveguide can be used to estimate the site-specific sediment attenuation properties, but the costs and time associated with such at-sea experiments using traditional measurement techniques can be prohibitive. Here a new measurement tool consisting of an autonomous underwater vehicle and a small, low noise, towed hydrophone array was developed and used to obtain accurate long-range sound transmission measurements efficiently and cost effectively. To demonstrate this capability and to determine the modal and intrinsic attenuation characteristics, experiments were conducted in a carefully surveyed area in Nantucket Sound. A best-fit comparison between measured results and calculated results, while varying attenuation parameters, revealed the estimated power law exponent to be 1.87 between 220.5 and 1228 Hz. These results demonstrate the utility of this new cost effective and accurate measurement system. The sound transmission results, when compared with calculations based on the modified Biot theory, are shown to explain the observed frequency dependence.

Advisor: William M. Carey

Investigations of indoor noise criteria systems based on human perception and task performance [43.50.Ba] —Erica Eileen Bowden, *Architectural Engineering Program, University of Nebraska–Lincoln, Lincoln, NE 68182, February 2006 (Ph.D.).* Several noise criteria methods commonly used in architectural acoustics have been quantitatively related to noise perception and task performance under a variety of ventilation systems-induced background noise conditions. Noise criteria, balanced noise criteria, room criteria, room criteria mark II, and A-weighted equivalent sound pressure level were examined. The first phase of the project included noise conditions controlled to be non-time-varying and nontonal, with neutral, rumbly, roaring, or hissy characteristics. An intermediate study examined exposure time length and types of performance tasks used. The final phase included noise conditions containing various levels of discrete tones from 120 to 595 Hz. Under each noise, subjects completed performance tasks and perception questionnaires. Results indicate task performance was significantly affected by perception of noise, but this relationship was not fully demonstrated by the criteria systems analyzed. The five criteria were generally well suited in describing subjective loudness perception, but some discrepancies in criteria spectral quality ratings and subjective perception existed. Finally, perception of annoyance changed based on the frequency and prominence of tones in noise, but these changes were not reflected in the criteria level or spectral quality ratings. Modifications to the existing criteria are recommended.

Advisor: Lily M. Wang

 0001-4966/2006/120(4)/1754/1/$22.50

BOOK REVIEWS

P. L. Marston
Physics Department, Washington State University, Pullman, Washington 99164

These reviews of books and other forms of information express the opinions of the individual reviewers and are not necessarily endorsed by the Editorial Board of this Journal.

Editorial Policy: *If there is a negative review, the author of the book will be given a chance to respond to the review in this section of the Journal and the reviewer will be allowed to respond to the author's comments. [See "Book Reviews Editor's note," J. Acoust. Soc. Am.* **81***, 1651 (May 1987).]*

The Physics of Birdsong (Biological and Medical Physics, Biomedical Engineering)

Gabriel B. Mindlin and Rodrigo Laje

Springer, Berlin, Heidelberg, 2005, 157 pp. $79.95 (hardcover), ISBN: 3-40-25399-8

Until a few years ago models concerning the way birds vocalize were based on indirect evidence from analyses of the structure of bird songs, on inferences from the morphology of the avian vocal organ, the syrinx, and from insight into physical principles. The major reason is that the syrinx is relatively inaccessible, although its anatomy, innervation, muscular control, and driving airflow are well described in some species. However, the bioacoustics and biomechanics of bird phonation are still poorly understood. The last decade has seen a renewed interest in the subject and the introduction of a number of new experimental techniques. So, now the model makers are on much firmer ground when proposing new mechanisms for how birds sing. The present book should be viewed as a progress report on this rapidly evolving subject written by two physicists who nicely sum up recent experimental and theoretical research covering not only the bioacoustics and biomechanics, but also the neurobiology of birdsong.

The three introductory chapters set the scene. In the first, readers are reminded of the basics of sound as a physical phenomenon, the different measures of sound, and of general principles such as superposition. In addition, the traditional bioacoustic representation of sound signals as spectrograms is introduced, though the authors use the old term "sonogram." The second chapter introduces sound sources and filters and associated concepts such as resonators, modes, and traveling and standing waves, which leads to a general source-filter model. Except for the unnecessarily complicated derivation of wave equations, this chapter is tutorial in nature and easy to follow with nice examples and good illustrations. The third chapter is the only one without equations and gives a thorough description of the anatomy and function of the avian vocal organ, the syrinx. In songbirds this is a bipartite organ situated at the junction between the trachea (the windpipe) and the bronchi. On each side two structures known as labia, reminiscent of the mammalian vocal folds, form pneumatic valves. These modulate the airflow, which is created by expiratory muscles rhythmically compressing the inhaled air in the air sacs, through the vocal tract leading to generation of birdsong.

In my opinion Chapters 4–6 form the core of the book describing how song most likely is produced in the syrinx and associated structures under control of the delicate syringeal muscles. Through an understandable description of linear and nonlinear oscillators via the dynamics of the van der Pol oscillator we arrive at the oscillations in the syrinx. A simplified model describes the self-sustained oscillations of the labia and how this dynamics depend on the reconstitution constant, k, of the labia and on the subsyringeal pressure, p. After a pedagogical introduction to the Hopf bifurcation, a simple model of sound production is presented. The authors show with the model how simple changes of the timing between k and p produce different paths in the parameter space that lead to complex patterns in frequency and time, i.e., in the sound spectrogram. Simple gestures of the syringeal muscles furthermore can adduct or abduct the labia and thus control the airflow through the syrinx. Most convincingly, the authors show that spectrograms very similar to those of cardinals are produced by feeding actual experimentally derived values of air sac pressure and syringeal muscle activity, expressed as rectified EMG's, into the model.

After the thorough treatment of the workings of the syrinx I had expected to enjoy reading about the last link in the chain of events from motor command to emission of song, that involving the acoustics and mechanics of the trachea, mouth cavity, and beak. Instead, the authors choose to spend most of Chapter 7 on describing how to synthesize birdsong both by computer and by an analog model. After an educational introduction of the acoustic impedance, however, they do briefly refer to new studies on sound radiation and bandpass filtering by avian air sacs. This leads to an informative Chapter 8 describing the present understanding, from a computational perspective, of the complex and highly nonlinear interactions of neural circuits in controlling birdsong. The text includes basic neurobiology starting with a description of the membrane potential and the Hodgkin-Huxley equations, but quickly develops into a presentation of a number of different computational models of the behavior of interconnected neurons, and discusses the advantages and shortcomings of using such models. The authors conclude by pointing out how hopelessly far we are from understanding the workings of the brain and discuss what level of computational units, neurons or circuits, should be used in attempting to model brain dynamics. With the extensive literature on the function and physiology of avian sound communication the song control system might constitute a good system for continuing this endeavor.

The theme of the final chapter concerns how complex patterns can emerge from simple inputs to systems exhibiting nonlinear dynamics. In an amusing journey from basic forced oscillations through the duetting of some South American birds to the dynamics of Poincaré oscillators, the authors elegantly lead the reader through some complicated mathematics. The reader gets a feeling of understanding and ends up agreeing that the complex interactions between neural circuits and vocal system provides biology and physics with exciting challenges for years to come.

So, for whom is this book intended? Despite its merits it is not well suited as a student text mostly because it suffers from the lack of an index and a glossary would have helped. On the back cover it is stated that the "book provides fascinating reading for physicists, biologists and general readers alike." This statement is probably too optimistic, since the book attempts to solve a classic dilemma: the need for expressing complex problems by means of equations versus the ordinary biologist's lack of command of mathematics. To physicists it is second nature to express a thought in a few equations, but most biologists are unable to appreciate the beauty of equations unless these are accompanied by educational graphics. Long passages of the book are a true pleasure to read as they are written broadly and philosophically, almost lyrically, but this does not help the biologist when he fails to see how one equation leads to the next.

Physicists unfamiliar with the generation of birdsong will definitely enjoy reading the book but should skip the first basic chapters. They will learn a lot about this fascinating topic and will easily pick up the contents of the book in a few hours. Biologists unfamiliar with the acoustics, mechanics, and neurobiology of birdsong may be frightened by the thorough mathematical descriptions, but if they choose to abstract from the equations and concentrate on text and graphics they will be richly rewarded. General readers may make better use of their time by perusing one of the recent review papers on birdsong generation written by biologists.

Biologists studying bioacoustics will gain much inspiration from reading the book. And for the growing number of biologists specializing in the large and diverse field of birdsong research this book is an indispensable contribution, which deserves many readers.

OLE NÆSBYE LARSEN
Center for Sound Communication, Institute of Biology
Campusvej 55, DK-5230 Odense M, Denmark

 0001-4966/2006/120(4)/1755/1/$22.50

Communications Acoustics

Jens Blauert

Springer, Berlin, Heidelberg, New York. 379 pp. Price $129.00 (hardcover). ISBN: 354022162X.

According to the definition provided by the author of this book, the term *Communication Acoustics* refers to "those areas of acoustics which relate to the modern communication and information sciences and technologies." The term itself is not a new one. Notably, both J. L. Flanagan and Jont Allen have used the term "communication acoustics" at recent meetings of the ASA to refer to the groundbreaking work in speech perception performed at Bell Laboratories in the first half of the 20th century. However, this collection of contributed chapters seeks to bring the concept of communications acoustics into the 21st century by giving a broad overview of the current state of the art in all the research areas concerning the production, analysis, transmission, synthesis, and perception of sounds by human talkers and listeners. While no single book could possibly cover all aspects of such a wide range of topics, this book provides a great deal of useful information about many diverse areas of acoustics, and I believe that it will make a major contribution to the literature in at least three ways.

The first major contribution of the book is that it brings into one collection an extremely diverse set of topics that might not ordinarily be considered by any single practitioner in the field of acoustics. The fourteen chapters of the book span a broad range of areas, from a chapter by Georg Klump on the evolutionary physiology of the auditory transmission and reception organs, to a chapter by Ute Jekocsh on the science of "semiotics" and its application in the field of sound design, to a chapter by Inga Holube and Volkmar Hammacher on hearing aid technology. In-depth coverage is provided on various aspects of psychoacoustics (including models of binaural hearing and aspects of sound quality and audio-visual perception), virtual audio environments (including binaural synthesis as well as applications), speech perception (including production mechanisms and speech quality), and digital audio technology and coding. All of the chapters are well-written and clearly illustrated, and a few of the chapters are truly outstanding in terms of their depth and breadth of coverage of the topic area. For example, I was particularly impressed with the chapter on "Binaural Technique" by Henrik Møller and Dorte Hammershøi, which provides a clear and concise summary of all of the known factors involved in the collection of high-quality head-related transfer functions. While the chapters are easy enough to understand, they are not really intended for complete novices. In general, the chapters of the book seem to be geared toward those readers who have some knowledge in one particular area of acoustics but would like to obtain a broad overview of the current state of the art in other areas.

The second major contribution of the book is that it provides a detailed overview of the current state of acoustic research in Europe. Fifteen of the seventeen contributors to the book are currently affiliated with European institutions, and the remaining two have strong historical ties to European universities. Consequently, many of the chapters are filled with outstanding references to prior work that has been published in non-English journals or in European doctoral dissertations that may be unfamiliar to many American readers. I personally found a number of useful references that were relevant to my own research areas, and I believe that many other readers of the book might have the same experience. Thus, I believe the book will prove to be an invaluable resource for acoustic researchers to broaden the scope of their knowledge about prior work in their own specific areas of expertise.

The final and perhaps the greatest contribution of this book is that it serves as a poignant reminder of the breadth and the depth of the contributions that Jens Blauert has made in the study of acoustics over the course of his illustrious career at the Ruhr-University Institute of Communication Acoustics in Bochum, Germany. Blauert's 1983 book on Spatial Hearing and its 1997 revision are still widely considered to be the gold standard for understanding the processes involved in auditory localization, and he and his collaborators have made immeasurable contributions to the development of binaural models. At least four of the contributors to this book are Blauert's students, and in his opening chapter he references 43 of the more than 50 Ph.D. dissertations he has supervised. Though we certainly have not heard the last from him, this book serves as a fitting tribute to the tremendous impact Jens Blauert has had on the worldwide study of acoustics in the many distinguished years he has served as Professor at the Ruhr University of Bochum.

DOUGLAS BRUNGART
Air Force Research Laboratory,
WPAFB, OH 45433

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
JOHN M. EARGLE, *JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068*
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*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *University of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

7,031,155

43.25.Nm ELECTRONIC THERMAL MANAGEMENT

Ioan Sauciuc and Gregory M. Chrysler, assignors to Intel Corporation
18 April 2006 (Class 361/695); filed 6 January 2003

This patent discloses the use of a piezoelectric bending element as a single-blade flapping fan. The idea is not new, but was developed for just the application described here (attachment to integrated circuit packages for cooling) by Purdue researchers in 2001. What may be new is the packaging of the flapping element inside a chamber.—JAH

7,035,166

43.30.Pc 3-D FORWARD LOOKING SONAR WITH FIXED FRAME OF REFERENCE FOR NAVIGATION

Matthew Jason Zimmerman and James Henry Miller, assignors to FarSounder, Incorporated
25 April 2006 (Class 367/88); filed 17 October 2003

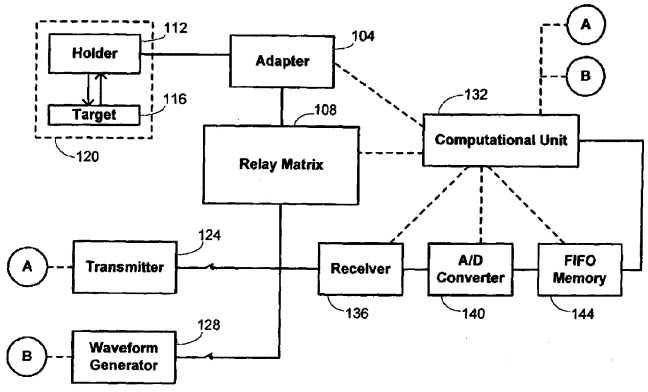
A forward-looking (or side-looking, or bottom-looking) sonar system incorporates a transmit transducer for projecting a signal into the water ahead of the vessel and a phased array of receivers to provide return signals to a computer, which then determines azimuthal and elevation angles and times of arrival of the return signals. The system is equipped with roll and tilt sensors, GPS receiver, and compass. The acoustic information plus the roll and tilt information is processed to create a 3D image of the space ahead of the vessel relative to a fixed frame of reference, i.e., the Earth. Advanced signal processing techniques allow the computer to extract targets from the raw data and other features of the system enable the suppression of multi-path targets.—WT

7,028,529

43.35.Yb APPARATUS AND METHODS FOR TESTING ACOUSTIC PROBES AND SYSTEMS

James M. Gessert *et al.*, assignors to Sonora Medical Systems, Incorporated
18 April 2006 (Class 73/1.82); filed 28 April 2003

This is the latest in a series of patents describing a fairly straightforward method for sequentially testing the individual probes and cables used in an ultrasonic imaging system. The invention itself is not as interesting as



the patent filing strategy. Of the 13 references cited, all but one were supplied by the examiner, including the earlier patents assigned to Gessert *et al.* Apparently, inventors and patent attorneys have very short memories.—GLA

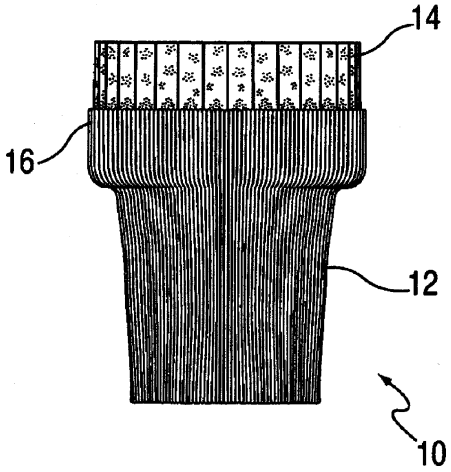
 0001-4966/2006/120(4)/1757/18/$22.50

7,029,446

43.35.Yb STANDOFF HOLDER AND STANDOFF PAD FOR ULTRASOUND PROBE

Martin Edmund Wendelken, Elmwood Park and Charles Pope, Guliford, both of New Jersey
18 April 2006 (Class 600/459); filed 30 October 2003

Disclosure is made for an accommodative standoff ultrasound probe holder. The standoff holder can be mounted and utilized with transducers of different sizes and shapes. The standoff holder includes a removable gel



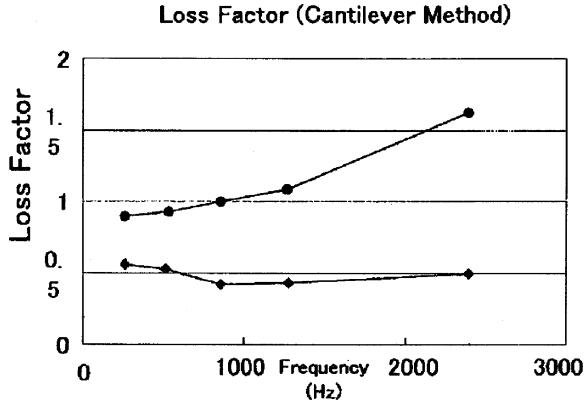
insert that can self-adjust the contact between the transducer's acoustic window and an ultrasonic target such as a human body, an animal, or any other object.—DRR

7,029,598

43.38.Ar COMPOSITE MATERIAL FOR PIEZOELECTRIC TRANSDUCTION

Masao Sato, assignor to Fuji Photo Film Company, Limited
18 April 2006 (Class 252/62.9R); filed in Japan 19 June 2002

This patent will be of some interest to acousticians, as it describes a piezoelectric composite material that is suitable for damping flexural vibrations. The emphasis here is on the chemistry of the matrix material that serves as the medium in which piezoelectric particles are embedded. The



authors find that liquid crystals of various types have good properties for such a matrix, and proceed to describe four liquid crystal polymers that function well as such a host material. The actual gains in damping over standard piezoelectric laminates appear to be modest. The nature of the piezoelectric material used is not disclosed.—JAH

7,029,850

43.38.Ar TRAVERSE SHEAR MODE PIEZOELECTRIC CHEMICAL SENSOR

Michael Thompson and Gordon L. Hayward, assignors to SencorChem International Corporation
18 April 2006 (Class 435/6); filed 29 September 2000

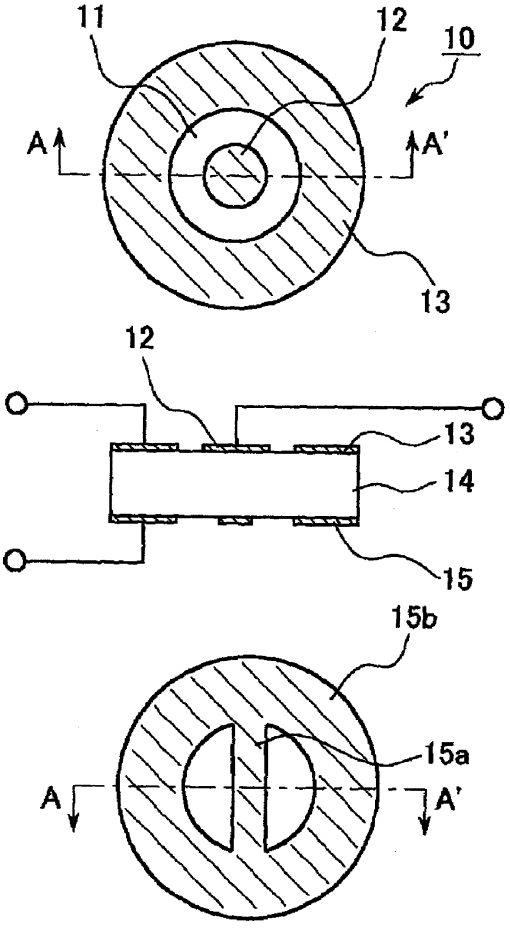
This patent describes a shear mode sensor that is used to monitor the binding of biological molecules to its sensitized surface. The resulting sensor can be used for highly specific detection and identification of biological molecules, DNA fragments, viruses, and the like. The patent is scarce on the acoustics of the system but does give a specific example of the detection signals for HIV binding. Not really a novel idea.—JAH

7,030,538

43.38.Ar PIEZOELECTRIC TRANSFORMER, PIEZOELECTRIC TRANSFORMER UNIT, INVERTER CIRCUIT, LIGHT EMISSION CONTROL DEVICE, AND LIQUID CRYSTAL DISPLAY DEVICE

Hiroshi Nakatsuka *et al.*, assignors to Matsushita Electric Industrial Company, Limited
18 April 2006 (Class 310/312); filed in Japan 14 June 2001

Can this be a record for the most claims included in a patent title? This patent discloses a piezoelectric step-up transformer of the Rosen type, but in a disk configuration. A piezoelectric disk is driven into radial oscillation by the voltage imposed on electrodes **12** and **15**, and a stepped-up output volt-



age is developed across **13** and **15**. The device is similar to many others of this type, but does have simplified fabrication possibilities. Unfortunately, no test results are given so we have to take the performance claims as unproven.—JAH

7,034,370

43.38.Bs MEMS SCANNING MIRROR WITH TUNABLE NATURAL FREQUENCY

Ting-Tung Kuo, assignor to Advanced Nano Systems, Incorporated
25 April 2006 (Class 257/414); filed 22 November 2002

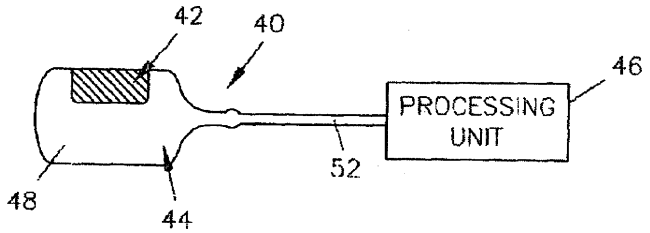
The author has designed an electrostatically driven MEMS mirror. This mirror and drive arrangement is distinguished by the presence of an additional electrode that is used to bias the electrostatic drive and change the effective spring constant of the mechanical suspension by acting in parallel with it. There is nothing new to any of this.—JAH

7,037,270

43.38.Fx SMALL ULTRASOUND TRANSDUCERS

James B. Seward, assignor to Mayo Foundation for Medical Education and Research
2 May 2006 (Class 600/459); filed 13 June 2003

The patent deals with miniaturized ultrasound transducers (e.g., less than $4\times4\times10$ mm) which are usable in small spaces, such as those within a surgical field or placed upon, within, attached to, embedded within, etc.



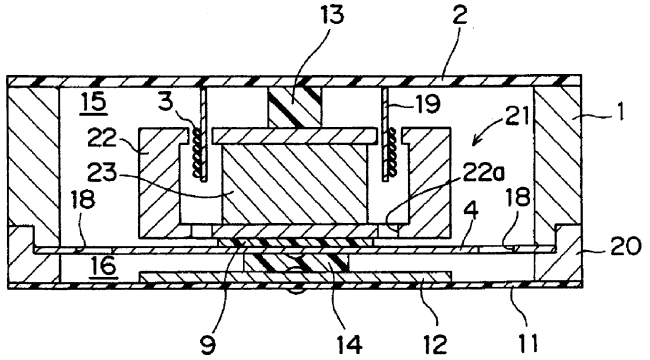
structures, organs, or devices. The ultrasonic transducer can be mounted onto or incorporated into a holding device to allow for easy manipulation of that transducer. The miniaturized transducer can be made to communicate with a processing unit wirelessly or via an electrical wire or cable.—DRR

7,006,651

43.38.Ja SPEAKER

Masataka Ueki, assignor to Uetax Corporation
28 February 2006 (Class 381/396); filed in Japan 26 February 2001

A waterproof loudspeaker is described whose main use appears to be for pagers and similar devices, although the patent discusses divers, which



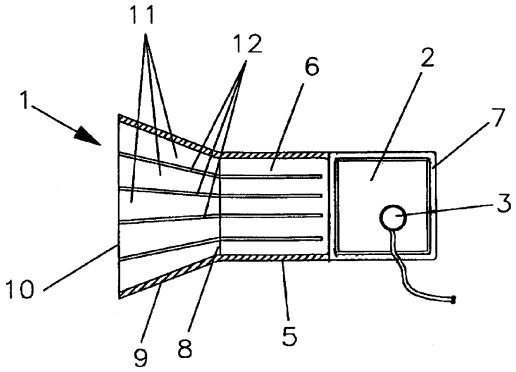
the reviewer reads to mean SCUBA. The device can also find use "in bad environments of air that includes dust and the like." The prose is dense, but the device appears to be workable.—NAS

7,010,138

43.38.Ja LOUDSPEAKERS

Neil Harris and Graham Bank, assignors to New Transducers Limited
7 March 2006 (Class 381/337); filed in the United Kingdom 6 November 1998

A distributed mode loudspeaker is mounted to various types of ducts in various ways. A variation that builds on simpler figures and claims shows the radiator **2** mounted such that sound radiates out to ducts **8**. By closing one duct, a chamber behind the radiator is formed. By rotating one duct **8** to the other side, sound is radiated in another direction. By extending duct **8**,



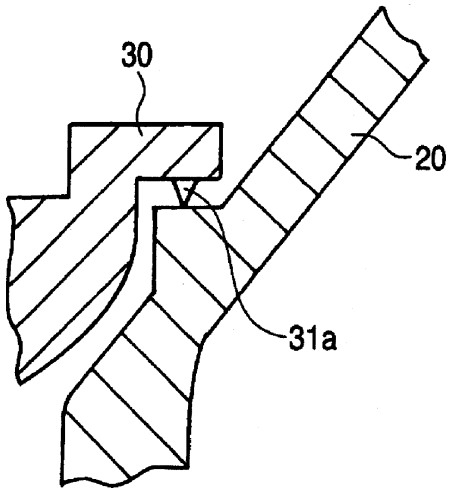
and adding openings along the length of the duct, a distributed sound source system can be made. And so on for 11-plus figures and 27 claims. An obvious use of this type of arrangement, which does not appear to be described or mentioned in the specifications or the claims, would be for active noise control in ducts, which you first heard here.—NAS

7,010,141

43.38.Ja SPEAKER DEVICE

Hideaki Sugiura *et al.*, assignors to Pioneer Corporation
7 March 2006 (Class 381/404); filed in Japan 6 December 2002

By using a spider holder (or damper holder **30** as used in the patent) that is supported at three places around the perimeter of the shoulder on speaker frame **20**, bad vibrations are canceled and good vibrations are enhanced. The damper holder **30** is attached using screws that lie below the damper. One assumes that the damper is placed on the damper holder after



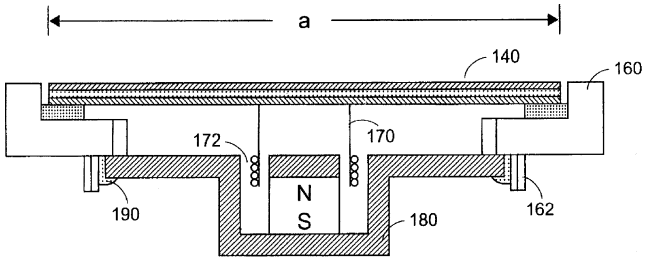
the screws are installed. In addition to the benefits listed in the specifications and claims of the patent, there is also the increase in the constituent parts for the assembly, the additional alignment of the damper and the three damper holder projections, additional care to insure that the adhesive that attached the damper to the holder does not bridge the gap between the holder and the frame, among others.—NAS

7,010,143

43.38.Ja RECTANGULAR PANEL-FORM LOUDSPEAKER AND ITS RADIATING PANEL

Tai-Yan Kam, Hsin Chu, Taiwan, Province of China
7 March 2006 (Class 381/426); filed 22 August 2002

Despite a lengthy prosaic recitation of what one would call a selective review of sound radiation from a flat panel, which invokes the hallowed name of Lord Rayleigh, and a detailed specification of the uniaxial laminate



B-B Cross section

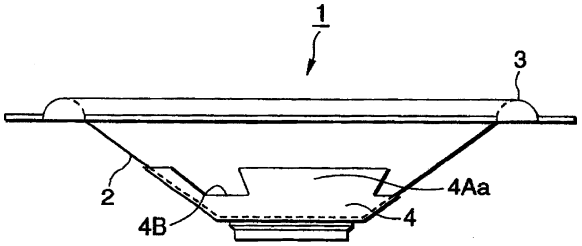
that is used for the panel skins, what is actually described appears to be very similar to inventions patented by NXT, Armstrong, and Slab, among many others.—NAS

7,027,609

43.38.Ja DIAPHRAGM FOR SPEAKERS

Yoshimi Kudo *et al.*, assignors to Pioneer Corporation
11 April 2006 (Class 381/423); filed in Japan 31 August 2000

A patent document written by an experienced patent attorney can be almost impossible for a layman to decipher. In the case at hand, the patent appears to have been written by a Japanese patent attorney and then translated into English. In spite of this, a careful reading of the claims section in relation to the illustrations yields a pretty good idea of what has been pat-



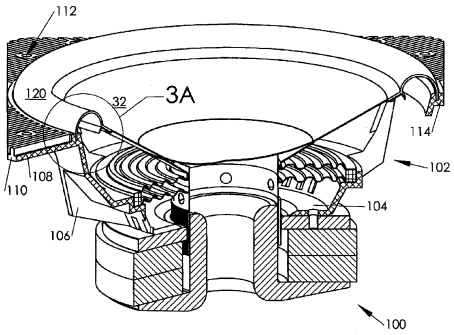
ented. Loudspeaker cone **2** is made of metal. An irregular auxiliary cone **4** is bonded to the inner portion of the metal cone. This technique is said to attenuate the sharp high-frequency peak that normally shows up in the response of a metal cone. However, what is patented is not the general idea of such an auxiliary cone but rather an auxiliary cone made of paper—the claims are quite specific about this point.—GLA

7,031,487

43.38.Ja TABBED SPEAKER FRAME WITH OVERSIZED DIAPHRAGM

Enrique M. Stiles, assignor to STEP Technologies, Incorporated
18 April 2006 (Class 381/398); filed 14 May 2003

This loudspeaker design includes two important features. The first—a square (tabbed) frame—is well-known prior art dating back at least 30 years. However, the second feature is both novel and interesting. The outer



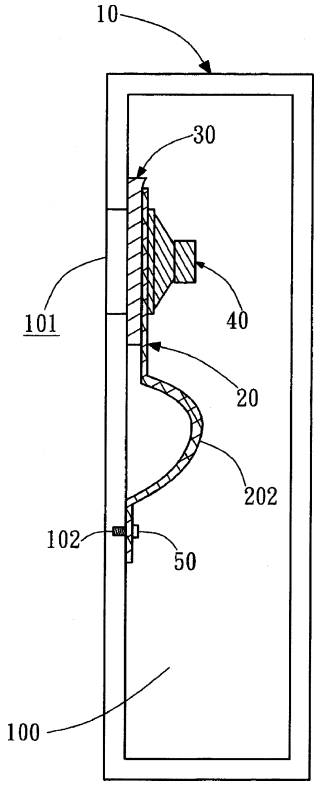
cone suspension terminates in vertical slot **114** rather than a flat flange. This may offer performance advantages in addition to simply saving a bit of space.—GLA

7,035,419

43.38.Ja ANTI-RESONANT STRUCTURE FOR SPEAKERS

Ching-Hsiang Yu *et al.*, assignors to Uniwill Computer Corporation
25 April 2006 (Class 381/353); filed 9 January 2004

This miniature loudspeaker system attempts to minimize cabinet vibra-



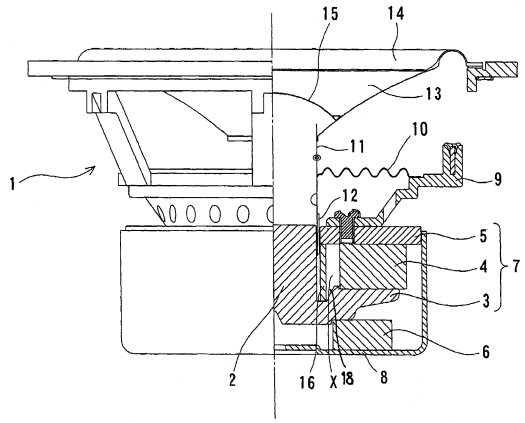
tion by including a resilient sealing ring **30** and then holding the speaker in place with spring clip **202**.—GLA

7,031,489

43.38.Ja MAGNETIC CIRCUIT FOR SPEAKER WITH SHORT-CIRCUITING RING

Kenta Amino, assignor to Minebea Company, Limited
18 April 2006 (Class 381/414); filed in Japan 28 August 2002

A shorting ring is sometimes inserted into the gap of a moving coil loudspeaker to increase high-frequency efficiency. The ring acts as a shorted



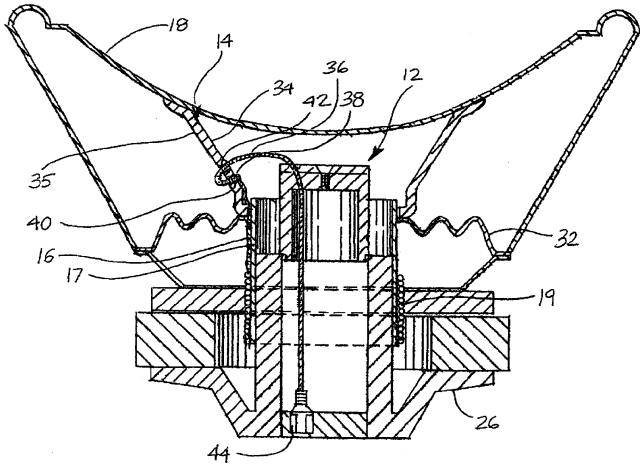
turn, counteracting the inductance of the coil and thereby reducing impedance in the high-frequency region. The trick is to position the shorting ring close to the voice coil, where it is effective, yet not significantly increase the width of the magnetic gap. A number of patented geometries address this trade-off. This patent describes a shorting ring **16** that takes the form of an aluminum or copper cylinder inserted below the magnetic gap. More specifically, as set forth in the patent claims, it ''fits into grooves located at the bottom face of the top plate and the top face of the bottom yoke.''—GLA

7,035,424

43.38.Ja LOUDSPEAKER HAVING AN INNER LEAD WIRE SYSTEM AND RELATED METHOD OF PROTECTING THE LEAD WIRES

Eugene P. Brandt, West Paducah, Kentucky
25 April 2006 (Class 381/409); filed 9 April 2002

It is difficult to pin down exactly what has been patented here. An otherwise conventional moving coil loudspeaker has its electrical connections brought out through the center pole piece, as is done in some commer-



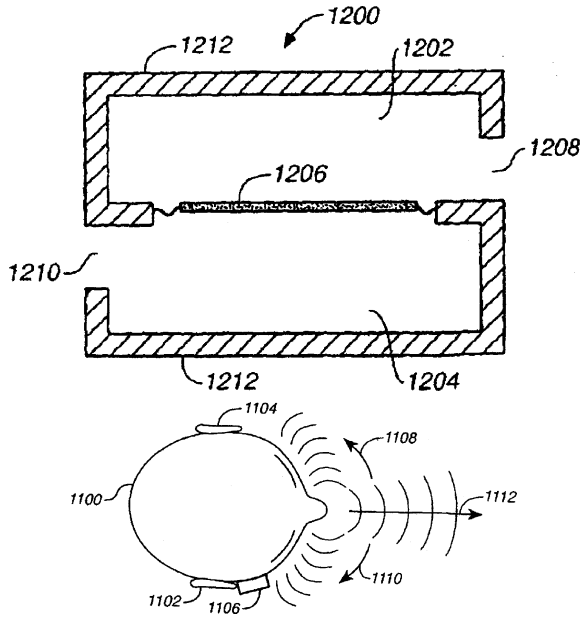
cial loudspeakers. However, the patent claims describe five possible embodiments, most of which include some kind of ''guide'' or ''adapter'' to hold the wires in place.—GLA

7,027,603

43.38.Kb EAR LEVEL NOISE REJECTION VOICE PICKUP METHOD AND APPARATUS

Jon C. Taenzer, assignor to GN Resound North America Corporation
11 April 2006 (Class 381/92); filed 21 February 2003

The patent describes a small first-order gradient microphone located at the side of the user's head and whose principal axis is oriented fore-aft.



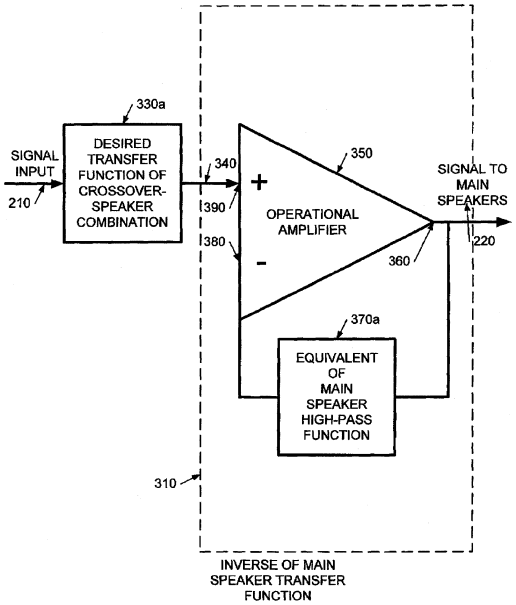
Because of its proximity to the mouth of the user, it can further be equalized to act as a conventional noise-canceling microphone. Because of its location at the side of the user's head it is well out of the way of interfering breath blasts and should work very well in typical broadcast news gathering applications.—JME

7,003,124

43.38.Lc SYSTEM AND METHOD FOR ADJUSTING FREQUENCY RESPONSE CHARACTERISTICS OF HIGH-PASS CROSSOVERS SUPPLYING SIGNAL TO SPEAKERS USED WITH SUBWOOFERS

James Thiel, assignor to Thiel Audio Products
21 February 2006 (Class 381/99); filed 20 April 2001

A means is described of building and adjusting a crossover network, by use of, but not necessarily limited to, the characteristic frequency, Q, and enclosure type of the main speaker, so that the low-frequency response of



the main loudspeaker system better ''blends'' with the sound output of a subwoofer. The adjustment is accomplished using a microcontroller to translate the input parameters into the equivalent resistors and capacitors needed to tune the filters. A clear discussion of the filter design is included in the

body of the patent, as well the use of an equivalent-resistance circuit that can be used for some of the resistors in the filter circuits. The microcontroller can adjust the effective resistances in the equivalent-resistance circuit.—NAS

7,013,011

43.38.Lc AUDIO LIMITING CIRCUIT

William A. Weeks and William R. Morrell, assignors to Plantronics, Incorporated
14 March 2006 (Class 381/98); filed 28 December 2001

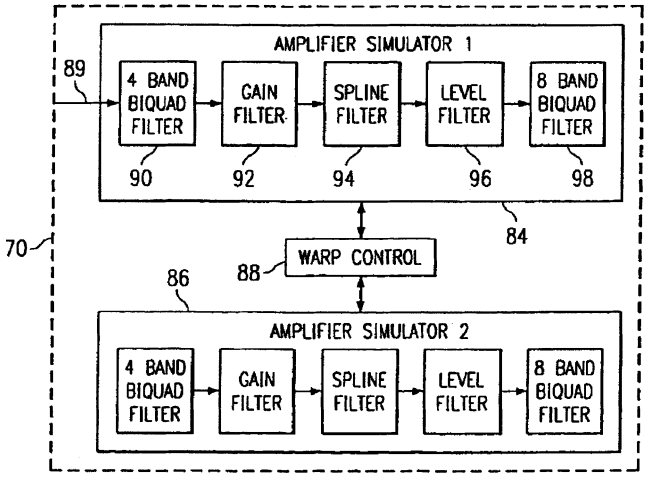
A smart limiting circuit is proposed which would decrease the amplitude of a signal according to predicted output sound pressure from a known loudspeaker system, thereby providing a new output that is specifically tailored to the frequency and amplitude response of the reproduction system.—SAF

7,026,539

43.38.Lc MUSICAL EFFECT CUSTOMIZATION SYSTEM

James D. Pennock *et al.*, assignors to Harman International Industries, Incorporated
11 April 2006 (Class 84/662); filed 19 November 2003

Once again, the issue is modeling the effect of "classic" tube/value amplifiers and loudspeakers via DSP techniques. A combination of FIR **88**



and IIR **90**, **98** filters can be used to filter the input. In addition, "blends" can be customized by the user.—MK

7,027,239

43.38.Md DATA ACQUISITION SYSTEM

William B. Priester, Memphis, Tennessee *et al.*
11 April 2006 (Class 360/6); filed 13 March 2002

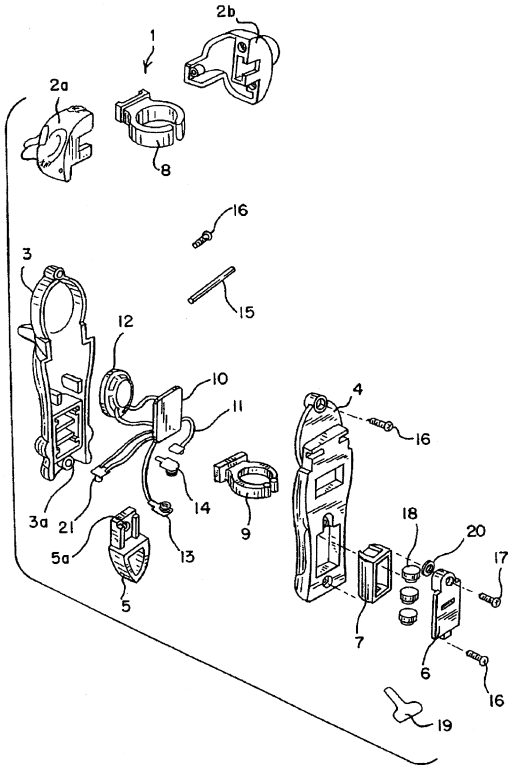
This is a simple proposal: use an existing sound card as data acquisition system. But, before we do that, let us convert the input voltage to frequency (via an LM331 V/F converter). Now the original voltage can be recovered via a FFT. Seems like a lot of trouble where a simple amplifier would suffice.—MK

7,029,361

43.38.Md FINGER PUPPETS WITH SOUNDS

Amy M. Seibert *et al.*, assignors to The Marketing Store Worldwide, L.P.
18 April 2006 (Class 446/327); filed 9 September 2003

Place a ubiquitous sound chip inside a finger puppet. The printed circuit board **10** is connected to the speaker **12** and finger switch **21**. The batteries **18** are visible inside the compartment. If each finger was a note in a scale, then with two toes a full scale would be possible.—MK



6,967,276

43.38.Si PORTABLE TELEPHONY APPARATUS WITH MUSIC TONE GENERATOR

Tsuyoshi Futamase *et al.*, assignors to Yamaha Corporation
22 November 2005 (Class 84/622); filed in Japan 28 July 1999

Cell phones have always had reasonably capable audio output devices in order to produce audible speech. Then, why did so many early phones use a harsh monotone buzzer to produce ring tones? One reason is that it takes a lot of memory to store an audio waveform. This patent represents a response to that situation. For the sake of storage efficiency, the sound is still produced by a parametric mechanism and, so, is still subject to the limits of what the parameters can generate. But, that includes simultaneous tones, along with fairly elaborate mechanisms to control timing and pitch. Options are also available to allow karaoke and video sync. The device would drive the speaker, not a buzzer.—DLR

7,027,604

43.38.Si CIRCUIT TO PREVENT ACOUSTIC FEEDBACK FOR A CELLULAR SPEAKERPHONE

Kee Eng Soo *et al.*, assignors to Motorola, Incorporated
11 April 2006 (Class 381/93); filed 24 September 2001

Most cellular telephones can be used with a speakerphone accessory for hands-free operation. When used as a speakerphone, the earpiece is disabled and an external speaker takes its place. Concurrently, the internal

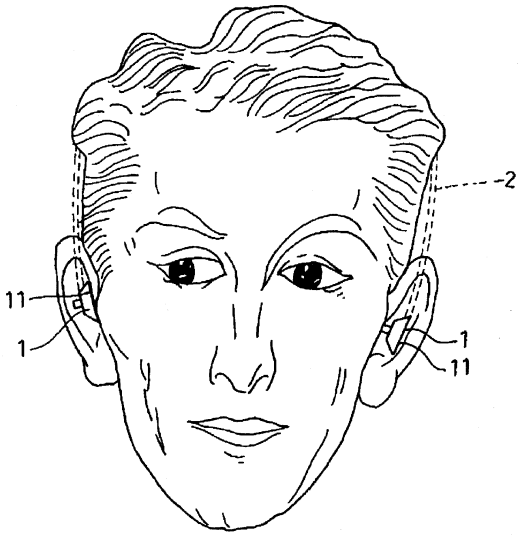
microphone is switched off and an external microphone is enabled. However, many cellular phones include a short delay when switching off the microphone. Thus, during the transition from earpiece to external speaker, acoustic feedback can be triggered, causing an annoying squeal. One obvious solution would be to briefly mute the speakerphone, and this patent describes a practical method for doing just that.—GLA

7,035,421

43.38.Si EARPHONE SET

Hsi Kuang Ma, Taipei, Taiwan, Province of China
25 April 2006 (Class 381/372); filed in Taiwan, Province of China 5 December 2001

Headset transducers **1** are oriented to point in the same direction. That is, the sound output of one transducer is directed toward the ear while that of the other transducer is directed away from the other ear. "In this way, the



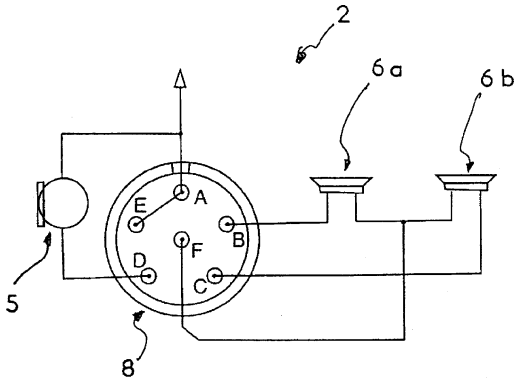
sound signals can move in the same direction to enhance the clearness of the sound out in addition to removing the ear pain." Actually, it is a headache that bothers me after reading the patent.—GLA

7,035,597

43.38.Si UNIVERSAL COMMUNICATION DEVICE

Bernard Maden, assignor to Gallet SA
25 April 2006 (Class 455/90.2); filed in France 12 December 2000

A military intercom set may be used with a variety of microphones and headsets. If these are connected to appropriate pins of a universal connector,



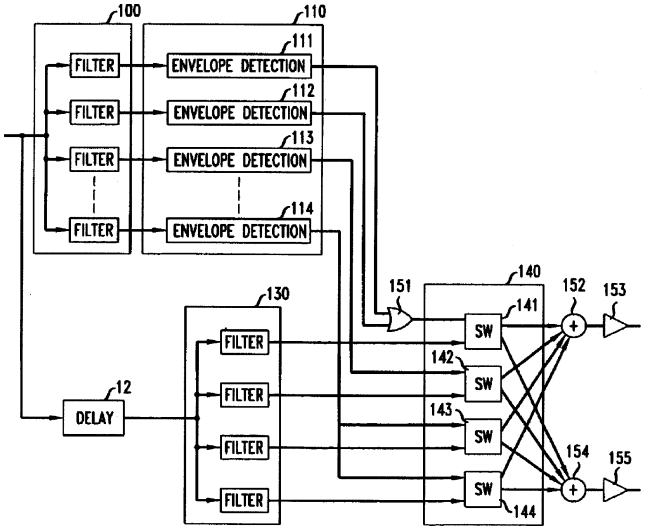
then it becomes possible to "automatically configure the electronic box without requiring special keypads or selector knobs." Such a scheme is worked out in considerable detail in this short patent.—GLA

7,027,601

43.38.Vk PERCEPTUAL SPEAKER DIRECTIVITY

James David Johnston, assignor to AT&T Corporation
11 April 2006 (Class 381/56); filed 3 May 2000

Filter banks **100** and **130** correspond more or less to the critical bands of human hearing. At higher frequencies, filter bank **100** detects leading edges in the signal's envelope. The presence of a strong leading edge implies a directional source, and the output of the corresponding reconstruction filter **130** is routed through amplifier **153** to a directional loudspeaker. If



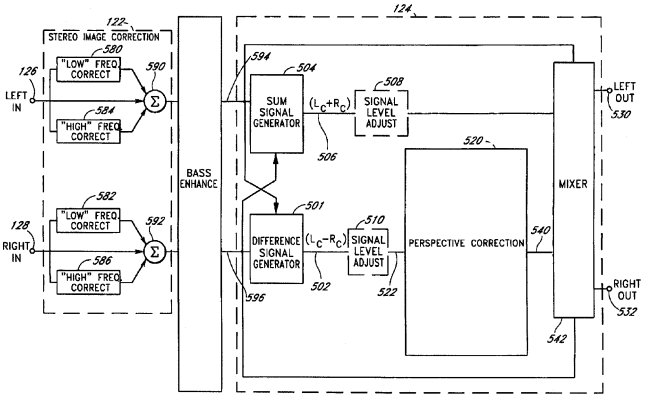
there is no strong leading edge, the signal is assumed to be nondirectional and is routed through amplifier **155** to a diffuse-source loudspeaker. Although not specifically mentioned in the patent, this might be an interesting way to extract a surround sound signal from conventional two-channel stereo program material.—GLA

7,031,474

43.38.Vk ACOUSTIC CORRECTION APPARATUS

Thomas C. K. Yuen *et al.*, assignors to SRS Labs, Incorporated
18 April 2006 (Class 381/1); filed 4 October 1999

This is a long patent. It includes more than 40 figures, 24 pages of descriptive text, and 35 claims. Like dozens of earlier patents in this field, it describes a method for enhancing the reproduction of conventional two-channel stereo program material. The method described here is a fairly



complicated mixture. It includes some multiband processing, a dab of dynamic expansion, a bit of HRTF-based image correction, and a healthy serving of synthetic bass boost. Unlike many patents, the claims provide a clear description of what has been patented. However, what is described seems to be familiar prior art in more than one instance.—GLA

7,030,718

43.38.WI APPARATUS AND METHOD FOR EXTENDING TUNING RANGE OF ELECTRO-ACOUSTIC FILM RESONATORS

Dieter Scherer, assignor to National Semiconductor Corporation
18 April 2006 (Class 333/188); filed 9 August 2002

This patent describes a circuit for use with a film bulk acoustic resonator that extends the tuning range of such a device. It comprises an inductor, a varactor diode, and the FBAR to be tuned. There is not much new here, but the exposition is clear and detailed for those who want to see how it is done.—JAH

7,028,969

43.40.Tm SEISMICALLY RESTRAINED VIBRATION ISOLATING MOUNTING DEVICE

Paul W. Meisel and Richard S. Sherren, assignors to Kinetics Noise Control, Incorporated
18 April 2006 (Class 248/638); filed 14 August 2002

A three-axis snubber that is attached to the isolated object is arranged near the support housing of an isolator so that the object's (and the snubber's) travel is restricted if the object's motion exceeds the clearance space between the snubber and the housing. In a representative configuration the snubber is suspended from a rod that extends through a spring; the latter provides vertical isolation and the rod acts like a pendulum to provide horizontal isolation.—EEU

7,028,997

43.40.Tm VIBRATION DAMPING TOOL

Hidebumi Takahashi and Yoichi Ishikawa, assignors to Mitsubishi Materials Corporation
18 April 2006 (Class 267/137); filed in Japan 13 June 2001

A cylindrical cavity that houses a dynamic absorber is provided in an essentially cylindrical holder for a machining tool. This absorber consists of a mass that is enclosed by a visco-elastic material.—EEU

7,032,723

43.40.Tm BRAKE ASSEMBLY WITH TUNED MASS DAMPER

Ronald Louis Quaglia *et al.*, assignors to Ford Global Technologies, LLC
25 April 2006 (Class 188/73.37); filed 22 October 2002

A dynamic absorber is located in a hole in a brake component, such as a backplate that supports a brake pad, in order to suppress brake squeal noise. The absorber consists of a resiliently supported mass and is tuned to the frequency of the noise that is to be suppressed. Locating the absorber in a hole has packaging and manufacturing advantages and also makes the absorber less susceptible to damage in use.—EEU

7,033,140

43.40.Tm COOLED ROTOR BLADE WITH VIBRATION DAMPING DEVICE

Shawn J. Gregg, assignor to United Technologies Corporation
25 April 2006 (Class 416/135); filed 19 December 2003

The device described in this patent differs only in minor ways from that described in Unites States Patent 6,929,451 [reviewed in J. Acoust. Soc. Am. **119**(2), 688 (2006)]. It deals with a ''stick damper'' for a rotor blade that is provided with gas-flow passages and orifices, with the stick damper arrangement configured so that it interferes minimally with the cooling gas flow.—EEU

7,027,353

43.40.Yq METHOD AND APPARATUS FOR REAL-TIME VIBRATION IMAGING

Philip Melese *et al.*, assignors to SRI International
11 April 2006 (Class 367/7); filed 2 April 2004

A detector array is configured to receive light or other radiation reflected from a target object. The signals correspond to the received radiation, which is modulated as the object vibrates, high-pass filtered to remove the effects of ambient radiation, and sampled at a predetermined frequency, such as 1 kHz. The data are stored periodically, such as once per second, and then subjected to various types of analysis, such as Fourier analysis. Since each detector of the array receives radiation from a particular region of the target object, a series of full images of the object's vibrations can be generated. The apparatus may be used, for example, to detect problems with machinery or for seismic exploration.—EEU

7,028,015

43.50.Ki FILTERING DEVICE AND METHOD FOR REDUCING NOISE IN ELECTRICAL SIGNALS, IN PARTICULAR ACOUSTIC SIGNALS AND IMAGES

Rinaldo Poluzzi *et al.*, assignors to STMicroelectronics S.r.l.
11 April 2006 (Class 706/1); filed in the European Patent Office 29 November 2000

This patent presents one of the more complicated entries in the adaptive noise-filtering genre. The signal is filtered by reconstruction as a moving average model whose weights are set by a complicated scheme involving fuzzy logic, which responds to the noise or other signal corruption (such as competing signals) in a neural network implementation—a ''neuro-fuzzy network.'' The stated goal is to be able to preserve ''steep edges'' in the target signal which would be filtered out by standard adaptive filtering techniques.—SAF

7,035,796

43.50.Ki SYSTEM FOR NOISE SUPPRESSION, TRANSCEIVER AND METHOD FOR NOISE SUPPRESSION

Ming Zhang and Hui Lan, assignors to Nanyang Technological University
25 April 2006 (Class 704/226); filed 6 May 2000

This is a rather strange patent in that the noise cancellation method described appears to be specific to a particular noisy environment which includes ''narrow band noise from rotating machine'' [sic] and audio signals from loudspeakers. The method involves somehow determining the noise signals directly from the offending devices, and using this information to set the parameters of typical adaptive filters, together with a third adaptive filtration scheme to cope with additional ambient noise.—SAF

6,968,738

43.58.Gn ACOUSTIC FLUID-GAUGING SYSTEM

Harry Atkinson, assignor to Smiths Group plc
29 November 2005 (Class 73/290 V); filed in the United Kingdom 2 October 2001

This patent describes an ultrasonic fuel gauge for use in aircraft fuel tanks. A probe immersed below the operating fuel level in the tank transmits a pulse toward the surface and measures the reflected signal. Various analy-

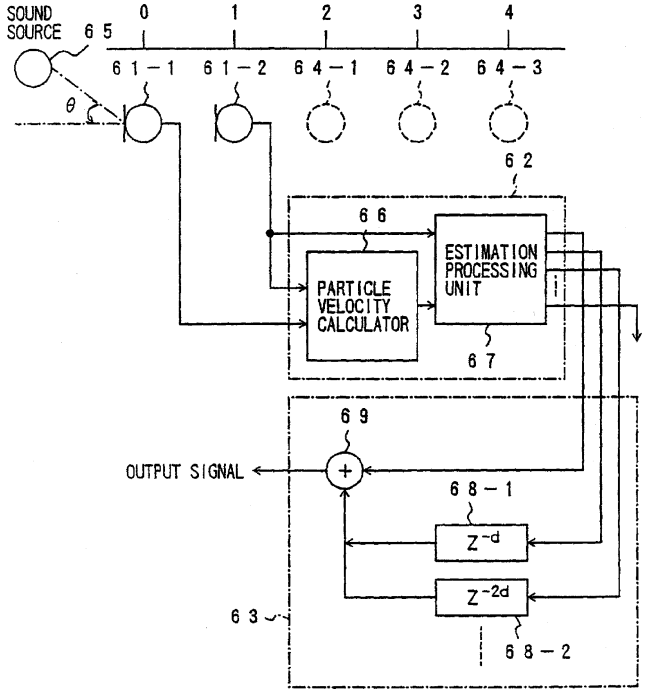
ses are performed on the received signal, including comparisons with reflections from multiple probes in the same tank, to determine the fuel level.—DLR

7,035,416

43.60.Bf MICROPHONE ARRAY APPARATUS

Naoshi Matsuo, assignor to Fujitsu Limited
25 April 2006 (Class 381/92); filed in Japan 26 June 1997

This patent deals with adaptive microphone arrays for use in video conferencing, a class of developments we are seeing more of in these pages.



Methods are described for converging on an arbitrary sound source while ''stably and precisely suppressing noise, emphasizing a target sound and identifying the position of a sound source.''—JME

7,035,417

43.60.Cg SYSTEM FOR REDUCING NOISE IN THE REPRODUCTION OF RECORDED SOUND SIGNALS

Thomas N. Packard, Syracuse, New York
25 April 2006 (Class 381/94.1); filed 5 April 1999

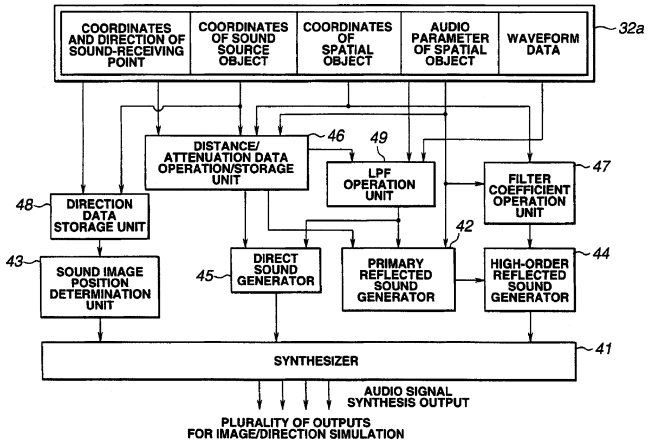
Those of you who are old enough may remember the spate of transient noise reduction devices that were introduced during the sunset years of the LP record. Some of these devices compared a signal with that same signal, slightly delayed, toggling between the two as required in order to minimize the audible effect of a tick or pop. This patent describes various elaborations on a tried and true process that may have uses well beyond the playback of mechanical recordings.—JME

7,027,600

43.60.Ek AUDIO SIGNAL PROCESSING DEVICE

Toshiyuki Kaji *et al.*, assignors to Kabushiki Kaisha Sega
11 April 2006 (Class 381/17); filed in Japan 16 March 1999

This patent deals with synthesizing and enhancing of sound effects for computer games. A given sound effect is built from the bottom up, so to speak, through the interaction of a number of spatial coordinates and basic



waveform data, along with a number of modifying functions. The aim is to increase realism and speed of response to rapidly changing commands. The conceptual nature of this process is shown in the figure.—JME

7,028,559

43.60.Mn METHOD OF AND A DEVICE FOR ACOUSTICALLY MONITORING THE COURSE OF A PROCESS, SUCH AS A MILKING PROCESS

Dick M. Oort and Karel Van den Berg, assignors to Lely Research Holding AG a Swiss Limited Liability Company
18 April 2006 (Class 73/861.18); filed 14 May 2002

The patent relates to a method of monitoring a process whereby the amplitude differences or the intensities of the sound, or both, are measured. These characteristics of the sound or vibration are measured continuously or with an adjustable sample rate by means of a sound and/or vibration sensor, such as a piezoelectric transducer, during the entire process or part of the process, and are compared mutually and/or with a predetermined threshold value or reference pattern for the purpose of establishing conclusions relating to the course of the process.—DRR

7,028,633

43.60.Mn DEVICE FOR KEEPING BIRDS AWAY WITH DIFFERENTIAL MANAGEMENT FUNCTIONS

Marco Pinton and Luciano Santarelli, assignors to Aviotek Engineering S.r.l.
18 April 2006 (Class 116/22 A); filed in Italy 23 January 2003

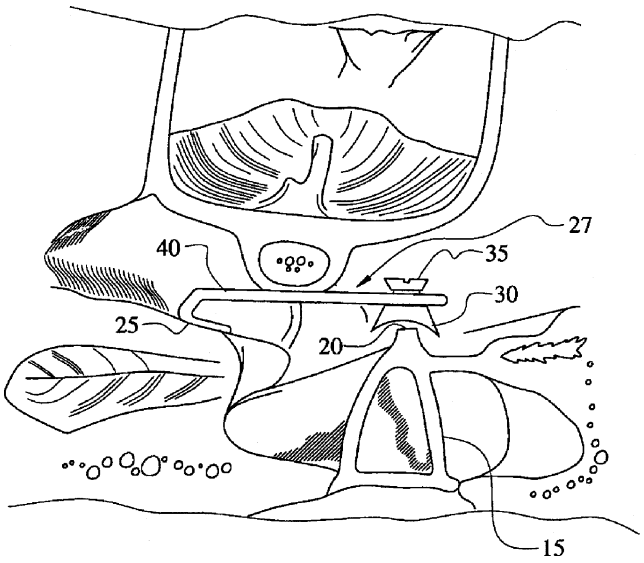
This is a device for keeping birds away from a designated region through differential management functions. The device consists of a control unit connected to a pilot system that controls one or more emission units. The emission units may emit sound, or light, or both. The pilot system and each emission unit consist of a casing with an optional protection cover, a control circuit, and one or more light and sound emitters. The control unit activates the light and sound emitters, receives operating instructions from the pilot system, and verifies the correctness of the operation of all components in the unit.—DRR

7,025,785

43.66.Ts INCUS REPLACEMENT PROSTHESIS

K. Paul Boyev, assignor to University of South Florida
11 April 2006 (Class 623/10); filed 30 December 2004

This device is designed to restore hearing to individuals who have a discontinuity in the middle-ear sound-conduction mechanism. The device is meant to address a specific problem arising often in middle-ear surgery. Currently, middle-ear prostheses may be inadequate to remedy the specific problem of a lateral relationship of the stapes capitulum to the malleus,



thereby necessitating a cartilage graft that can result in poor sound-conductive properties. This auditory prosthesis, adapted to be inserted into the ear, includes an adjustable pivot element configured to contact the stapes capitalum, a positioning element integral to the adjustable pivot element, and an anchoring device (with two legs) that forms a bearing surface to contact the inner wall of the tympanum.—DRR

7,027,607

43.66.Ts HEARING AID WITH ADAPTIVE MICROPHONE MATCHING

Brian Dam Pedersen and René Mortensen, assignors to GN Resound A/S
11 April 2006 (Class 381/313); filed in Denmark 22 September 2000

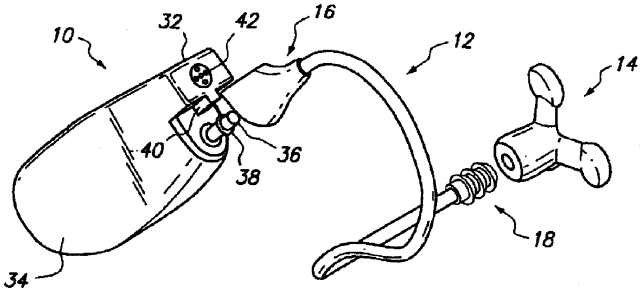
In order to maintain maximum directional performance using two omnidirectional microphones and two amplifier channels, signal processing circuitry senses the difference over time of the average signal levels of the microphone outputs and makes corrections as a function of frequency to keep the two channels matched.—DAP

7,027,608

43.66.Ts BEHIND THE EAR HEARING AID SYSTEM

Robert J. Fretz *et al.*, assignors to GN ReSound North America
11 April 2006 (Class 381/330); filed 17 July 1998

The normal acoustic termination consisting of an ear hook attached to a hearing aid receiver output nozzle and connecting through tubing to a



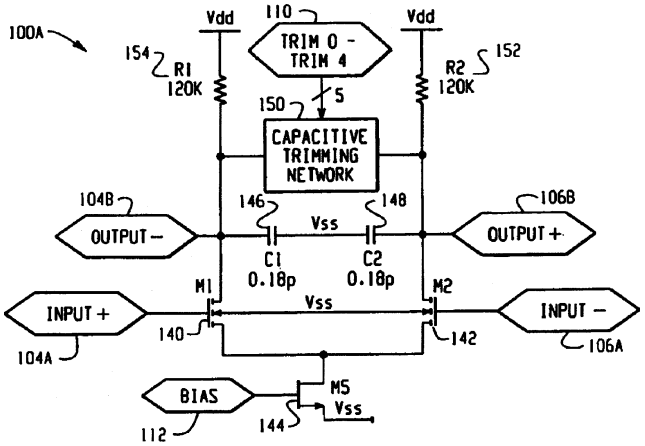
custom earmold is replaced by a preformed semirigid tubing attached to a noncustom eartip.—DAP

7,031,482

43.66.Ts PRECISION LOW JITTER OSCILLATOR CIRCUIT

Wei Yang and Frederick Edward Sykes, assignors to Gennum Corporation
18 April 2006 (Class 381/312); filed 10 October 2003

Differential inverters such as used in hearing aids are configured in a feedback loop to generate a clock signal. A capacitive trimming network in



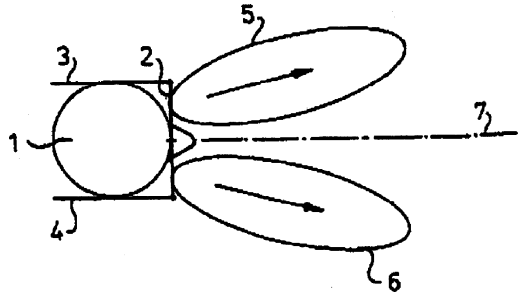
the differential inverters allows the frequency of the clock to be adjusted and resistive loads are used to minimize jitter.—DAP

7,031,483

43.66.Ts HEARING AID COMPRISING AN ARRAY OF MICROPHONES

Marinus Marias Boone *et al.*, assignors to Technische Universiteit Delft
18 April 2006 (Class 381/313); filed in the Netherlands 20 October 1997

Building on the work of Soede, two beam signals are derived by summing the outputs of multiple microphones configured on the hearing aid wearer as an end-fire array and/or a broadside array. The main sensitivity



directions of the two outputs occur symmetrically at an angle to the left and to the right relative to the main axis of the array. The left and right signals are selectively amplified and fed to the wearer's left and right ear, respectively.—DAP

7,031,481

43.66.Ts HEARING AID WITH DELAYED ACTIVATION

Rene Mortensen, assignor to GN Resound A/S
18 April 2006 (Class 381/312); filed 7 February 2003

To avoid acoustic feedback oscillation while the hearing aid is being placed on the wearer's head or in an ear of the wearer, the hearing aid circuit

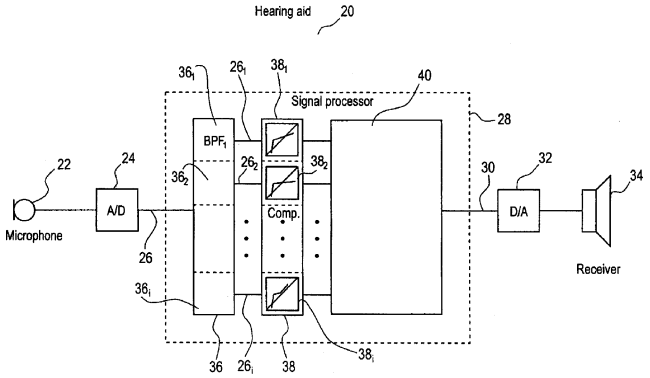
temporarily blocks the signal path following power on for an adjustable delay time. To notify the wearer of the duration of the delay, the hearing aid emits a special power-on delay signal.—DAP

7,031,484

43.66.Ts SUPPRESSION OF PERCEIVED OCCLUSION

Carl Ludvigsen, assignor to Widex A/S
18 April 2006 (Class 381/316); filed in Denmark 13 April 2001

To equalize out the head-shadow effect on the hearing-aid wearer's own voice, the low-frequency acoustic input level is decreased if the wearer is speaking by increasing the compression ratio of the hearing-aid amplifier



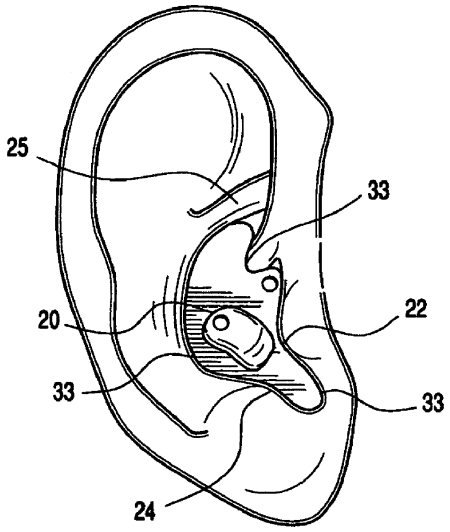
and optionally applying an offset gain in at least one low-frequency channel. The system analyzes the input signal to determine if the wearer's own voice is present so as to not apply low-frequency reduction of the speech signals from others.—DAP

7,025,061

43.66.Vt CUSTOMIZED PASSIVE HEARING PROTECTION EARPLUG, USE OF THE SAME AND METHOD FOR MANUFACTURING THE SAME

Mathias Haussmann, assignor to Phonak AG
11 April 2006 (Class 128/864); filed 25 August 2004

By manufacturing a custom-fit earmold from a pattern determined by a laser scan of the cavum concha and a subsequent buildup of material, a passive earplug is constructed that can be partly displaced to permit communication. The earplug is held in place by the rim of the cavum concha, which forces it to return to the fully inserted position when a tab **20** is released. Communication circuitry can also be incorporated in the earplug.—JE



7,013,276

43.72.Ar METHOD OF ASSESSING DEGREE OF ACOUSTIC CONFUSABILITY, AND SYSTEM THEREFOR

Corine A. Bickley and Lawrence A. Denenberg, assignors to Comverse, Incorporated
14 March 2006 (Class 704/255); filed 5 October 2001

Here, we find a rare combination of an obvious method which is public domain, and a largely ineffective one which is prior art. The former is the idea of using the minimum edit distance between two phoneme strings to measure confusability between the two. The latter is the idea to apply a published method of acoustic feature detection to convert speech signals into strings of acoustic features, and to then also use minimum edit distance to measure the confusability.—SAF

7,016,832

43.72.Ar VOICED/UNVOICED INFORMATION ESTIMATION SYSTEM AND METHOD THEREFOR

Yong Soo Choi, assignor to LG Electronics, Incorporated
21 March 2006 (Class 704/208); filed in the Republic of Korea 22 November 2000

This incomprehensible document might be proposing a method for measuring the degree to which a given speech frame is voiced, by somehow comparing the spectral energy in harmonics of the input to those in a corresponding synthesized signal. But, it is hard to say anything for certain about this patent.—SAF

7,016,839

43.72.Ar MVDR BASED FEATURE EXTRACTION FOR SPEECH RECOGNITION

Satayanarayana Dharanipragada and Bhaskar Dharanipragada Rao, assignors to International Business Machines Corporation
21 March 2006 (Class 704/245); filed 31 January 2002

The authors appear to be patenting a speech recognition front end which calculates the usual sorts of acoustic features such as linear predictive, cepstral, etc. using a promising (according to the authors) new spectral modeling technique found in the literature and known as minimum variance distortionless response (MVDR). This technique seeks to minimize the variance in a sequence of acoustic features over successive frames. Since the MVDR method was made public in a paper about speech modeling, it is surprising a patent could be granted which applies it in a speech recognition context and which adds little more than a few standard things.—SAF

7,027,980

43.72.Ar METHOD FOR MODELING SPEECH HARMONIC MAGNITUDES

Tenkasi V. Ramabadran *et al.*, assignors to Motorola, Incorporated
11 April 2006 (Class 704/217); filed 28 March 2002

A procedure is described for iteratively improving a linear predictive signal model to determine the magnitudes of harmonics as correctly as possible. The procedure is described in some detail, but it involves a complicated sequence of calculations which includes the unusual computation of a ''pseudo auto-correlation sequence'' resulting from an inverse Fourier transform of spectral harmonic magnitudes, and then performing linear prediction on this sequence.—SAF

7,035,792

43.72.Ar SPEECH RECOGNITION USING DUAL-PASS PITCH TRACKING

Eric I-Chao Chang and Jian-Lai Zhou, assignors to Microsoft Corporation
25 April 2006 (Class 704/207); filed 2 June 2004

A method for pitch tracking is described which uses the output from a first pitch tracking algorithm as a set of candidate pitches for a second more computationally expensive pitch tracking algorithm. The first algorithm suggested is the ''average magnitude difference function'' technique, which is cheap but less than effective. The second algorithm is suggested as a ''normalized cross-correlation'' technique, which is more expensive but would operate on a smaller sample size than the first-pass algorithm.—SAF

7,013,272

43.72.Dv AMPLITUDE MASKING OF SPECTRA FOR SPEECH RECOGNITION METHOD AND APPARATUS

Changxue Ma, assignor to Motorola, Incorporated
14 March 2006 (Class 704/227); filed 14 August 2002

The general idea here is to mitigate noise in speech signals by means of spectral masking. One possible method involves downscaling those parts of the spectrum that fall below a threshold, thereby literally clipping the valleys between spectral peaks. It is not clear how this idea would extend to noise of comparable amplitude to the speech.—SAF

7,024,359

43.72.Fx DISTRIBUTED VOICE RECOGNITION SYSTEM USING ACOUSTIC FEATURE VECTOR MODIFICATION

Chienchung Chang *et al.*, assignors to Qualcomm Incorporated
4 April 2006 (Class 704/251); filed 31 January 2001

This patent pursues a currently popular aim, namely that of somehow applying speaker-independent speech recognition approaches in a speaker-dependent way. The acoustic features of the speech received are used to select an adaptation model from a database to help characterize the speaker. The features are then modified by some transformation based on this model to put them into a form more amenable to the general speech recognition algorithm. The patent is short on details, and contains no mathematics.—SAF

7,024,366

43.72.Fx SPEECH RECOGNITION WITH USER SPECIFIC ADAPTIVE VOICE FEEDBACK

Scott A. Deyoe and Tuan A. Hoang, assignors to Delphi Technologies, Incorporated
4 April 2006 (Class 704/270.1); filed 14 January 2000

The authors propose to augment a speech recognition system with multiple response possibilities, in a fashion that adapts itself to a particular recognized user by selecting the most appropriate responses based upon the user's prior patterns of input. This would entail the ability to indeed recognize a particular speaker, a fact which is mentioned in the patent, though not addressed.—SAF

7,031,926

43.72.Gy SPECTRAL PARAMETER SUBSTITUTION FOR THE FRAME ERROR CONCEALMENT IN A SPEECH DECODER

Jari Mäkinen *et al.*, assignors to Nokia Corporation
18 April 2006 (Class 704/500); filed 30 July 2001

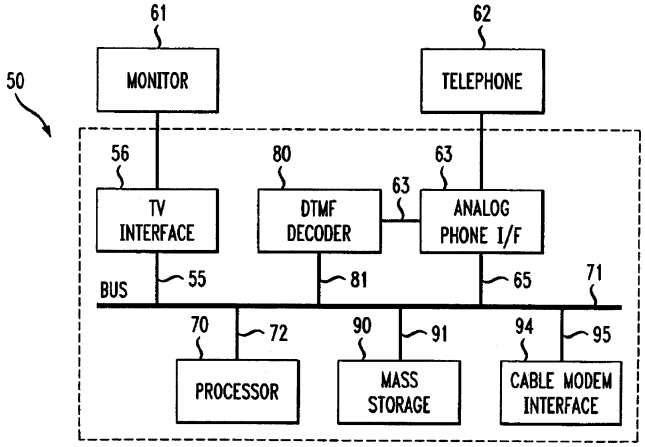
Transmission of coded speech over wireless networks involves inevitable lost or corrupted speech frames. This patent presents a technique for interpolating lost speech frames from previous good frames, and for using partial information that is received in corrupted frames to further optimize the ''concealment procedure.'' The method is presented as an equation, which presumably speaks for itself, but evaluation results are also reported.—SAF

7,027,986

43.72.Kb METHOD AND DEVICE FOR PROVIDING SPEECH-TO-TEXT ENCODING AND TELEPHONY SERVICE

Charles David Caldwell *et al.*, assignors to AT&T Corporation
11 April 2006 (Class 704/235); filed 22 January 2002

The patent covers a method and an apparatus for providing automated speech-to-text encoding and decoding for hearing-impaired people. A broadband subscriber terminal interfaces with (a) a network to transmit speech packets; (b) a telephone to convey speech information; and (c) a display to present textual information representing the spoken words. A speech buffer



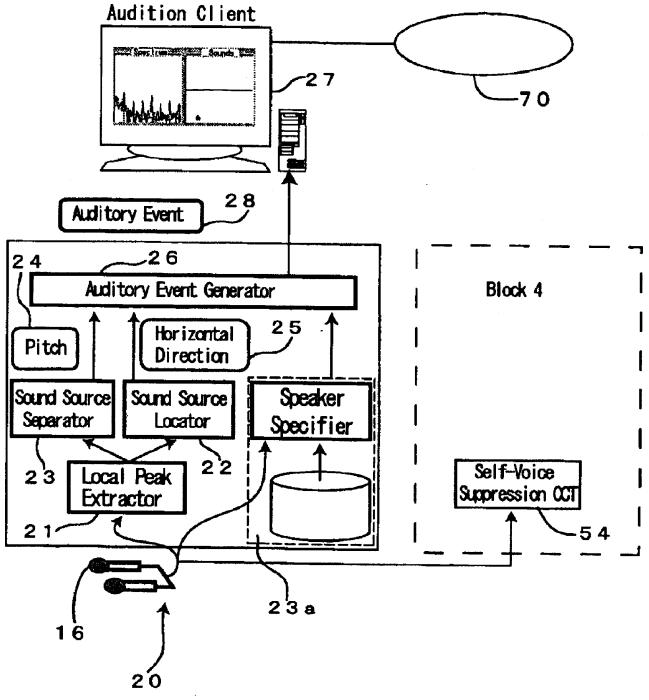
in the subscriber terminal receives speech data and a processor decodes and displays a textual representation of the speech. A database stores the voice and/or speech patterns that are utilized by a speech analyzer to recognize an incoming caller and to associate a name, or at least the gender, with the incoming call. A pitch and inflection analyzer evaluates speech to add punctuation to the displayed text.—DRR

6,967,455

43.72.Ne ROBOT AUDIOVISUAL SYSTEM

Kazuhiro Nakadai *et al.*, assignors to Japan Science and Technology Agency
22 November 2005 (Class 318/568.12); filed in Japan 9 March 2001

This robot control system includes a sound detection mechanism geared primarily to detect, identify, and locate speech sounds. When a voice source has been located, the robot can orient toward the sound. An initial response includes head-turning and, if the situation seems to call for it, the



Block 1

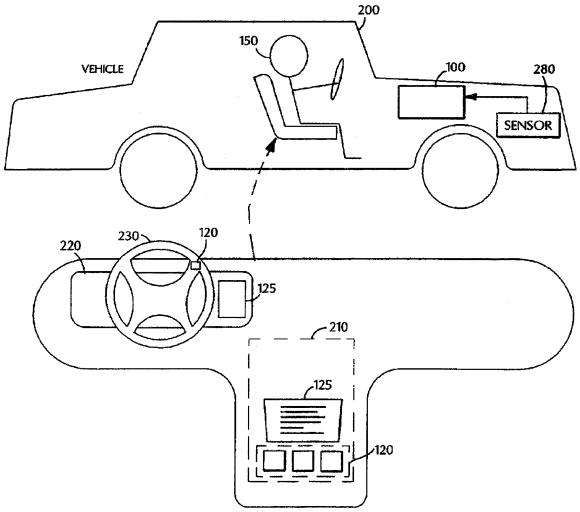
device can turn toward the voice and move in that direction. The speech perception mechanisms are described in some detail.—DLR

6,968,311

43.72.Ne USER INTERFACE FOR TELEMATICS SYSTEMS

Ronald P. Knockeart *et al.*, assignors to Siemens VDO Automotive Corporation
22 November 2005 (Class 704/270); filed 30 July 2001

Apparently, the term telematics, as used here, refers, in general, to computer control of various kinds of equipment. However, examples given are as diverse as cell phones and automobiles. In the car, the issue seems to



involve navigation more than driving. The patent describes a collection of fairly obvious techniques involving both audible and visual cues and feedback to help verify and confirm the accuracy of system commands being entered by voice recognition or by manual operations.—DLR

7,016,837

43.72.Ne VOICE RECOGNITION SYSTEM

Hiroshi Seo *et al.*, assignors to Pioneer Corporation
21 March 2006 (Class 704/233); filed in Japan 18 September 2000

The authors propose to modify the standard sort of hidden Markov model (HMM) speech recognition system by additionally modeling additive noise and multiplicative distortion in distinct ways. Additive noise is given its own HMM, which is then added (literally) to the speech HMM to yield an initial ''combination'' HMM that is then adaptively modified in response to changing noise. The modeling of multiplicative distortion is more obscure, but refers to the ''cepstrum means normalization'' method that is presumably prior art.—SAF

7,027,979

43.72.Ne METHOD AND APPARATUS FOR SPEECH RECONSTRUCTION WITHIN A DISTRIBUTED SPEECH RECOGNITION SYSTEM

Tenkasi Ramabadran, assignor to Motorola, Incorporated
11 April 2006 (Class 704/205); filed 14 January 2003

This patent builds upon a previous patent describing a distributed speech recognition system in which just 13 mel-frequency cepstral coefficients defining the speech frames are transmitted from the mobile unit to the central decoding system. The patent now describes a new method for reconstructing the other coefficients which are not transmitted. This involves determining the dependence of the missing coefficients on the transmitted ones and on the pitch period, by exploiting the information in a large database of speech signals.—SAF

7031,915

43.72.Ne ASSISTED SPEECH RECOGNITION BY DUAL SEARCH ACCELERATION TECHNIQUE

James K. Baker, assignor to Aurilab LLC
18 April 2006 (Class 704/231); filed 23 January 2003

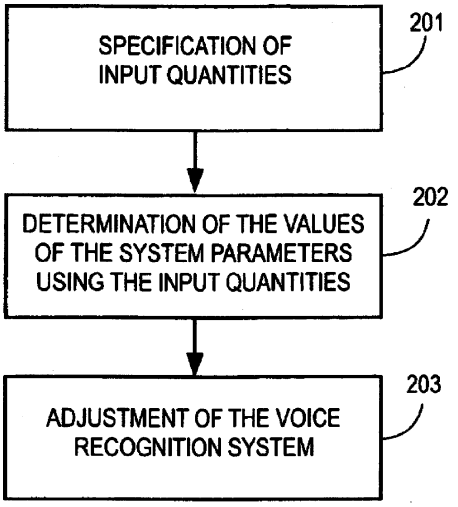
Speech recognition search involves a trade-off between considering numerous alternatives in order of likelihood and just pushing ahead with a first choice. It is considered impractical to consider alternatives in a theoretically ideal fashion in real-time search. This patent puts forth a new scheme for combining the best attributes of a priority queue search and a beam search into a ''dual search'' algorithm. The algorithm can in principle be used to combine other pairs of speech recognition search strategies.—SAF

7,035,799

43.72.Ne METHOD AND DEVICE FOR VOICE RECOGNITION

Alfred Hauenstein, assignor to Siemens Aktiengesellschaft
25 April 2006 (Class 704/246); filed in Germany 21 November 1997

The accuracy of a voice recognition system is adjusted automatically with a prescribed input quantity based on the application and computing



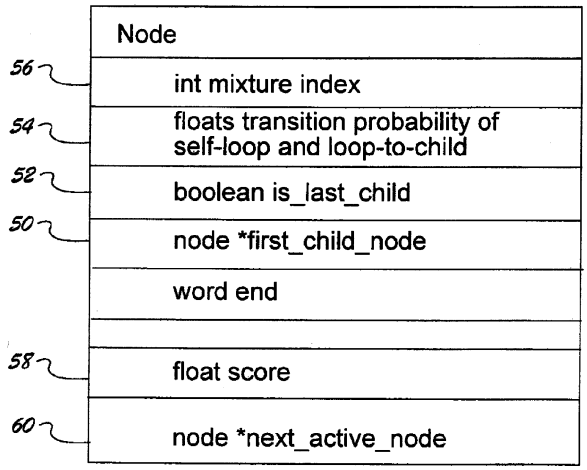
power available. One or more system parameters can be specified and weighted in conjunction with the input quantity. The system can be adapted with an adjusting element to individual requirements by trading off accuracy for speed.—DAP

7,035,802

43.72.Ne RECOGNITION SYSTEM USING LEXICAL TREES

Luca Rigazio and Patrick Nguyen, assignors to Matsushita Electric Industrial Company, Limited
25 April 2006 (Class 704/256); filed 31 July 2000

To improve on the Viterbi algorithm in isolated word or continuous speech recognition systems having limited processing power, pattern classification is performed by measuring the closeness between feature vectors extracted from the input and reference patterns. After aligning two speech patterns which may differ in duration and rate of speaking, a lexical tree structure makes a decision on what is the closest reference pattern.—DAP



7,016,833

43.72.Pf SPEAKER VERIFICATION SYSTEM USING ACOUSTIC DATA AND NON-ACOUSTIC DATA

Todd J. Gable *et al.*, assignors to The Regents of the University of California
21 March 2006 (Class 704/209); filed 12 June 2001

A truly novel scheme is proposed for incorporating the data from a ''glottal electromagnetic microsensor'' alongside acoustic speech data to verify a speaker's identity. The method is completely specified and test results are reported which show dramatic improvement of verification performance as against traditional ''all-acoustic'' methods.—SAF

7,028,914

43.75.Mn PIANO HUMIDISTAT

Robert W. Mair and E. Keith Howell, assignors to Dampp-Chaser Electronics Corporation
18 April 2006 (Class 236/44 A); filed 29 September 2003

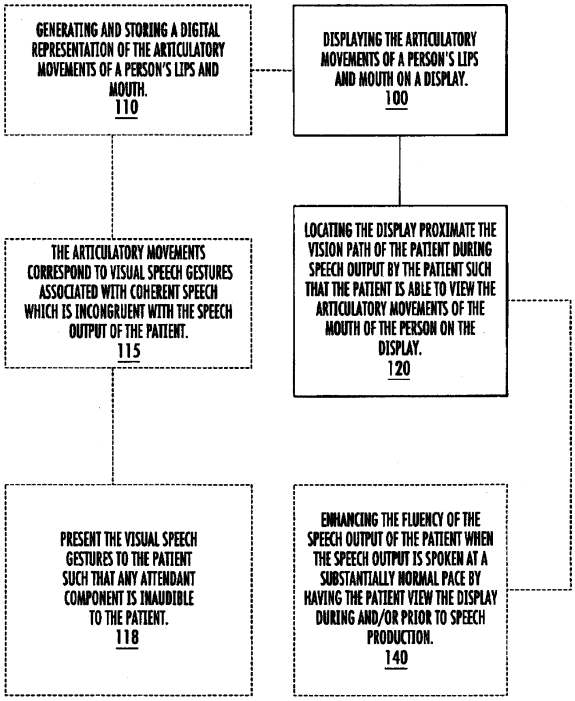
It is well known that humidity levels affect all wood instruments by swelling (and contraction) of the wood fibers. This proposal uses an automatic humidistat and heater to control the humidity levels inside the piano. Next step: the harpsichord, where temperature and humidity control is not a laughing matter.—MK

7,031,922

43.75.Rs METHODS AND DEVICES FOR ENHANCING FLUENCY IN PERSONS WHO STUTTER EMPLOYING VISUAL SPEECH GESTURES

Joseph Kalinowski *et al.*, assignors to East Carolina University
18 April 2006 (Class 704/271); filed 20 November 2000

Embodiments of this device use visual choral speech (defined here as visual speech gestures) as a fluency-enhancing stimulus relayed to a stuttering patient during and/or in advance of speech output by the patient. The



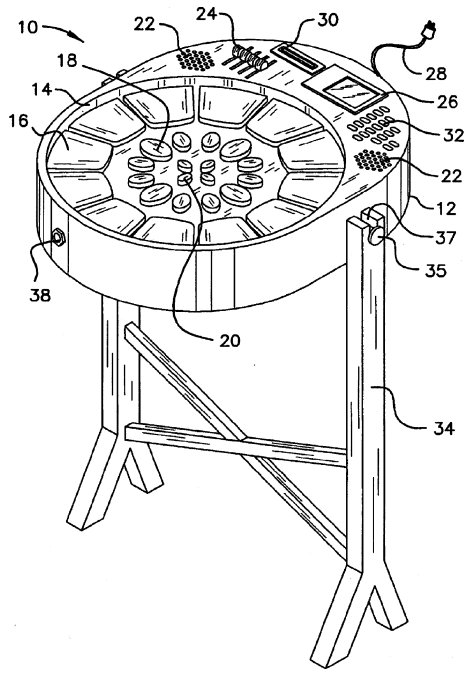
visual choral speech can be coherent or incoherent and can be incongruent with the speech content output by the patient. The visual choral speech stimulus can be used with forms of auditory feedback or other forms of treatment as well.—DRR

7,030,305

43.75.Tv ELECTRONIC SYNTHESIZED STEELPAN DRUM

Salmon Cupid, Whitby, Ontario, Canada
18 April 2006 (Class 84/411 R); filed 6 February 2004

The inventor proposes a steel pan electronic interface. The use of a ''pressure sensor'' is vaguely mentioned but all other details are conspicu-



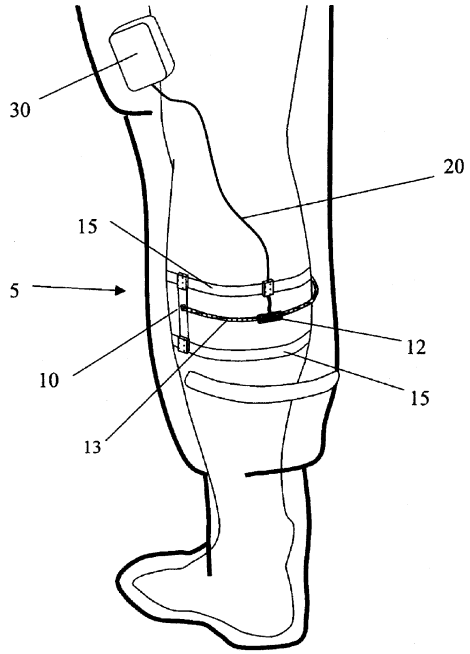
ously absent. Isn't this really just an extension of the numerous electronic drum patents?—MK

7,033,321

43.80.Ev ULTRASONIC WATER CONTENT MONITOR AND METHODS FOR MONITORING TISSUE HYDRATION

Armen P. Sarvazyan, assignor to Artann Laboratories, Incorporated
25 April 2006 (Class 600/449); filed 25 October 2004

We have here a device for determining the hydration and water content status of soft biological tissue. The method uses a pair of ultrasonic transducers **10** (one not seen in this figure) located with a known separation distance and held against the tissue of interest by an adjustable support



system. Measuring the transmission angle, the ultrasound velocity in the tissue of interest is measured. The water content is evaluated on the basis of time-of-flight results of the velocity measurement and the hydration status is thus determined. One embodiment is a wearable device attached to human calf for measuring water content in that muscle.—DRR

6,992,245

43.75.Wx SINGING VOICE SYNTHESIZING METHOD

Hideki Kenmochi *et al.*, assignors to Yamaha Corporation
31 January 2006 (Class 84/622); filed in Japan 27 February 2002

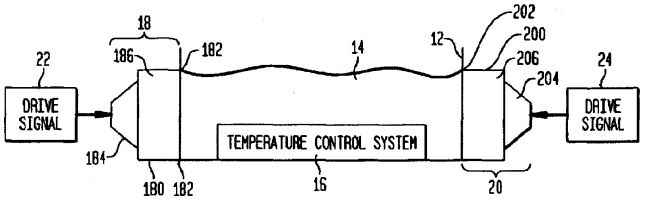
This patent uses the well-known phase vocoder to model the sound of a human singer. This is not new (as the inventors know, citing both Flanagan's United States Patent 3,360,610 and Laroche and Dolson's 1999 J. Audio Eng. Soc. paper.) After a short-time Fourier transform, the spectral peaks are identified. These peaks can be shifted up or down in the spectrum according to the desired pitch. Since the phase vocoder is a FFT-based method, it depends on both magnitude and phase, so the phase must be interpolated along with the spectrum. Both magnitude and phase interpolation are covered by Laroche and Dolson. Rodet and Depalle used the phase vocoder to model the voice of Farinelli (the last castrato) in 1994. What is marginally new here is the use of various tabular ''throb data'' (i.e., vibrato or tremolo) to modify the spectrum before applying the inverse transform.—MK

7,029,452

43.80.Qf ACOUSTICALLY-DRIVEN HYDROTHERAPY SYSTEM

Noyal John Alton, Jr., Virginia Beach, Virginia
18 April 2006 (Class 601/47); filed 20 February 2003

This sounds (no pun intended) like an acoustic version of a Jacuzzi system. The system consists of a rigid tank filled with a liquid. At least one



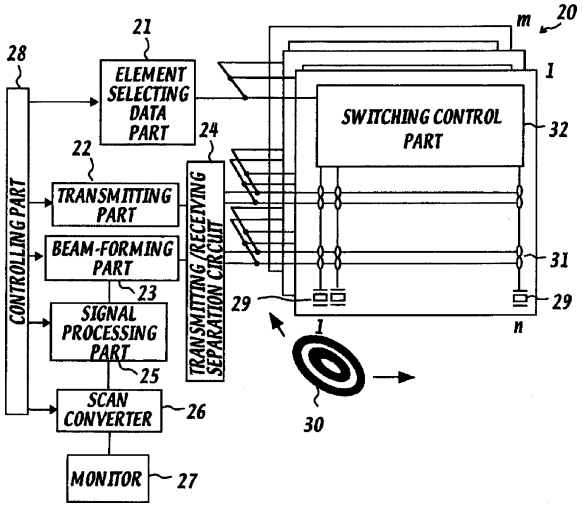
acoustic wave impinges on the tank's exterior walls, where it couples with and is transmitted through the liquid. The walls of the tank act as reflectors to enhance the hydrotherapeutic massage.—DRR

7,029,445

43.80.Qf ULTRASONIC DIAGNOSING APPARATUS

Ryuichi Shinomura *et al.*, assignors to Hitachi Medical Corporation
18 April 2006 (Class 600/443); filed in Japan 12 January 2000

This is a medical diagnostic device for acquiring an ultrasonic image by scanning the interior of the target object to be examined in real time with an ultrasonic beam formed by a two-dimensional transducer array. The system includes means for correcting a focusing error that can occur due to the difference in the propagation-path length of ultrasound transmitted or received by a concentric multiring-type transducer.—DRR

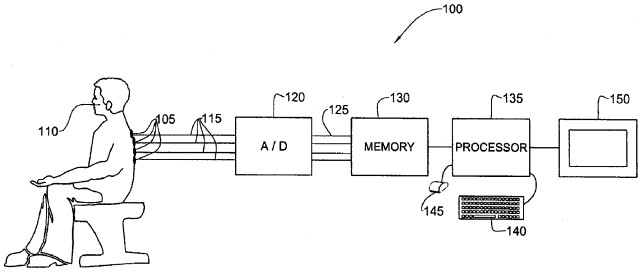


7,033,323

43.80.Qf METHOD AND SYSTEM FOR ANALYZING RESPIRATORY TRACT AIR FLOW

Meir Botbol and Igal Kushnir, assignors to Deepbreeze Limited
25 April 2006 (Class 600/538); filed 4 February 2004

The operation of this analytical device is based on the finding that the average acoustic energy over a region of an individual's back or chest during a given time interval can be correlated with the airflow portion in the respiratory tract underlying that region during that time interval. To perform the measurement, multiple microphones are fixed onto a subject's back or chest over the portion of the respiratory tract. Respiratory tract sounds are recorded from the region over a time interval, and the average acoustical energy during the interval is determined at multiple locations in the region. The average acoustical energy, summed over these locations, is then correlated with the airflow in the portion of the respiratory tract. In a preferred embodiment, an airflow value is calculated equal to the logarithm of the total acoustic energy.—DRR

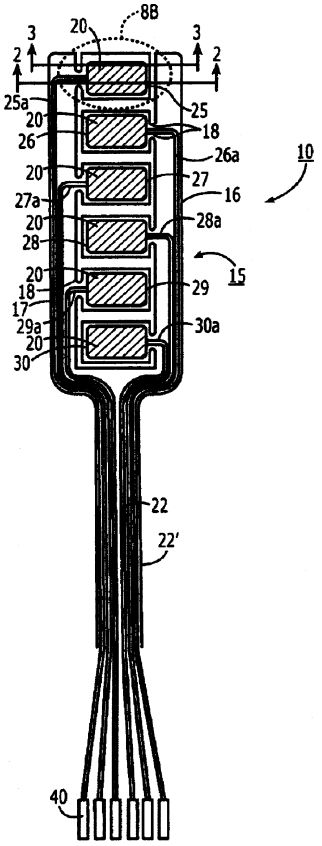


7,037,268

43.80.Qf LOW PROFILE ACOUSTIC SENSOR ARRY AND SENSORS WITH PLEATED TRANSMISSION LINES AND RELATED METHODS

Michael Z. Sleva *et al.*, assignors to MedAcoustics, Incorporated
2 May 2006 (Class 600/459); filed 29 February 2000

The patent relates to disposable, low-profile acoustic sensors for capturing sounds from within the human body. The acoustic sensors are especially designed for noninvasive digital acoustic cardiography, phonography, and acoustic spectral applications. The low-profile acoustic array is configured to selectively respond to shear waves while rejecting compression wave energy in the frequency range of interest. One sensor may be configured as a linear strip with a frame segment having at least one rail extending longitudinally and multiple sensor elements extending therefrom. These sensor elements each have a resilient core and opposing PDVF outer layers configured with opposing polarities onto the core. A detachable carrier member carries the discrete sensors to maintain the positional alignment until they are secured to a patient.—DRR

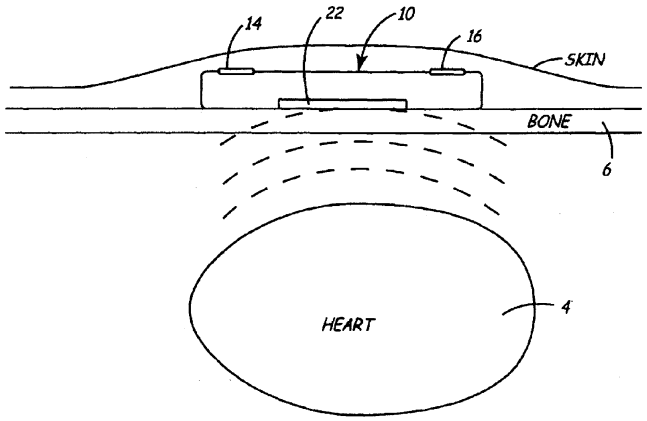


7,035,684

43.80.Qf METHOD AND APPARATUS FOR MONITORING HEART FUNCTION IN A SUBCUTANEOUSLY IMPLANTED DEVICE

Brian B. Lee, assignor to Medtronic, Incorporated
25 April 2006 (Class 600/513); filed 26 February 2003

An implantable monitor is asserted to be minimally invasive. The method for chronically monitoring a patient's hemodynamic function is



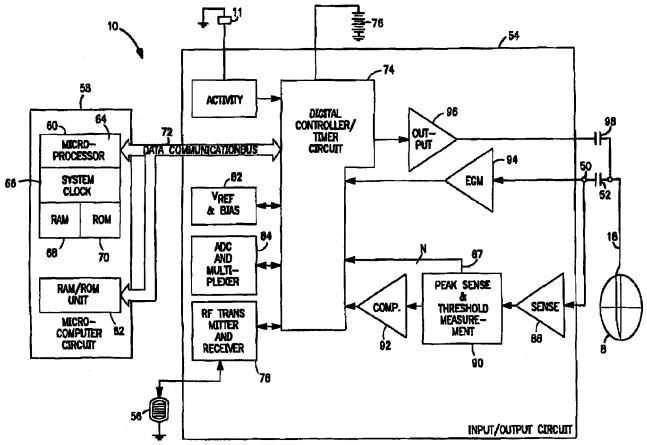
based on signals sensed by one or more acoustical sensors. The monitor may be implanted subcutaneously or submuscularly relative to the heart to allow acoustic signals generated by the heart or blood motion to be received by a passive or active acoustic sensor. Circuitry for filtering and amplifying acoustical data is provided and sampled data may be continuously or intermittently written to a looping memory buffer. ECG electrodes and associated circuitry may also be included to simultaneously record ECG data. Upon a manual or automatic event trigger, acoustical and ECG data may be stored in long-term memory for future uploading to an external device. The external device may present acoustical data visually and acoustically along with associated ECG data to allow interpretation of both electrical and mechanical aspects of heart function.—DRR

7,037,266

43.80.Qf ULTRASOUND METHODS AND IMPLANTABLE MEDICAL DEVICES USING SAME

Bozider Ferek-Petric and Branko Breyer, assignors to Medtronic, Incorporated
2 May 2006 (Class 600/453); filed 25 April 2002

The implantable system described in this patent uses ultrasound and other techniques for measuring blood flow velocities at several sampling rates. In order to minimize energy required for ultrasonic sampling, pulsed Doppler signal packages provided by a pulsed ultrasound circuit are switched at the lowest possible repetition rate such as to be able to record the blood flow velocity within the heart. For example, the ultrasound circuitry may be activated only within the part of the cardiac cycle designated as the Doppler measurement interval (DMI). The ultrasound may be switched off and on during the DMI and/or the ultrasound circuit may be switched on and off at different sampling modes.—DRR

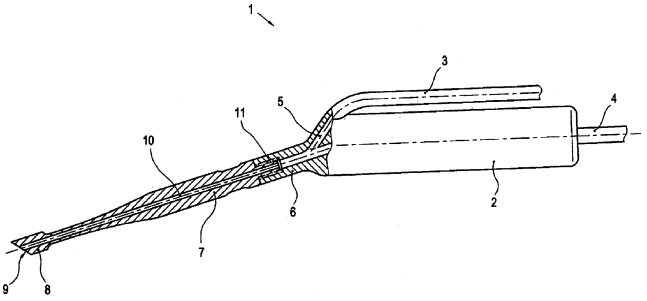


7,025,735

43.80.Sh ULTRASONIC APPARATUS FOR THE TREATMENT OF SEPTIC WOUNDS

Holger Soring and Jorg Soring, assignors to Soring GmbH Medizintechnik
11 April 2006 (Class 601/2); filed in Germany 20 November 2000

This ultrasonic apparatus, designed to treat septic wounds, is essentially a handpiece **1** containing a handle region **2** and a connecting tube **3** to attach to a supply tube providing rock salt solution and medical treatment



agents such as heparin or antibiotics that are fed through the feed channel **5** to the sonotrode **7** or ultrasonic treatment head. The ultrasound energy is fed in a conventional manner to the handpiece through a line **4** connecting to an ultrasonic generator.—DRR

7,025,724

43.80.Vj WAVELET DEPULSING OF ULTRASOUND ECHO SEQUENCES

Dan Adam and Oleg Michailovich, assignors to Technion Research and Development Foundation Limited
11 April 2006 (Class 600/437); filed 27 April 2001

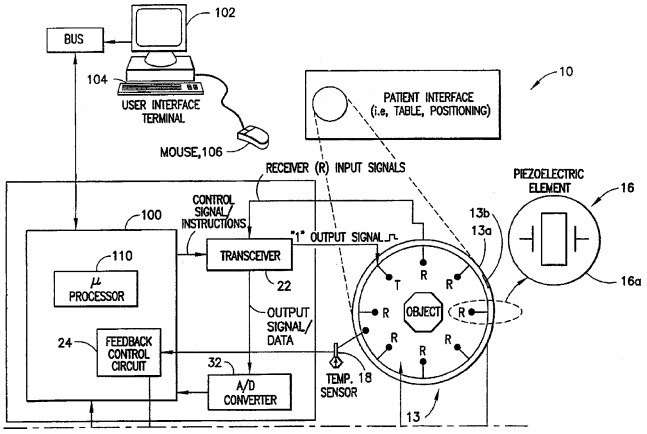
A log spectrum (known as the cepstrum) of an echo sequence is computed and a point-spread function is estimated using one of two methods. In one method, a low-resolution wavelet projection of the log spectrum is used to estimate the point-spread function spectrum. In the other method, the log spectrum is thresholded to remove outlying components. From either method, the frequency-domain phase of the point-spread function is obtained and the point-spread function is deconvolved from the echo sequence.—RCW

7,025,725

43.80.Vj THREE-DIMENSIONAL ULTRASOUND COMPUTED TOMOGRAPHY IMAGING SYSTEM

Donald P. Dione *et al.*, assignors to Ultrasound Detection Systems, LLC
11 April 2006 (Class 600/443); filed 13 September 2004

Piezoelectric elements arranged in rings are used to transmit a cone-shaped beam and also to receive ultrasound echos with a cone-shaped beam.



The echoes are processed to construct a three-dimensional image of the scattering object. The resulting image is displayed for analysis.—RCW

7,033,320

43.80.Vj EXTENDED VOLUME ULTRASOUND DATA ACQUISITION

Patrick L. Von Behren and Jian-Feng Chen, assignors to Siemens Medical Solutions USA, Incorporated
25 April 2006 (Class 600/443); filed 5 August 2003

Multiple overlapping volumes are registered and combined to form an extended volume. The extended volume is larger than the volume that a multidimensional or an oscillating array is capable of imaging without repositioning the probe. The registration is accomplished by sensing the transducer position associated with each of the different volumes. Overlapping parts of the volumes are compounded and the extended volume is displayed.—RCW

7,037,263

43.80.Vj COMPUTING SPATIAL DERIVATIVES FOR MEDICAL DIAGNOSTIC IMAGING METHODS AND SYSTEMS

Thilaka Sumanaweera *et al.*, assignors to Siemens Medical Solutions USA, Incorporated
2 May 2006 (Class 600/443); filed 20 August 2003

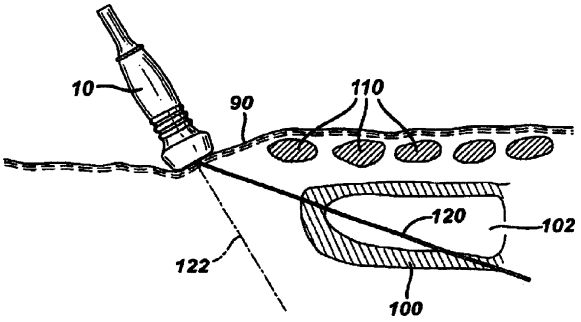
Spatial gradients are computed in various ways for use in volume rendering three-dimensional images with shading or for use in other image processing. The gradients are found in the coordinates of the scanned beam or in the coordinates of the display. The methods permit use of the gradient information in the coordinate system of the scan without prior scan conversion.—RCW

7,037,264

43.80.Vj ULTRASONIC DIAGNOSTIC IMAGING WITH STEERED IMAGE PLANE

McKee Dunn Poland, assignor to Koninklijke Philips Electronics N.V.
2 May 2006 (Class 600/447); filed 30 May 2003

The beam from a two-dimensional array transducer is steered through an acoustic window by adjusting parameters that allow the image plane to be tilted in the elevation dimension, moved laterally, or rotated about the axis of the transducer to position the image plane so that a particular anatomic region is effectively imaged for diagnosis.—RCW



7,037,265

43.80.Vj METHOD AND APPARATUS FOR TISSUE HARMONIC IMAGING WITH NATURAL (TISSUE) DECODED CODED EXCITATION

Xiaohui Hao *et al.*, assignors to GE Medical Systems Global Technology Company, LLC
2 May 2006 (Class 600/447); filed 6 October 2003

Coded pulses are transmitted with a time-bandwidth greater than one and with a phase-inverted version of the coded pulses. Backscattered echoes are received and filtered. Echo decoding is implemented through the propagation of the specially designed wideband waveforms used in the method and avoids costly decoding filters.—RCW

7,037,269

43.80.Vj ULTRASONIC IMAGINING CATHETERS

Elvin Nix *et al.*, assignors to Volcano Corporation
2 May 2006 (Class 600/459); filed in the United Kingdom 20 July 2000

An ultrasound transducer array is mounted on the distal end of a catheter using a special flexible coupling cable that is helically wound around the catheter to enhance flexibility.—RCW

7,037,271

43.80.Vj MEDICAL IMAGING DEVICE

Robert J. Crowley, assignor to Boston Scientific Corporation
2 May 2006 (Class 600/463); filed 27 May 2003

A disposable sheath contains a liquid and supports an elongated flexible drive shaft carrying an ultrasound transducer and a trocar adapted to the sheath. The trocar, which permits the device to be inserted where no natural passage exists, is also adapted to maintain registration of the imaging device while the imaging device rotates. An inflatable balloon may be included.—RCW

LETTERS TO THE EDITOR

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

Comment on "Broadband matched-field processing: Coherent and incoherent approaches" [Journal of the Acoustical Society of America 113, 2587–2598 (2003)] (L)

Saralees Nadarajah[a)]
University of Manchester, Manchester M13 9PL, UK

Samuel Kotz
George Washington University, Washington, D.C. 20052

(Received 5 May 2006; revised 29 June 2006; accepted 13 July 2006)

The recent paper by Soares and Jesus [J. Acoust. Soc. Am. **113**, 2587–2598 (2003)] discussed the probability density functions (pdfs) of the envelope and phase of signals for modeling in broadband match-field processing. In this Comment, we provide an explicit expression for the pdf of phase involving the standard normal distribution function. Several elementary approximations are also suggested and references given.
 [DOI: 10.1121/1.2335274]

PACS number(s): 43.60.Cg [JJM] Pages: 1777–1778

The recent paper by Soares and Jesus (2003) considered modeling of observed signals in broadband match-field processing. Appendix A of the paper derived expressions for the probability density functions (pdfs) of the envelope (R) and phase (Φ) of the signals. The pdf of the phase (Φ) is given as [see Eq. (A14) in the paper]

$$p_\Phi(\phi) = \frac{1}{2\pi\sigma^2} \times \exp\left(-\frac{m_a^2+m_b^2}{2\sigma^2}\right) + \frac{m_a\cos\phi + m_b\sin\phi}{2\pi\sigma} \times \exp\left\{-\frac{(m_a\sin\phi - m_b\cos\phi)^2}{2\sigma^2}\right\} \times \int_{-\infty}^{(m_a\cos\phi+m_b\sin\phi)/\sigma} \exp\left(-\frac{\lambda^2}{2}\right)d\lambda \quad (1)$$

for some parameters m_a, m_b, and σ. Following this expression, the paper states "It is not possible to continue any further knowing the difficulties to calculate the integral in the second term. Available approximate expressions exist for large $(m_a\cos\phi+m_b\sin\phi)/\sigma$ but …."

We would like to point out that Eq. (1) can be easily computed and that elementary approximations applicable for wide range of values of $(m_a\cos\phi+m_b\sin\phi)/\sigma$ are available. First, note that Eq. (1) can be reexpressed in terms of the cumulative distribution function (2) of the standard normal distribution. If $\Psi(\cdot)$ denotes standard normal cdf

$$\Psi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \exp\left(-\frac{z^2}{2}\right)dz \quad (2)$$

then Eq. (1) can be reexpressed as

$$p_\Phi(\phi) = \frac{1}{2\pi\sigma^2} \times \exp\left(-\frac{m_a^2+m_b^2}{2\sigma^2}\right) + \frac{m_a\cos\phi + m_b\sin\phi}{\sqrt{2\pi}\sigma} \times \exp\left\{-\frac{(m_a\sin\phi - m_b\cos\phi)^2}{2\sigma^2}\right\} \times \Psi\left(\frac{m_a\cos\phi + m_b\sin\phi}{\sigma}\right). \quad (3)$$

The expression above is elementary except for the standard normal cdf in Eq. (3). Numerical routines for the computation of $\Psi(\cdot)$ are available in almost every statistical package, e.g., GenStat, Minitab, R, SAS, S-Plus, SPSS. Even most hand calculators have functions to compute $\Psi(\cdot)$.

As far as approximations for $\Psi(\cdot)$ are concerned, there are many from which one can choose. Some of them are

a)Electronic mail: saralees.nadarajah@manchester.ac.uk

$$\Psi(x) \approx \frac{1}{2}\left[1 + \left\{1 - \exp\left(-\frac{2x^2}{\pi}\right)\right\}^{1/2}\right]$$

for $-\infty < x < \infty$ [due to Polya (1945)] with $\epsilon = 0.003$ (where ϵ denotes the absolute error of the approximation);

$$\Psi(x) \approx \frac{1}{2}[1 + \{1 + (\alpha - \beta x)^c\}^{-k} - \{1 + (\alpha + \beta x)^c\}^{-k}]$$

for $-\infty < x < \infty$ [due to Burr (1967)] with $\alpha = 0.644\,693$, $\beta = 0.161\,984$, $c = 4.874$, $k = -6.158$, and $\epsilon = 0.000\,46$;

$$\Psi(x) \approx \frac{\exp(2y)}{1 + \exp(2y)}$$

for $-\infty < x < \infty$ [due to Page (1977)] with $y = a_1 x(1 + a_2 x^2)$, $a_1 = 0.7988$, $a_2 = 0.044\,17$, and $\epsilon = 0.00014$;

$$\Psi(x) \approx \frac{1}{2}\left[1 + \left\{1 - \exp\left(-\frac{2x^2}{\pi} - \frac{2(\pi - 3)x^4}{3\pi^2}\right)\right\}^{1/2}\right]$$

for $0 < x < 3.5$ [due to Cadwell (1951)] with $\epsilon = 0.0007$;

$$\Psi(x) \approx 1 - (a_1 t + a_2 t^2 + a_3 t^3)\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

for $x > 0$ [due to Zelen and Severo (1964)] with $t = 1/(1 + 0.332\,67x)$, $a_1 = 0.436\,183\,6$, $a_2 = -0.120\,167\,6$, $a_3 = 0.937\,298$, and $\epsilon = 0.000\,01$;

$$\Psi(x) \approx 1 - \frac{x(x^6 + 6x^4 + 14x^2 - 28)}{\sqrt{2\pi}(x^2 + 2)(x^6 + 5x^4 - 20x^2 - 4)} \exp\left(-\frac{x^2}{2}\right)$$

for $5 \leq x \leq 10$ [due to Schucany and Gray (1968)] with $\epsilon = 0.000\,005$; and,

$$\Psi(x) \approx 1 - \frac{1}{2}(a_1 + a_2 x + \cdots + a_6 x^5)^{-16}$$

for $x > 0$ [due to Carta (1975)] with $a_1 = 0.999\,999\,858\,2$, $a_2 = 0.048\,738\,579\,5$, $a_3 = 0.021\,098\,110\,45$, $a_4 = 0.003\,372\,948\,927$, $a_5 = -0.000\,051\,728\,977\,42$, $a_6 = 0.000\,085\,695\,794\,2$, and $\epsilon = 0.000\,001\,2$. For a comprehensive account of the known approximations for Eq. (2), the reader is referred to Chap. 13 of Johnson *et al.* (1994) and references therein.

The purpose of this correspondence is not to criticize the paper by Soares and Jesus (2003). The overall contribution of the paper is highly commendable. However, we feel the references as well as the results mentioned above can help the readers and authors of this journal in making better choices with regard to similar problems.

Burr, I. W. (**1967**). "A useful approximation to the normal distribution function with application to simulation," Technometrics **9**, 647–651.

Cadwell, J. H. (**1951**). "The bivariate normal integral," Biometrika **38**, 475–479.

Carta, D. G. (**1975**). "Low-order approximations for the normal probability integral and the error function," Math. Comput. **29**, 856–862.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (**1994**). *Continuous Univariate Distributions*, 2nd ed. Vol. **1** (Wiley, New York).

Page, E. (**1977**). "Approximations to the cumulative normal function and its inverse for use on a pocket calculator," Appl. Stat. **26**, 75–76.

Polya, G. (**1945**). "Remarks on computing the probability integral in one and two dimensions," *Proceedings of the 1st Berkeley Symposium on Mathematical Statistics*, pp. 63–78.

Schucany, W. R., and Gray, H. L. (**1968**). "A new approximation related to the error function," Math. Comput. **32**, 1232–1240.

Soares, C., and Jesus, S. M. (**2003**). "Broadband matched-field processing: Coherent and incoherent approaches," J. Acoust. Soc. Am. **113**, 2587–2598.

Zelen, M., and Severo, N. C. (**1964**). "Probability functions," *Handbook of Mathematical Functions*, edited by M. Abramowitz, and I. A. Stegun (U.S. Department of Commerce, Washington, D.C.), pp. 925–995.

A prototype acoustic gas sensor based on attenuation (L)

Andi Petculescu[a)]
Department of Mechanical Engineering, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208

Brian Hall, Robert Fraenzle, and Scott Phillips
Commercial Electronics, Broken Arrow, Oklahoma 74012

Richard M. Lueptow
Department of Mechanical Engineering, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208

(Received 27 March 2006; revised 21 July 2006; accepted 21 July 2006)

Acoustic attenuation provides the potential to identify and quantify gases in a mixture. We present results for a prototype attenuation gas sensor for binary gas mixtures. Tests are performed in a pressurized test cell between 0.2 and 32 atm to accommodate the main molecular relaxation processes. Attenuation measurements using the 215-kHz sensor and a multiseparation, multifrequency research system both generally match theoretical predictions for mixtures of CO_2 and CH_4 with 2% air. As the pressure in the test cell increases, the standard deviation of sensor measurements typically decreases as a result of the larger gas acoustic impedance. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2336758]

PACS number(s): 43.35.Fj, 43.35.Yb [RR] Pages: 1779–1782

I. INTRODUCTION

The propagation of acoustic waves in gases is intimately linked to gas composition, temperature, and pressure. This may enable the use of acoustic waves as a quantitative probe of the thermodynamic and molecular properties of gas mixtures. The classical sound attenuation arises by transport phenomena (e.g., heat conduction, viscosity, and diffusion) while the nonclassical absorption is due to thermal relaxation involving the molecules' vibrational and rotational levels. In a gaseous medium, the molecules exchange energy via collisions. In the absence of any external perturbations, the internal degrees of freedom of the gas molecules (e.g., vibration and rotation) are rapidly thermalized. As a result, the gas is in thermal equilibrium at a constant temperature. When an acoustic wave is launched in the gas, this equilibrium is broken: the internal degrees of freedom (DOF) activated by collisions now have to relax to the acoustic temperature. Generally speaking, the wave energy expended to activate the molecular internal degrees of freedom is not promptly returned to the acoustic wave. This gives rise to a net absorption of acoustic energy—the nonclassical, or molecular absorption of sound. The sound absorption is largest when the acoustic period is commensurate with the molecular relaxation time. It is this feature that makes acoustic studies of molecular relaxation possible.

Over the years, many carefully thought-out experiments have been designed for the accurate measurement of attenuation and sound speed, only a few of which are mentioned here.[1–10] In his book, Lambert[11] describes various acoustic experimental techniques to measure molecular relaxation. The experiments, coupled with a strong modeling effort, have yielded insights into the physical mechanisms responsible for molecular relaxation in various gas mixtures.

Typical acoustic gas monitors operate by tracking changes in the adiabatic sound speed c that are induced largely by changes in the molecular weight according to the ideal-gas formula $c=\sqrt{\gamma RT/M}$, where $\gamma=C_P/C_V$ is the isobaric-to-isochoric specific heat ratio, $R=8.3144$ J/mol/K is the universal gas constant, T is the ambient temperature, and M is the molecular weight of the gas. For some applications requiring a quantitative analysis of the monitored environment, the speed of sound alone cannot provide sufficient information. That is where measurements of the acoustic attenuation come into play. As shown in our previous work,[12] precise measurements of sound speed *and* attenuation can lead to (quasi) quantitative gas analysis via the identification of the molecular symmetry properties, if not of the molecules themselves. In the future, this could lead to the development of rugged "smart" acoustic sensors capable of quantitatively determining gas composition in various environments and processes.

In this paper we report on the development of a prototype dual-path acoustic sensor assembly that measures the acoustic attenuation in gas mixtures. From an experimental standpoint, the main challenge is extracting the acoustic attenuation from only two fixed-length sensing paths, as opposed to multiple-separation measurements where the separation distance between the emitting and receiving transducers is increased incrementally.[9] To compare the two cases, measurements are performed inside a pressurized test cell for different gas mixtures. Since relaxation times vary inversely with ambient pressure,[13] decreasing the pressure p is equivalent to increasing the frequency f and vice versa. By changing the pressure in the test cell incrementally over al-

[a)]Currently at University of Louisiana at Lafayette, Department of Physics; electronic mail: andi@louisiana.edu

most two orders of magnitude, we attained an f/p range that is wide enough to cover the main molecular relaxation processes of the gas mixtures.

II. THEORETICAL MODEL

In the theoretical model,[14–16] we fit a shifted-exponential function to the Lennard-Jones interaction potential[17] to obtain the transition probabilities between vibrational levels. For gas mixtures containing at least one polyatomic (thus relaxing) component, this is done for *each* energy exchange process,[14] whether interspecies or intraspecies. The fact that there exists a time delay or phase shift between the acoustic fluctuations and the exchange of energy from the internal to the external DOF makes the specific heat per mole complex-valued and frequency- dependent, i.e., it becomes an *effective specific heat*, $C_V^{\text{eff}}(f)$. The subscript V denotes isochoric processes. Using the equations of linear acoustics coupled with the equations for molecular relaxation, we have shown that the propagation of sound in excitable gas mixtures is ultimately characterized by the *effective wave number*,[14] $\tilde{k}(f)=2\pi f\sqrt{\rho_0 p_0^{-1} C_V^{\text{eff}}(f)[C_V^{\text{eff}}(f)+R]^{-1}}$, where ρ_0 and p_0 are the equilibrium density and pressure of the gas. The real and imaginary parts of the effective wave number $\tilde{k}$ determine the phase velocity c and attenuation $\alpha=\alpha_{\text{class}}+\alpha_m$ of the acoustic wave,via

$$\tilde{k}(f)=\frac{2\pi f}{c(f)}-i\alpha(f),$$

where the classical attenuation, α_{class}, results from viscous, thermal, and diffusional losses, and the molecular attenuation, α_m, results from relaxation processes. A good way to emphasize relaxational effects on the acoustic attenuation is by representing the *normalized attenuation* $\alpha\lambda$, where λ is the acoustic wavelength, against the frequency-to-pressure ratio. The presence of relaxing gases is manifested by relaxational peaks in the normalized attenuation.

It is important to note that for certain gas mixtures and for certain ambient conditions of temperature and pressure, additional mechanisms such as chemical reactions[18] and thermal radiation[19] may become important and thus may need to be accounted for in the modeling. We estimated theoretically that the effect of infrared radiation on the relaxation times and sound absorption is too small to be of any importance for the gas compositions and ambient conditions encountered in these experiments. Also, if present in the sensing volume, water vapor could influence the molecular sound absorption noticeably.[20,21] Nevertheless, dry gases were used in the experiments presented here to avoid having any water vapor present.

III. APPARATUS

In order to (i) check the validity of the model and (ii) establish experimental attenuation "references" for various gas mixtures, we use a pressurized test cell connected to a precision gas mixing system. For reference measurements in the test cell, four transmitter-receiver (T/R) pairs of narrowband piezoelectric transducers operating at 92, 149, 215, and 1000 kHz are used. The transducers are mounted inside the pressure chamber on a linear translation stage enabling 75 different T/R separations measured precisely to within 0.03 mm. For each pair, one transducer is fixed while the other is incrementally displaced, thus changing the transmitter-receiver separation length. The operating frequencies are chosen such that, together with the attainable pressure range (roughly 0.2–32 atm), they provide a frequency-to-pressure ratio span that covers the primary relaxation processes for the gases of interest (methane, carbon dioxide, ethylene, nitrogen, air). A detailed description of the apparatus is given in Ref. 9.

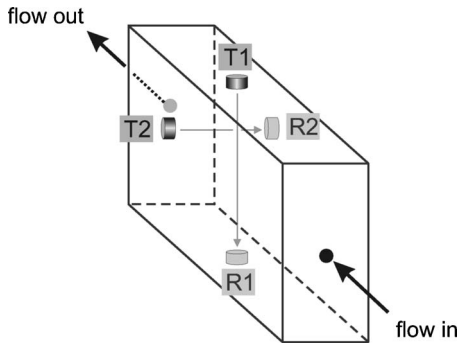


FIG. 1. Schematic diagram of the sensing geometry. The two sensing paths (T1/R1 and T2/R2) lie in planes 2 cm apart. Flow is perpendicular to both sensing paths.

The sensor assembly operates at a fixed frequency of 215 kHz. A diagram of the sensing geometry of the prototype acoustic sensor is shown in Fig. 1. The hardware consists of four ultrasonic piezoelectric transducers arranged on two sensing paths inside a small rectangular chamber (approximately 70 mm × 43 mm × 69 mm) that allows the monitored gas to flow through the sensing area. The two sensing paths, T1/R1 and T2/R2 with separation lengths of 30 and 60 mm, respectively, are staggered by 2 cm in the flow direction. Two identical end plates provide the gas inlet and outlet ports. A metal screen between the inlet and the sensing area diffuses the flow. The transmitters (T1, T2) are energized with square-wave pulse-trains of either five or ten cycles. The ultrasonic bursts travel through the gas medium to the receiving transducers (R1, R2) where they are converted to electrical signals sent to the processing electronics.

IV. DATA ANALYSIS AND RESULTS

In order to remove any dc offsets and spurious high-frequency components, a bandpass filter is applied to the signals received on the two paths. To correct for diffraction effects due to the finite size of the transmitter, the amplitude A_{rec} of each received signal is multiplied by a correction factor based on the far-field diffraction from a circular source:[9,22]

$$A(x)=A_{\text{rec}}(x)\left\{\sin\left[\frac{\pi f}{c}(\sqrt{x^2+R_E^2}-x)\right]\right\}^{-1}, \quad (1)$$

where x is the propagation distance and R_E is the emitter radius.

For the multiple-separation measurements in the test cell, 75 transmitter-receiver separations can be achieved us-

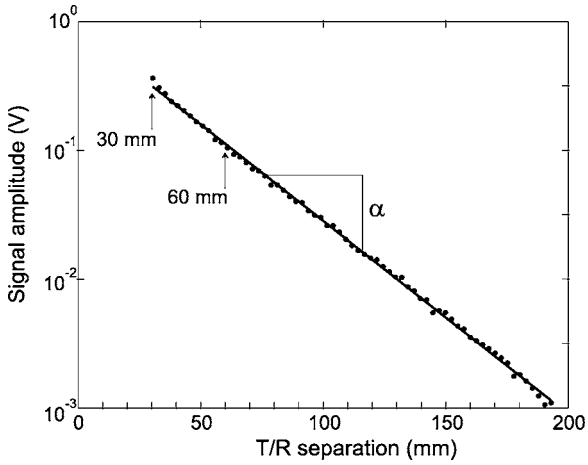


FIG. 2. Amplitude of the received signal as a function of transmitter-receiver (T/R) separation for a typical multiple-separation experiment inside the test cell. The curve is a linear fit to the 75 amplitude values. The arrows indicate the two sensing paths implemented in the sensor assembly.

ing a linear translation stage. However, the sensor prototype relies on only two sensing paths, which makes the measurements more challenging. Figure 2 shows the signal amplitude (in volts) as a function of transmitter-receiver separation for a typical multiple-separation experiment, along with the values for the two sensing paths implemented in the sensor (arrows). A linear fit is sought for the logarithmic amplitude versus separation length, whose slope yields the attenuation coefficient α.[9] In the test cell, the linear fit is based on many data points (typically 75 T/R separations), whereas the fit is based on only two data points for the prototype sensor. It is useful to note that the sound speed c enters the normalized attenuation (i) via the diffraction correction (1) and (ii) via the wavelength $\lambda = cf^{-1}$. In the analysis of the sensor data, we use the theoretical frequency-dependent phase velocity obtained using the above-described model.

Figures 3 and 4 show multiple-separation, multifrequency measurements of the normalized attenuation in the test cell, along with model predictions for N_2-C_2H_4 mixtures and for air-CH_4 and air-CO_2 mixtures, respectively. There is a good agreement between the experimental data and the

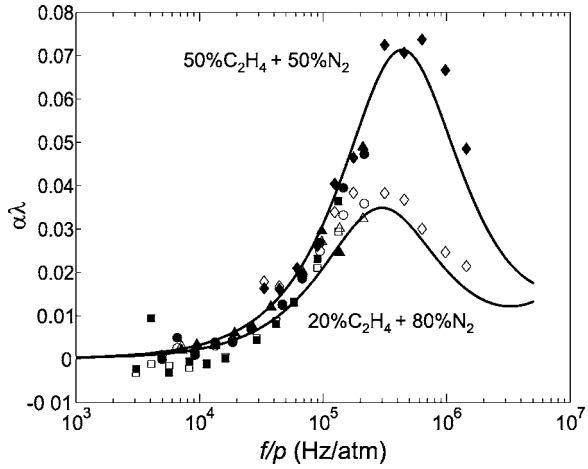


FIG. 3. Multiple-separation, multifrequency, test-cell measurements of the normalized attenuation for mixtures of 50% and 20% ethylene in nitrogen (symbols: □, 92 kHz; ○, 149 kHz; △, 215 kHz; ◇, 1 MHz; filled for 50% ethylene). Curves represent the model predictions.

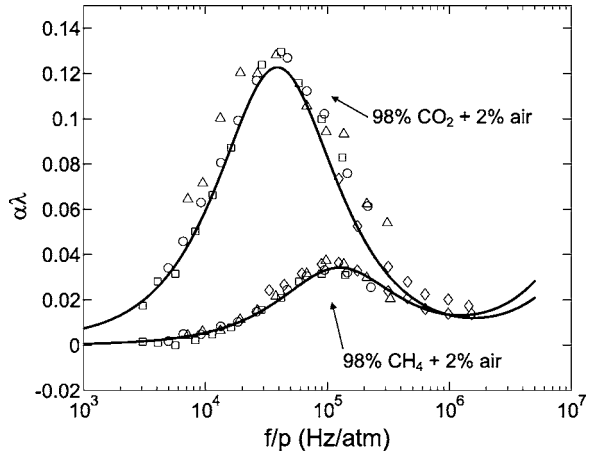


FIG. 4. Multiple-separation, four-frequency, test-cell measurements of the normalized attenuation for mixtures of CO_2 and CH_4 with air (symbols: □, 92 kHz; ○, 149 kHz; △, 215 kHz; ◇, 1 MHz). Curves represent the model predictions.

theory, especially in the region of vibrational relaxation as indicated by the normalized attenuation peaks. The vibrational modes considered in the model at room temperature are, in units of cm^{-1}:[23]

C_2H_4:	826 (rock), 943 (wag), 949 (wag), 1023 (twist), 1236 (rock), 1342 (scissor), 1444 (scissor),
CO_2:	667 (bend), 1337 (s-stretch), 2349 (a-stretch),
CH_4:	1307, 1535,
N_2:	2331.

In Fig. 3, as the ethylene concentration changes from 20% to 50%, the effective relaxation frequency shifts from 240 to 350 kHz/atm, corresponding to a decrease in the effective relaxation time from 0.66 to 0.45 μs at 1 atm, since $\tau = (2\pi f)^{-1}$.[13] This is attributed to the higher relaxational component of the 50–50 mixture, due largely to the ethylene molecules. The key point here is that the theoretical model predicts fairly well the measured attenuation, an important result given the large number of significant vibrational modes of ethylene. Prior to this, the model was only applied to gases with one to three significant vibrational modes.[9] The molecular absorption of sound by the air-CO_2 mixture in Fig. 4 is visibly larger than that of the air-CH_4 mixture, owing largely to the low-lying bending mode vibrational level of CO_2.

Comparisons of the sensor dual-path data against the multiple-separation measurements are obtained by placing the sensor assembly inside the test cell and varying the pressure over the same range as in the multiple-separation measurements (0.2–32 atm). This allows precise control of the gas composition and permits testing the sensor over a wide range of f/p (including the main relaxation processes) even though the sensor operates at a single frequency. As shown in Fig. 5, the sensor is able to capture the major relaxational features for both gas mixtures. The sensor measurements match the theory quite well for the CH_4-air mixture demonstrating the ability of the dual-path arrangement to measure the attenuation nearly as well as the multiple-separation

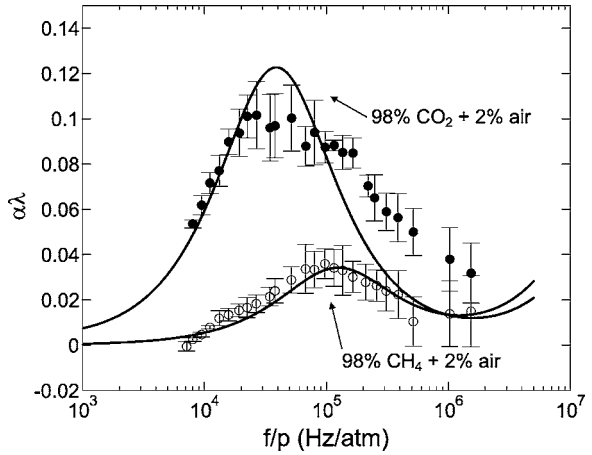


FIG. 5. Measurements with the dual-path sensor assembly at 215 kHz taken inside the test-cell over a pressure range of 0.2–32 atm (error bars represent the standard deviation over four to five tests). Curves represent the model predictions.

"laboratory" setup. The CO_2-air data are consistent with the model only in the high-pressure (small f/p) range. The departure from model predictions for the CO_2-air data at high values of f/p may stem from a number of causes. First, low-pressure measurements are challenging especially if CO_2 is present due to the high attenuation coefficient of CO_2 coupled with the large gas-transducer acoustic impedance mismatch, affecting both acoustic generation and detection. Second, any foreign molecules present in the test-cell (e.g., from small oil leaks from the vacuum pump) may induce multiple molecular relaxation pathways leading to possible broadening of the $\alpha\lambda$ curve. These issues are probably also responsible for the generally larger standard deviations at low pressures. Ambient temperature fluctuations can also affect measurements primarily via their effect on the sound speed in the medium. However, the ambient temperature varied by less than 1% during the tests so the errors introduced by temperature resulted in errors smaller than symbol sizes in Fig. 5.

V. CONCLUSIONS

We show here the basic concept for a simple dual-path acoustic gas sensor for measuring acoustic attenuation in gas mixtures. For a clear picture, we compare attenuation measurements between the dual-path, single-frequency sensor and a multiseparation, multifrequency research system for the same gas mixtures. The sensor works well, but there are limitations inherent to using only two sensing paths to obtain the attenuation coefficient, especially in the presence of a lossy gas-like CO_2. To obtain reliable quantitative real-time sensing devices, efforts must be made in the future on several fronts. In the processing electronics, filtering must be implemented in the electronics module instead of the post-processing filtering use in this analysis. Additionally, sound speed must also be measured accurately for quantitative acoustic gas sensing. For accurate time-of-flight measurements, tight thermal stability must be enforced in the sensing volume.[12] Future "smart" sensors will need to be "taught" how to interpret the acquired data based on preprogrammed physics concepts. Using sound speed and attenuation data streams, the sensor could ultimately be able to make inferences about the nature of the sensed gases based on a refined theoretical model. This constitutes a complex task, given that this is an inverse problem: it is much easier to determine sound speed and attenuation from the gas composition than vice versa. Nevertheless, combining sound speed and attenuation measurement in a simple sensor offers the potential to identify and quantify the gases making up a multicomponent mixture.

ACKNOWLEDGMENT

This work was funded by a grant from the National Aeronautics and Space Administration.

[1]F. D. Shields, "Thermal relaxation in carbon dioxide as a function of temperature," J. Acoust. Soc. Am. **29**, 450–454 (1957).
[2]T. G. Winter and G. L. Hill, "High-temperature ultrasonic measurements of rotational relaxation in hydrogen, deuterium, nitrogen, and oxygen," J. Acoust. Soc. Am. **42**, 848–858 (1967).
[3]J. C. F. Wang and G. S. Springer, "Vibrational relaxation times in some hydrocarbons in the range 300–900 K," J. Chem. Phys. **59**, 6556–6562 (1973).
[4]D. Chang, F. D. Shields, and H. E. Bass, "Sound-tube measurements of the relaxation frequency of moist nitrogen," J. Acoust. Soc. Am. **62**, 577–581 (1977).
[5]A. J. Zuckerwar and W. A. Griffin, "Resonant tube for measurement of sound absorption in gases at low frequency/pressure ratios," J. Acoust. Soc. Am. **68**, 218–226 (1980).
[6]A. J. Zuckerwar and R. W. Meredith, "Acoustical measurements of vibrational relaxation in moist N_2 at elevated temperatures," J. Acoust. Soc. Am. **71**, 67–73 (1982).
[7]J. E. Carlson and P.-E. Martinsson, "Exploring interaction effects in two-component gas mixtures using orthogonal signal correction of ultrasound pulses," J. Acoust. Soc. Am. **117**, 1961–1968 (2005).
[8]P.-E. Martinsson, "Ultrasonic measurements of molecular relaxation in ethane and carbon monoxide," Proc.-IEEE Ultrason. Symp., **511–516** (2002).
[9]S. G. Ejakov *et al.*, "Acoustic attenuation in gas mixtures with nitrogen: Experimental data and calculations," J. Acoust. Soc. Am. **113**, 1871–1879 (2003).
[10]A. Cottet *et al.*, "Acoustic absorption measurements for characterization of gas mixing," J. Acoust. Soc. Am. **116**, 2081–2088 (2004).
[11]J. D. Lambert, *Vibrational and Rotational Relaxation in Gases* (Clarendon, Oxford, 1977).
[12]A. Petculescu and R. M. Lueptow, "Synthesizing primary molecular relaxation processes in excitable gases using a two-frequency reconstructive algorithm," Phys. Rev. Lett. **94**, 238301 (2005).
[13]K. F. Herzfeld and T. H. Litovitz, *Absorption and Dispersion of Ultrasonic Waves* (Academic Press, 1959), Sec. 37.
[14]A. Petculescu and R. M. Lueptow, "Fine-tuning molecular acoustic models: Sensitivity of the predicted attenuation to the Lennard-Jones parameters," J. Acoust. Soc. Am. **117**, 175-184 (2005).
[15]Y. Dain and R. M. Lueptow, "Acoustic attenuation in three-component gas mixtures—Theory," J. Acoust. Soc. Am. **109**, 1955–1964 (2001).
[16]Y. Dain and R. M. Lueptow, "Acoustic attenuation in a three-gas mixture—Results," J. Acoust. Soc. Am. **110**, 2974–2979 (2001).
[17]R. N. Schwartz, Z. I. Slawsky, and K. F. Herzfeld, "Calculation of vibrational relaxation times in gases," J. Chem. Phys. **20**, 1591–1599 (1952).
[18]Reference 13, pp. 138–156.
[19]T. H. Ruppel and F. D. Shields, "Sound propagation in vibrationally excited N_2/CO and H_2/He/CO gas mixtures," J. Acoust. Soc. Am. **87**, 1134–1137 (1990).
[20]A. J. Zuckerwar and W. A. Griffin, "Effect of water vapor on sound absorption in nitrogen at low frequency/pressure ratios," J. Acoust. Soc. Am. **69**, 150–154 (1981).
[21]A. J. Zuckerwar and K. W. Miller, "Vibrational-vibrational coupling in air at low humidities," J. Acoust. Soc. Am. **84**, 970–977 (1988).
[22]J. M. M. Pinkerton, "On the pulse method of measuring ultrasonic absorption in liquids," Proc. Phys. Soc. London **62**, 286–299 (1949).
[23]NIST Chemistry WebBook (http://webbook.nist.gov/chemistry), accessed 20 December 2005.

On analysis of exponentially decaying pulse signals using stochastic volatility model. Part II: Student-*t* distribution (L)

C. M. Chan
Hong Kong Community College, Hong Kong, China

S. K. Tang[a)]
Department of Building Services Engineering The Hong Kong Polytechnic University, Hong Kong, China

(Received 3 March 2006; revised 5 July 2006; accepted 8 July 2006)

The authors have recently demonstrated how the stochastic volatility model incorporating the exponential power distribution can be used to retrieve the instant of initiation of an exponentially decaying pulse and its decay constants in the presence of background noises. In the present study, the Student-*t* distribution, which can be expressed in a two-stage scale mixtures representation, is adopted in the stochastic volatility model. It is found that the corresponding performance is comparable to that for the case of the exponential power distribution when the signal-to-noise ratio is larger than or equal to 3 dB. The performance deteriorates quickly when the signal-to-noise ratio drops below 0 dB. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2266455]

PACS number(s): 43.60.Uv, 43.60.Jn, 43.60.Cg, 43.60.Wy [EJS] Pages: 1783–1786

I. INTRODUCTION

Pulses and their decays are important in the studies of acoustics and building services engineering. In a diffused system, such decay is an exponential function of time and the rate of the decay, which is commonly described by the decay constant, contains information on the system damping and thus pulse decay analysis has attracted great attention in the past few decades (for instance, Heckl[1] and Tang[2]).

The recent study of the authors[3] illustrates that the stochastic volatility (SV) model incorporating the exponential power distribution (EP) is able to retrieve the instant of the pulse initiation and the decay constant within engineering tolerance even when the background noise level is comparable to that of the pulse. Its performance is much better than that of the conventional short-time Fourier transform when there is a small fluctuation in the frequency of the decaying pulse. Details on the properties of the EP distribution and its application in the Bayesian analysis can be found, for instance, in Choy and Walker.[4]

The Student-*t* distribution is a conventional distribution form and has been used in many applications (for instance, Lange *et al.*,[5] and Chan[7]). It can be expressed in a two-stage scale mixtures representation and can in principle be an alternative to the EP distribution.[6] In the present study, the performance of the SV model incorporating the Student-*t* distribution in analyzing decaying pulses is investigated. The results supplement those of the previous study of the authors.[3]

II. TWO-STAGE SCALE MIXTURES REPRESENTATION

A standard random variable X having the normal scale mixtures representation can be expressed in the form of $X = Z \times \lambda$ where Z is the standard normal random variate and λ is a positive random variate known as the mixing variable having a probability density function g, which can be either continuous or discrete.

Let θ and σ be the location and scale parameter of a scale mixture distribution, respectively. The probability density function of X takes the mixture form

$$f(x) = \int_{\Re^+} N(x|\theta, \lambda\sigma^2) g(\lambda) d\lambda, \tag{1}$$

where $N(\cdot|\cdot)$ denotes the normal density defined on $\Re^+ = (0, \infty)$. In the Bayesian framework, the mixture density in Eq. (1) can be expressed into a two-stage hierarchy of the form

$$X|\theta, \sigma^2, \lambda \sim N(\theta, \lambda\sigma^2), \quad \lambda \sim g(\lambda). \tag{2}$$

The Student-*t* distribution with degrees of freedom α corresponds to an inverse gamma mixing distribution

$$\lambda \sim g(\lambda) = G_{\text{inv}}(0.5\alpha, 0.5\alpha), \tag{3}$$

where $G_{\text{inv}}(a,b)$ is the inverse gamma distribution with density ($a>0$ and $b>0$)

$$g(\lambda) = b^a e^{-b/\lambda} \lambda^{-(a+1)} / \Gamma(a). \tag{4}$$

To facilitate an efficient computation for the SV models, use is made of the class of scale mixtures of uniform representation for the normal density. Since X is a normal random variable with mean θ and variance σ^2, its density function can be rewritten into

$$N(x|\theta, \sigma^2) = \int_{\theta-\sigma\sqrt{u}}^{\theta+\sigma\sqrt{u}} \frac{1}{2\sigma\sqrt{u}} G(u|1.5, 0.5) du, \tag{5}$$

where $G(u|a,b)$ is the gamma density function with parameter a and b.[8] The Student-*t* distribution with a degree of freedom α can be expressed into the following hierarchy:

a)Author to whom correspondence should be addressed. Electronic mail: besktang@polyu.edu.hk

TABLE I. Full conditionals in the Gibbs sampling.

Conditional	Remarks	Distribution
$h_t\|\bar{h}_{-t},\bar{\lambda},\bar{u},\sigma^2,\phi,\bar{y}\sim N(a_t,b_t\sigma^2)$	$h_t>\ln y_t^2-2\ln\beta\ln\lambda_t-\ln u_t$, $a_t=\begin{cases}\phi h_{t+1}-0.5\sigma^2 & t=1\\ \dfrac{\phi(h_{t-1}+h_{t+1})-0.5\sigma^2}{1+\phi^2} & 2\le t\le n=1,\\ \phi h_{t-1}-0.5\sigma^2 & t=n\end{cases}$ $b_t=\begin{cases}1 & t=1,n\\ (1+\phi^2)^{-1} & 2\le t\le n-1\end{cases}$	Truncated normal
$\sigma^2\|\bar{h},\bar{\lambda},\bar{u},\phi,\bar{y}\sim G_{\text{inv}}\left(a_\sigma+\frac{n}{2},b_\sigma+\frac{1}{2}\left[(1-\phi^2)h_1^2+\sum_{t=2}^{n}(h_t-\phi h_{t-1})^2\right]\right)$	N.A.	Inverse gamma
$\lambda_t\|\bar{h},\bar{\lambda}_{-t},\bar{u},\phi,\sigma^2,\bar{y}\sim G_{\text{inv}}\left(\frac{\alpha+1}{2},\frac{\alpha}{2}\right)$	$\lambda_t>\frac{y_t^2}{\beta^2H_tu_t}$	Truncated inverse gamma
$u_t\|\bar{h},\bar{\lambda},\bar{u}_{-t},\phi,\sigma^2,\bar{y}\sim\exp(0.5)$	$u_t>\frac{y_t^2}{\beta^2H_t\lambda_t}$	Truncated exponential
$\phi\|\bar{h},\bar{\lambda},\bar{u},\sigma^2,\bar{y}\sim N\left(\frac{\sum_{t=2}^{n}h_{t-1}h_t}{\sum_{t=2}^{n}h_t^2},\frac{\sigma^2}{\sum_{t=2}^{n}h_t^2}\right)(1+\phi)^{a_\phi-1/2}(1-\phi)^{b_\phi-1/2}$	$\|\phi\|\le1$	Product of normal and shifted beta

$$X|\theta,\sigma^2,\lambda,u\sim U(\theta-\sigma\sqrt{\lambda u},\theta+\sigma\sqrt{\lambda u})$$

$$,\lambda\sim G_{inv}(0.5\alpha,0.5\alpha),\quad\text{and}\quad u\sim G(1.5,0.5),\tag{6}$$

where $U(a,b)$ is a uniform distribution defined on the interval (a,b).

III. BAYESIAN STUDENT-*T* SV MODEL AND GIBBS SAMPLING

Let H_t and h_t be the volatilities, and log-volatilities, respectively. In the SV model, the signal data, y_t (where $t=1,2,\ldots,n$), is defined as

$$y_t=\beta\sqrt{H_t}\varepsilon_t\quad\text{and}\quad h_t=\begin{cases}\sigma\eta_1/\sqrt{1-\phi^2} & t=1\\ \phi h_{t-1}+\sigma\eta_t & t>1\end{cases},\tag{7}$$

where $\{\varepsilon_t\}$ and $\{\eta_t\}$ are independent standard Gaussian processes. β is a constant representing the model instantaneous volatility, σ the variance of the log volatilities, and $\phi\in(-1,1)$ the persistence of the volatility.

In the present study, the signal data set is modeled by a Student-t distribution while the log volatility is assumed to follow a normal distribution

$$y_t|h_t\sim t_\alpha(0,\beta^2H_t)\tag{8}$$

and

$$h_t|h_{t-1},\phi,\sigma^2\sim N(\phi h_{t-1},\sigma^2),\tag{9}$$

while the marginal distribution is $h_t|\phi,\sigma^2\sim N[0,\sigma^2/(1-\phi^2)]$.

To complete the Bayesian framework, the shifted beta and inverse gamma distributions are assigned to be the independent priors for ϕ and σ^2, respectively

$$\phi\sim 2Be(a_\phi,b_\phi)-1\quad\text{and}\quad\sigma^2\sim G_{\text{inv}}(a_\sigma,b_\sigma).\tag{10}$$

The SV model can then be rewritten hierarchically as

$$y_t|h_t,\lambda_t,u_t\sim U(-\beta H_t^{1/2}\lambda_t^{1/2}u_t^{1/2},\beta H_t^{1/2}\lambda_t^{1/2}u_t^{1/2})\quad\text{and}$$

$$h_t|h_{t-1},\phi,\sigma^2\sim N(\phi h_{t-1},\sigma^2)\tag{11}$$

with $\lambda_t\sim G_{inv}(0.5\alpha,0.5\alpha)$, $u_t\sim G(1.5,0.5)$, $\phi\sim 2Be(a_\phi,b_\alpha)-1$, and $\sigma^2\sim G_{inv}(a_\sigma,b_\sigma)$. It is implemented using the Gibbs sampling approach with the variables $\bar{y}=(y_1,\ldots,y_n)$, $\bar{h}=(h_1,\ldots,h_n)$, $\bar{\lambda}=(\lambda_1,\ldots,\lambda_n)$, $\bar{u}=(u_1,\ldots,u_n)$, $\bar{h}_{-t}=(h_1,\ldots,h_{t-1},h_{t+1},\ldots,h_n)$, $\bar{\lambda}_{-t}=(\lambda_1,\ldots,\lambda_{t-1},\lambda_{t+1},\ldots,\lambda_n)$, and $\bar{u}_{-t}=(u_1,\ldots,u_{t-1},u_{t+1},\ldots,u_n)$.

With arbitrarily chosen starting values for $\bar{h},\bar{\lambda},\bar{u},\sigma^2$, and ϕ, the Gibbs sampler iteratively sample random variates from a system of full conditional distributions and the resulting simulations are used to mimic a random sample from the

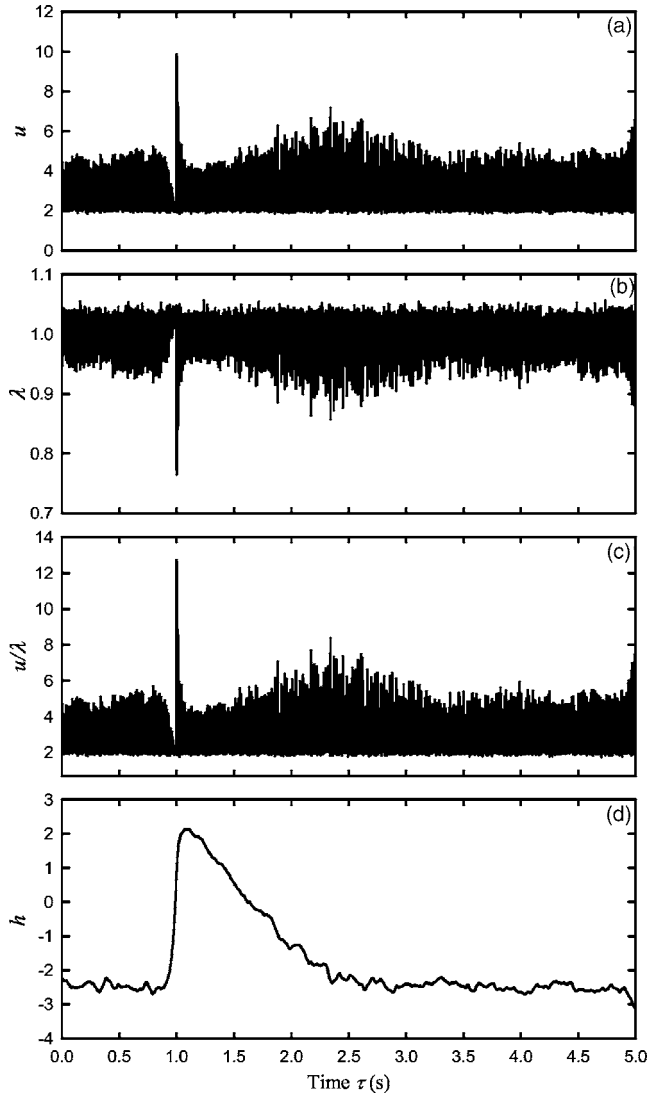


FIG. 1. Time variations of the mixing parameters for S/N=10 dB and α =30. (a) u; (b) λ; (c) u/λ; (d) h.

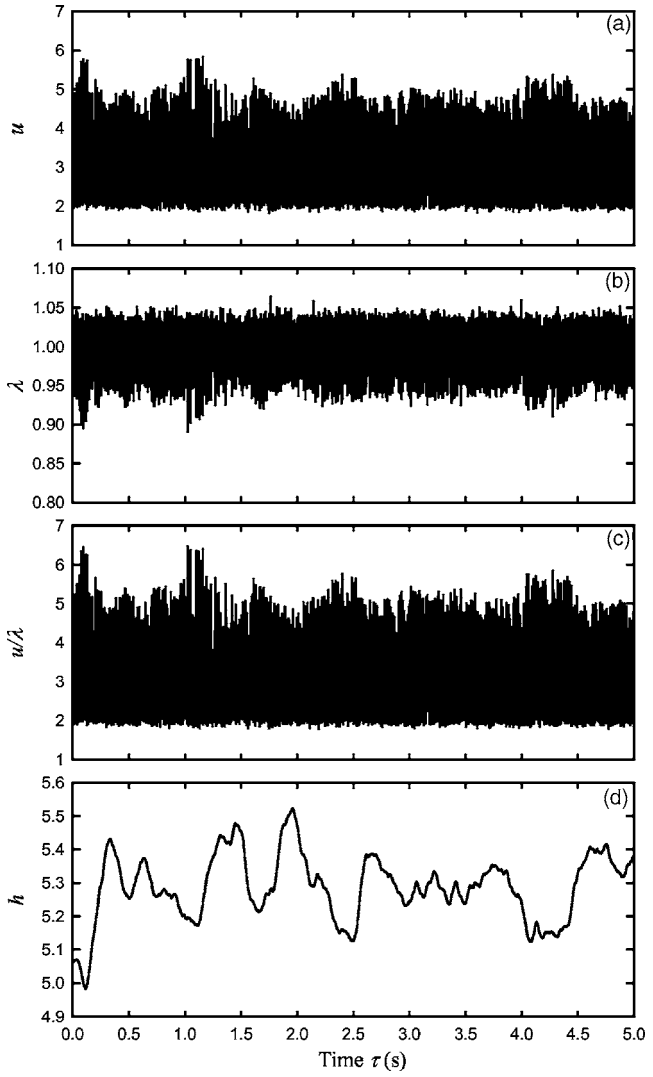


FIG. 2. Time variations of the mixing parameters for S/N=−3 dB and α =30. (a) u; (b) λ; (c) u/λ; (d) h.

targeted joint posterior distribution. The system of full conditionals is given in Table I. In the present study, β is fixed at unity.

IV. NUMERICAL EXAMPLES

The artificial signals in Chan *et al.*[3] are adopted again in the present study. They consist of an exponentially decaying harmonic wave s and Gaussian white noises v of various levels

$$y(\tau) = e^{-\eta(\tau-\tau_o)}\cos[2\pi f(\tau-\tau_o)]H(\tau-\tau_o) + v(\tau), \qquad (12)$$

where H denotes the Heavside step function, f the frequency of the wave, η the decay constant, τ_o the instant of pulse initiation, and τ the time in second. The signal-to-noise ratio (S/N) in decibels is defined as

$$\text{S/N} = 10\log_{10}(|s|_{\max}/|v|_{\max}). \qquad (13)$$

Without loss of generality, η is fixed at 2 and f is normally set at 50 Hz in the present study.

The parameter α in the SV model determines the shapes of the Student-t distributions and thus has a significant impact on the modeling of the decaying signal by the SV model. However, since the shape of the Student-t distribution does not vary much for $\alpha > 30$ and the computation with large α is in general impractical as it is very demanding on computing resources, the largest α included in the present investigation is 30. The incorporation of the Student-t distribution into the SV model gives rise to two mixing parameters, namely, u and λ, and a log-volatility h as shown in the previous section. The two mixing parameters illustrate fluctuation while the latter follows the envelope of the signal magnitude.

Figure 1 illustrates the time variations of $u, \lambda, u/\lambda$, and h for S/N=10 dB and α=30. One can observe that the variation patterns of u and λ are opposite. The initiation of the pulse results in a prominent upward and downward spike in u and λ, respectively. The parameter u/λ shows a more prominent spike at the instant of pulse initiation. The present u is basically the same as those obtained using the EP distribution with a kurtosis of 0.75.[3] The variation pattern of h is also very close to that shown in Chan *et al.*[3] Further increas-

TABLE II. Estimated τ_o and η.[a]

S/N (dB)	τ_o (s)	$\eta(s^{-1})$
$+\infty$	1.000(1.000)	1.95(2.00)
10	1.000(1.000)	2.30 (2.05)
3	1.000(1.000)	2.20(2.16)
0	1.010(1.001)	2.60 (2.42)
−3	1.022(1.002)	—(2.36)

[a]Numbers in parenthesis are those obtained with the EP distribution (Ref. 3).

ing α may result in a slightly better performance, but the very computer resources demanding calculation makes it impractical.

The performance of the SV model deteriorates when the background noise level increases. For S/N=−3 dB with $\alpha=30, u, \lambda$, and u/λ are unable to indicate without ambiguity the instant of the pulse initiation [Figs. 2(a) to 2(c)]. The log-volatility h does not even suggest the presence of a decaying pulse [Fig. 2(d)]. However, the pulse initiation and its eventual decay are still observable with the EP distribution at this signal-to-noise ratio level.[3]

Results for $\alpha<30$ are not presented as they are all worse than those shown in Figs. 1 and 2. Table II summarizes the performance of the present SV model and has it compared with that of the previous study of the authors.[3] One can find that the use of the EP distribution in the SV model results in a better analysis when the S/N drops below 3 dB. The somewhat worse performance of the Student-t distribution overall may be due to the relatively longer tail of the Student-t distribution compared to those of the EP family with small kurtosis even when the degree of freedom is large.

V. CONCLUSIONS

A stochastic volatility model which incorporates the Student-t distribution is used in the present study to retrieve the instant of initiation and the decay constant of an exponentially decaying signal in the presence of random background noises. It is found that the performance of the Student-t distribution is comparable to that of the EP distribution for a signal-to-noise ratio higher than or equal to 3 dB. It deteriorates quickly when this ratio falls below 0 dB.

ACKNOWLEDGMENT

C.M.C. is supported by a staff development program of the Hong Kong Community College, The Hong Kong Polytechnic University.

[1]M. Heckl, "Measurement of absorption coefficients on plates," J. Acoust. Soc. Am. **34**, 803–808 (1962).
[2]S. K. Tang, "On the time-frequency analysis of signals that decay exponentially with time," J. Sound Vib. **234**, 241–258 (2000).
[3]C. M. Chan, S. K. Tang, and H. Wong, "On analysis of exponentially decaying pulse signals using stochastic volatility model," J. Acoust. Soc. Am. **119**, 1519–1526 (2006).
[4]S. T. B. Choy and S. G. Walker, "The extended exponential power distribution and Bayesian robustness," Stat. Probab. Lett. **65**, 227–232 (2003).
[5]K. L. Lange, R. J. A. Little, and J. M. G. Taylor, "Robust statistical modeling using the t distribution," J. Am. Stat. Assoc. **84**, 881–896 (1989).
[6]J. Geweke, "Bayesian treatment of the independent Student-t linear model," J. Appl. Econ. **8**, Issue S, S19–40 (1993).
[7]C. M. Chan, "On a topic of Bayesian analysis using scale mixtures distributions," MPhil thesis, Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China (2001).
[8]S. T. B. Choy and A. F. M. Smith, "Hierarchical models with scale mixtures of normal distribution," TEST **6**, 205–211 (1997).

Auditory masking: Need for improved conceptual structure (L)[a)]

Nat Durlach[b)]
Hearing Research Center, Boston University Boston, Massachusetts 02215

(Received 27 January 2005; revised 14 July 2006; accepted 15 July 2006)

Even in simpler times, some people (e.g., Tanner) found it useful to ask "What is masking?" Independent of the extent to which this question was adequately answered even in those times, it is clear that the current expanded interest in central auditory processing has raised this question anew. In this note, comments are made about masking-related issues that illustrate the kinds of questions that need to be considered in attempting to develop a conceptual structure that can be effectively used to define, classify, study, and model auditory masking. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2335426]

PACS number(s): 43.66.Ba, 43.66.Dc [RAL] Pages: 1787–1790

Masking has long been of interest to auditory scientists. Recently, both activity level and range of topics in this area has increased to such an extent that the area appears to be in total disarray. Not only is there no overarching conceptual structure available to organize the area and provide it with scientific elegance, but there are few definitions that evidence even a modest degree of scientific stability (varying across both individuals and time). The material in this Letter in no way solves these problems; it is aimed merely at stimulating discussion that can help reduce them.[1]

In general, masking can be said to occur whenever the reception of a specified set of acoustic stimuli ("targets") is degraded by the presence of other stimuli ("maskers"). In this letter, attention is focused on behavioral (rather than physiological) aspects of masking and the magnitude of the masking is assumed to be measured by the elevation in threshold caused by the presence of the masker in the given behavioral (psychoacoustic) task.

Consistent with the effort to build a solid knowledge base, much past research on masking has made use of simple stimuli and tasks, and has been concerned with interactions between target and masker that occur in the auditory periphery (i.e., on *peripheral masking*). When central processing has been considered, rather than examining phenomena related to interference between target and masker in central portions of the system (i.e., *central masking*), attention has often been focused on how central processing reduces the effects of peripheral masking (i.e., central *unmasking*). This trend is evident, for example, in work on binaural unmasking (Durlach and Colburn, 1978; Gilkey and Anderson, 1997) and comodulation release from masking (Hall *et al.*, 1995; Verhey *et al.*, 2003). Although there have been important exceptions to this focus on peripheral *masking* and central *unmasking* (Carhart *et al.*, 1969; Pollack, 1975; Watson *et al.*, 1976; Relkin and Turner, 1988), the main thrust appears to the writer as just described.

Recently, however, attention to central masking has substantially increased. Presumably, this is due in part to an increased emphasis on central processing in general (driven by increased knowledge about peripheral processing and the desire to "work one's way up the system," as well as by major advances in brain imaging techniques). A further factor is the increased concern with realistic acoustic environments/tasks and "everyday problems" (reverberation, multiple maskers, etc.). Finally, of course, is the increased interest in cognitive "top-down" processing and auditory scene analysis, grouping, segregation, object formation, and auditory attention (Bregman, 1990; Ellis, 1999; Cooke and Ellis, 2001). In the following sections, the need for increased conceptual clarity is illustrated by considering some elementary nomenclature problems, informational versus energetic masking, and additivity of masking.

I. ELEMENTARY NOMENCLATURE PROBLEMS

The above definition of masking covers many phenomena. Some investigators would use another term (e.g., "interference") at this general level and reserve the term "masking" for situations that are more constrained. One constraint sometimes applied relates to the paradigm used to make the measurements. Thus, for example, masking appears more closely coupled to detection than to discrimination or recognition. Another constraint relates to beliefs about the level of processing being probed; the term "masking' is more likely to be used when it is thought that the experiments shed more light on peripheral than on central processing.

That such considerations cannot by themselves provide adequate guidance for a logical definitional structure, however, is suggested by noting the difficulties involved in the apparently simple task of distinguishing between detection and discrimination. It is obvious that any masked detection experiment can be thought of as a discrimination experiment (e.g., see Tanner, 1958); the task of detecting a target $T(t)$ in the presence of a masker $M(t)$ is identical to the task of discriminating between $M(t)$ and $M(t)+T(t)$. It is less obvious, but equally true, that any discrimination experiment can be thought of as a masking experiment; discriminating between signals $A(t)$ and $B(t)$ is identical to the task of detect-

[a)]Alternate subtitles include (a) Stroll through the conceptual woods: (b) Stirring the conceptual waters; (c) Conceptual problems by the ton; (d) Egads, what a mess!, and (e) Help!
[b)]Electronic mail: durlach@mit.edu

ing the target $T(t)=B(t)-A(t)$ in the presence of the masker $M(t)=A(t)$. Such ambivalence in nomenclature is clearly evident in the area of intensity resolution as well as detection. Although less evident, the same problem exists in frequency resolution; discriminating between tones of frequency f and $f+\Delta f$ is identical to detecting a target signal consisting of the difference between the two tone waveforms in the presence of a masker consisting of one of these waveforms. One approach to solving this nomenclature problem involves applying constraints to the relationship between $M(t)$ and $T(t)$ when the experiment is expressed as a detection experiment.[2] Note finally that the nomenclature problem becomes even larger when recognition (or equivalently, identification) is included. Just as "detection" and "discrimination are interchangeable, both of these terms are interchangeable with "recognition" when only two stimuli are involved.

A further general area in which basic nomenclature issues occur relates to the kind of disturbance suffered by the signal. In particular, consider the relationship between masking and distortion. Making use of frequency-domain representations, and restricting attention to distortions that are linear filters (e.g., the distortion caused by reverberation), one can contrast the disturbance of a target $T(\omega)$ caused by a masker $M(\omega)$ to the disturbance caused by a distortion $D(\omega)$ as follows:

$$\text{Masking} \quad T(\omega) \rightarrow T(\omega) + M(\omega)$$
$$\text{Distortion} \quad T(\omega) \rightarrow T(\omega)D(\omega).$$

Note, however, that the distortion case can also be written $T(\omega) \rightarrow T(\omega)+T(\omega)[D(\omega)-1]$, which transforms it into a masking case with $M(\omega)=T(\omega)[D(\omega)-1]$. Again, as in the attempt to distinguish between detection and discrimination, the key to distinguishing between masking and distortion appears to lie in the relationship between target and masker when the experiment is viewed as a masking experiment. Exactly how such relationships can be most usefully categorized and named, however, remains unclear.

II. INFORMATIONAL MASKING (IM) VERSUS ENERGETIC MASKING (EM)

A major contributor to the increased activity in masking, as well as the increased chaos, is informational masking (IM). (Important IM research has been done by Watson, Neff, Lutfi, Kidd, Brungart, Richards, Wright, Freyman, and Shinn-Cunningham, as well as their associates—see references in Lutfi *et al.*, 2003; Brungart *et al.*, 2005; Watson, 2005; Durlach *et al.*, 2005; Kidd *et al.*, 2006). Although efforts have been made to clarify definitional issues (Durlach *et al.*, 2003a; Watson, 2005), these issues are far from settled. Most investigators use "energetic masking" (EM) to refer to masking that occurs due to "overlap between target and masker at the periphery" and "informational masking" (IM) to refer to non-energetic masking.

Even if one defines IM as all masking that is not energetic (thus roughly equating EM-versus-IM with peripheral-versus-central), there remains the problem of specifying what is meant by "periphery" and "overlap." In Durlach *et al.* (2003a), it is noted that (1) although the periphery is usually identified with the auditory nerve, for some purposes identification with a higher site might be useful; and (2) at any site, "overlap" can be defined and measured by the extent to which the ideal processor operating at that site evidences poor performance. According to this definitional structure, what appears as IM at one level can appear as EM at a higher level (via overlap between target and masker in some more central channel), and that all masking is EM if examined at a sufficiently high level. [3]

Apart from issues related to the definition of EM, are issues concerned with what kinds of $\overline{\text{EM}}$ should be called IM (where $\overline{\text{EM}}$ denotes all masking that is *not* EM) and what the causes of such masking might be. The association of IM with the effects of stimulus uncertainty not only has a long history, but also forms the basis of most of the theoretical work on IM (Lutfi, 1993; Oh and Lutfi, 1998; Richards and Neff, 2004). However, it is obvious that there are factors other than uncertainty that influence the amount of $\overline{\text{EM}}$ observed. Prominent among these is target-masker "similarity" (Durlach *et al.*, 2003b; Watson, 2005). Among the difficulties here are the many dimensions along which target and masker can be similar, the lack of a generally applicable quantitative measure of similarity, and the lack of any model that takes account of both similarity and uncertainty (which is capable of predicting contours of constant $\overline{\text{EM}}$ in uncertainty-similarity space). Furthermore, there are undoubtedly factors beyond uncertainty and similarity that need to be considered in the $\overline{\text{EM}}$ domain. In general, $\overline{\text{EM}}$ occurs because the listener has difficulty segregating target and masker, identifying target and masker, and attending to the target (or disattending to the masker). As has already been demonstrated in auditory scene analysis, there are many factors that play important roles in $\overline{\text{EM}}$.

Beyond difficulties associated with definitions and relations among various concepts in the $\overline{\text{EM}}$ domain, there currently exist many experimental results in this area that are quite startling. For example, in Kidd *et al.* (2005), a monaurally presented speech target processed to have energy only in a set of narrow frequency bands was partially masked by an independent ipsilateral speech masker processed to have energy only in a disjoint complementary set of frequency bands (so that EM was minimal). It was then shown that the speech-target intelligibility could be increased by adding an ipsilateral noise masker processed to have energy only in the same set of bands as the speech masker (so that the speech masker, but not the speech target, was energetically masked by the noise masker). It was then further shown, surprisingly, that additional improvement in the speech target intelligibility could be achieved by switching the noise masker to the contralateral ear (a change that reduced the EM of the speech masker by the noise masker). In a second example (Durlach *et al.*, 2005), concerned with yes-no one-interval detection of a target tone in a simultaneous random broadband multitone masker (constrained to have minimal energy around the target frequency), it was found that the elevation in threshold (measured in terms of the sensitivity index d') associated with the uncertainty of the masker spectrum (i.e., the IM) was explainable (for different listeners) in terms of either (1) the listener opening up the acceptance filter to extend way

beyond the normal critical band or (2) the tails of the masking spectra "leaking through" the critical band. However, the response bias β measured in case (2) was exactly opposite to that which one would expect in this case. Thus, the results were surprising not only in that the d′ values obtained with the nominal IM situation could be explained for both cases (1) and (2) in terms of peripheral processing (although central processing obviously played a role in how the peripheral-processing resources were exploited), but also in that the β values obtained in case (2) were the negative of those expected. Again, as with the previous example, we do not even know how to talk about these results in a consistent and meaningful manner.

III. ADDITIVITY OF MASKING

Another major topic in masking that evidences conceptual problems is that of masking additivity. (Research on this topic has been done by Penner, Humes, Jesteadt, Lutfi, Moore, Neff, Green, Bilger, and Oxenham, as well as their associates—see references in Lutfi, 1985; Neff and Jesteadt, 1996; Oxenham and Moore, 1995). One complicating factor is that addition-of-masking effects can arise in connection with (1) the addition of different masking *signals* or (2) the addition of different masking *effects* (e.g., arising from energetic and informational masking) of a single masking signal.[4]

Futhermore, even when attention is restricted entirely to the former case, the situation is unclear. In particular, there is no general theory available for specifying how $m(M_1+M_2,T)$ is related to $m(M_1,T)$, and $m(M_2,T)$, where m denotes the amount of masking, M_1 and M_2 two masking signals, and T a target signal. Although there exist cases that satisfy the simple additivity rule $m(M_1+M_2,T)=m(M_1,T)+m(M_2,T)$, there are many more cases that do not.

Cases in which $m(M_1+M_2,T)\gg m(M_1,T)+m(M_2,T)$, often referred to as "excess masking," occur in a variety of situations. Even ignoring details of peripheral processing, one is led to excess masking by the multiprocessing notion of Green (1967); by the multimasker-penalty notion associated with listener Min behavior as described in Durlach *et al.* (2003a); and by the simple observation that excess masking will hold whenever T can be represented as the sum of two components $T=T_i+T_j$ with M_i chosen to be an effective masker of T_i but not of T_j $(i=1,2;j=1,2;i\neq j)$. Most research concerned with excess masking, however, has been focused on compressive nonlinearities and off-frequency listening. This research is concerned mainly with peripheral masking and encompasses both simultaneous masking and forward and backward masking. The model that appears most successful here is the modified power-law model studied by Humes and collaborators (Humes and Jesteadt, 1989). This model not only exploits the idea of nonlinear transformations in the periphery, but also considers the existence (and compression) of peripheral internal noise, interactions between different maskers, and non-energetic masking.

Cases in which $m(M_1+M_2,T)\ll m(M_1,T)+m(M_2,T)$ can occur when the maskers M_1 and M_2 not only mask T, but also each other. To the extent that M_1 masks M_2, the masking of T by M_2 is reduced; and to the extent that M_2 masks M_1 the masking of T by M_1 is reduced. In such circumstances, which are likely to arise whenever M_1 and M_2 are not widely separated in frequency and/or time, computation of $m(M_1+M_2,T)$ can become complex because of the iterative nature of the interaction. A simple example of this inequality can occur when the two maskers consist of tones close enough in frequency to cause perceptual beats. More complex examples can occur in the areas of IM and binaural interaction (Kidd *et al.*, (2005); Durlach and Colburn, 1978; Freyman *et al.*, 1999; Brungart *et al.*, 2005).

Beyond the violations of additivity cited above, there is a fundamental issue concerning the possibility that masking may not even be able to satisfy a generalized combination law of the type $m(M_1+M_2,T)=F[m(M_1,T),m(M_2,T)]$, where F is *any* function of two variables and m is *any* well-defined measure of masking effectiveness. A simple observation that suggests the inadequacy of even this general relationship is the following. Let T be any target and M any masker; let M_1 be any signal whatsoever and define $M_2=M-M_1$. To the extent that the above law can be realized, the quantity $F[m(M_1,T),m(M_2,T)]$ must be equal to the constant $m(M,T)$ for any signal M_1. In other words, no matter how a masker M is subdivided into two component maskers M_1 and $M-M_1$, the quantity $F[m(M_1,T),m(M-M_1,T)]$ must remain invariant.[5]

IV. CONCLUDING REMARKS

Sections I–III contain a sampling of issues related to auditory masking that illustrate some of the conceptual chaos in this area. Reducing this chaos will require the development of a coherent conceptual scheme for differentiating, classifying, and labeling various kinds of masking that is capable of transforming the current collection of masking configurations and results (i.e., the current masking "bestiary") into a hierarchical lattice that encompasses meaningful relationships among the entries in the lattice. Presumably, the key dimensions of this lattice will be those describing constraints on the relationship between $M(t)$ and $T(t)$, as well as those describing how the auditory system responds to the various types of masking configurations.

Although the distinction between EM and IM masking constitutes a beginning step in this direction, as currently constituted it is a very crude one. At the moment, many of us are just lazily dumping all masking that cannot be explained by "overlap at the periphery" into an undifferentiated "informational masking" basket. Notions related to uncertainty, similarity, attention, memory, etc., which are sometimes introduced in connection with such maskers, are obviously relevant to this classification problem, but they have not yet been adequately described in the masking context.

These same kinds of problems are also evident when one considers the results of combining different maskers. Although significant progress on developing models for additivity of masking was made a number of years ago for a relatively restricted set of masking configurations, there has been little recent follow through in this area. In addition to the problems recognized at the time of these developments, the growing concern with central masking has added a whole

new set of problems. In the author's opinion, an adequate model of additivity will now require an explicit partitioning of masking configuration into well-defined subsets and the assignment of a different combination rule for each subset. This area appears particularly complicated because specification of the masking configuration in this area includes specification of not only the relationship between the target and each masker, but also the relationship among the various maskers. Obviously, it would be helpful if large coverage could be obtained with relatively few combination rules and the associated classification schemes were relatively consistent with the classification schemes constructed in connection with theory development for the single-masker case.

Finally, although the masking area currently appears to be in chaos, it is not clear how this chaos can best be reduced. Although certain of the above comments may hint at paths to be explored, none really examines these paths nor provides an overview of what paths should be considered. Even if one decided that the term "masking" should be used only in connection with detection/peripheral processing and a term such as "interference" used for the general case (a stance already assumed by some auditory scientists), most of the problems mentioned above would remain.

ACKNOWLEDGMENTS

The author is indebted to Chris Mason, Erick Gallun, Steve Colburn, Gerald Kidd, Lorraine Delhorne, the JASA Reviewers, and the JASA Associate Editor for help in preparing this note (work supported by NIH and AFOSR).

[1]The author's disparaging remarks about the state of masking should not obscure the high regard the author has for the excellent past research in this area; it is precisely this high regard, combined with the major challenges now evident, that has motivated this letter.

[2]Note, however, that comparison of $M(t)$ and $T(t)$ for definitional purposes is distinct from comparison for performance-prediction purposes; in the latter case, it is the comparison of $M(t)$ with $M(t)+T(t)$ that counts. The extent to which auditory scientists tend to ignore this distinction is quite remarkable [see, for example, how some people, including the author, were misled into thinking that binaural unmasking could be understood by comparing the lateralization of $M(t)$ alone to that of $T(t)$ alone].

[3]Inasmuch as actual application of this definitional structure requires the acquisition and modeling of data on neural firing patterns, as well as complex mathematical analyses, the near-term value of this definitional structure appears more theoretical than practical.

[4]Furthermore, in the latter case, the extent to which it is possible to decompose the single masking signal into different masking signals corresponding to the different masking effects has not yet been determined.

[5]A second argument against the feasibility of such a law posits that M_1 and M_2 are identical except for an intermasker phase shift Ø. As Ø varies from π to 0, M_1+M_2 varies from O to $2M_1$, $m(M_1+M_2,T)$ varies from $m(0,T)$ to $m(2M_1,T)$, and (for many conditions) $m(M_1,T)$ and $m(M_2,T)$ remain constant. In other words, fixed values of $m(M_1,T)$ and $m(M_2,T)$ can lead to different values of $m(M_1+M_2,T)$, thus implying that F is not even a well-defined function.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).

Brungart, D. S., Simpson, B. D., and Freyman, R. L. (**2005**). "Precedence-based speech segregation in a virtual auditory environment," J. Acoust. Soc. Am. **118**, 3241–3251.

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

Cooke, M., and Ellis, D. P. W. (**2001**). "The auditory organization of speech and other sources in listeners and computational models," Speech Commun. **35**, 141–177.

Durlach, N. I., and Colburn, H. S. (**1978**). "Binaural Phenomena" in *Handbook of Perception*, *Vol*. **IV**, *Hearing*, edited by E. C. Carterette, and M. P. Friedman (Academic Press, New York), pp. 365–466.

Durlach, N. I., Mason, C. R., Gallun, F. J., Shinn-Cunningham, B., Colburn, H. S., and Kidd, G., Jr. (**2005**). "Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers," J. Acoust. Soc. Am. **118**, 2482–2497.

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

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G. Jr. (**2003b**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.

Ellis, D. P. W. (**1999**). "Using knowledge to organize sound: The prediction-driven approach to computational auditory scene analysis and its application to speech/nonspeech mixtures, Speech Commun. **27**, 281–298.

Freyman, R., Helfter, K., McCall, D., and Clifton, R. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3587.

Gilkey, R. H., and Anderson, T. R., eds. (**1997**). *Binaural and Spatial Hearing* (Erlbaum Press, Mahwah, NJ).

Green, D. M. (**1967**). "Additivity of masking," J. Acoust. Soc. Am. **41**, 1517–1525.

Hall, J. W., III, Grose, J. H., and Mendoza, L. (**1995**). "Across-channel processes in masking," in *Handbook of Perception*, *Vol*. **VI**, *Hearing*, edited by B. C. J. Moore (Academic Press, San Diego, CA).

Humes, L. E., and Jesteadt, W. (**1989**). "Model of the additivity of masking," J. Acoust. Soc. Am. **85**, 1285–1294.

Kidd, G., Jr., Mason, C. R., and Gallun, F. J. (**2005**). "Combining energetic and informational masking for speech identification," J. Acoust. Soc. Am. **118**, 982–992.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (**2006**). "Informational masking" in *Springer Handbook of Auditory Research: Auditory Perception of Sound Sources*, W. A. Yost, ed. (Springer, NY).

Lutfi, R. A. (**1985**). "A power-law transformation predicting masking by sounds with complex spectra," J. Acoust. Soc. Am. **77**, 2128–2136.

Lutfi, R. A. (**1993**). "A model of auditory pattern analysis based on component-relation-entropy," J. Acoust. Soc. Am. **88**, 2607–2610.

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

Neff, D. L., and Jesteadt, W. (**1996**). "Intensity discrimination in the presence of random-frequency multicomponent maskers and broadband noise," J. Acoust. Soc. Am. **100**, 2289–2298.

Oh, E. L., and Lutfi, R. A. (**1998**). "Nonmonotonicity of informational masking," J. Acoust. Soc. Am. **104**, 3489–3499.

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.

Pollack, I. (**1975**). "Auditory informational masking," J. Acoust. Soc. Am. **57**, S5.

Relkin, E. M., and Turner, C. W. (**1988**). "A reexamination of forward masking in the auditory nerve," J. Acoust. Soc. Am. **84**, 584–591.

Richards, V. M., and Neff, D. L. (**2004**). "Cuing effects for informational masking," J. Acoust. Soc. Am. **115**, 289–300.

Tanner, W. P., Jr. (**1958**). "What is Masking?", J. Acoust. Soc. Am. **30**, 919–921.

Verhey, J. L., Pressnitzer, D., and Winter, I. M. (**2003**). "The psychophysics and physiology of comodulation masking release," Exp. Brain Res. **153**, 405–417.

Watson, C. S. (**2005**). "Some comments on informational masking," Acta. Acust. Acust. **91**, 502–512.

Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). "Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty," J. Acoust. Soc. Am. **60**, 1176–1186.

Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans (L)

Erik Bresch, Jon Nielsen, Krishna Nayak, and Shrikanth Narayanan
Department of Electrical Engineering, University of Southern California, Los Angeles, California 90089

(Received 28 February 2006; revised 13 July 2006; accepted 14 July 2006)

This letter describes a data acquisition setup for recording, and processing, running speech from a person in a magnetic resonance imaging (MRI) scanner. The main focus is on ensuring synchronicity between image and audio acquisition, and in obtaining good signal to noise ratio to facilitate further speech analysis and modeling. A field-programmable gate array based hardware design for synchronizing the scanner image acquisition to other external data such as audio is described. The audio setup itself features two fiber optical microphones and a noise-canceling filter. Two noise cancellation methods are described including a novel approach using a pulse sequence specific model of the gradient noise of the MRI scanner. The setup is useful for scientific speech production studies. Sample results of speech and singing data acquired and processed using the proposed method are given. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2335423]

PACS number(s): 43.70.Jt [BHS] Pages: 1791–1794

INTRODUCTION

In recent years, magnetic resonance imaging has become a viable tool for investigating speech production. Technological advances have enabled studying the structure of the vocal tract, and its dynamical shaping, during speech production. For example, tongue deformation characteristics have been studied with a cine-magnetic resonance imaging (MRI) technique[1] and a real-time magnetic resonance (MR) imaging technique described in Ref. 2 has been successfully used to capture the changing midsagittal shape of the vocal tract during speech production. One methodological challenge, however, is in synchronizing the acquisition of an audio signal with the collection of time-varying vocal tract images, which is important for any subsequent analysis and modeling of the acoustic-articulatory relation. In Ref. 1 the audio signal was recorded in a separate procedure after the MR images were collected so that synchronicity of the signals and images could be only approximately achieved through extensive training of the subject and with a restriction to few utterances.

There have been few studies where MR images and audio signals were obtained simultaneously. The problem is posed by the high intensity gradient noise caused by the scanner, which is in the audible frequency range. This degrades the audio signal such that acoustic analysis of the speech content is difficult, if not impossible. Previous studies such as Ref. 3 have addressed this problem using a correlation-subtraction method, where one captures the noise signal separately and relies on its stationarity. This method does not, however, account for nonstationary noise sources such as body movement of the subject or vibration of the cooling pump.

There are commercially available noise mitigation solutions that have been used in some MRI studies, such as the one by Phone-OR[4,5] which provides an integrated MR-compatible fiber-optical microphone system that allows both real-time and offline noise cancellation. This proprietary system is described to use a special microphone assembly which houses two transducers, one to capture the speech signal and one to capture only the ambient noise. The two microphones are mounted in close proximity but their directional characteristics are at a 90 deg angle so that one (main) microphone is oriented towards the mouth of the subject to capture the speech signal and the other (reference) microphone is oriented such that it rejects the speech signal and captures only the ambient noise. In our own experiments with this system, however, the reference signal contained a strong speech signal component and the subsequent noise cancellation procedure would remove the desired speech signal in addition to the noise to an extent that was undesirable for further analysis of the signal!

The purpose of this letter is to describe the development of an alternative system in which a separate fiber optical microphone was located away from the subject and outside the magnet, but inside the scanner room, in a place where it captures almost exclusively the ambient noise and not the subject's speech. This system captures the audio and the MR images simultaneously and ensures absolute synchronicity for spontaneous speech and other vocal productions including singing.

SYSTEM LEVEL DESCRIPTION OF THE DATA ACQUISITION SYSTEM

Figure 1 illustrates how the various components of the data acquisition system are located in the scan room, the systems room, and the control room of the MRI facility. Two fiber optical microphones are located in the scan room. The main microphone is approximately 0.5 in. (1.3 cm) away from the subject's mouth at a 20 deg angle, and the reference microphone is positioned on the outside of the magnet, roughly 3 ft. (0.9 m) away from the sidewall at a height of

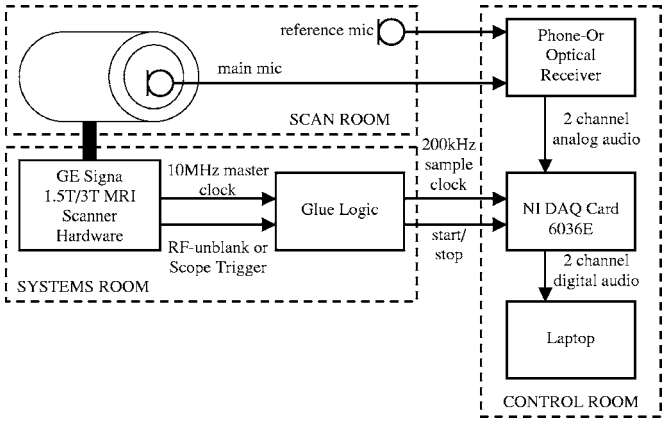


FIG. 1. System level diagram of the audio acquisition system.

about 4 ft. (1.2 m). The microphones connect to the optical receiver box, which is located in the MRI control room.

The data are recorded on a laptop computer using a National Instruments NI-DAQ 6036E PCMCIA card,[6] which provides a total sample rate of 200 kHz and supports up to 16 analog input channels. The main and the reference microphone signal are sampled at 100 kHz each.

In order to guarantee sample-exact synchronicity the audio sample clock is derived from the MRI scanner's 10 MHz master clock. Furthermore, the audio recording is started and stopped using the radio frequency (rf)-unblank signal of the scanner with the help of some interfacing glue logic. This mechanism is described in detail in the following section.

SYNCHRONIZING HARDWARE

The GE Signa scanner provides a digital 10 MHz master clock signal to its MRI excitation and readout sequencer circuits, which is also available on the scanner's service interface. Furthermore, the scanner allows access to the so-called rf-unblank TTL signal, which is a short low-pulse in the beginning of each MRI acquisition.

The key part of the data acquisition system is the field-programmable gate array (FPGA)-based digital glue logic that interfaces the MRI scanner hardware to the audio analog-to-digital converter (ADC) on the NI-DAQ card. The logic circuitry was implemented on a DIGILAB 2 XL board,[7] which contains a XILINX Spartan two FPGA.[8]

The digital glue logic consists of two independent systems, namely a clock divider and a retriggerable monostable. The clock divider derives a 200 kHz clock signal from the 10 MHz master clock, which is used to clock the ADC on the NI-DAQ card, resulting in a sampling rate of 100 kHz for each of the two microphone channels.

The retriggerable monostable vibrator has a time constant which equals the MRI repetition time, TR. The monostable is (re-)triggered on the falling edge of each rf-unblank pulse, i.e., in the beginning of each MRI acquisition.

If a number of MRI acquisitions are performed consecutively a train of rf-unblank low pulses is observed with a time distance of TR. Each rf-unblank pulse retriggers the monostable and keeps its output high during the entire acquisition period. This process is shown in Fig. 2, where we assume a series of three consecutive MRI acquisitions.

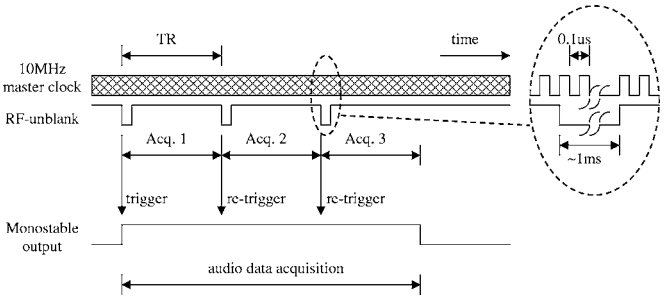


FIG. 2. Glue logic timing diagram.

The output of the monostable is used as an enable signal for the NI-DAQ ADC. This mechanism turns on the analog-to-digital conversion with the first MRI acquisition in a series and stops it as soon as the rf-unblank pulses disappear, i.e., exactly one TR after the last unblank pulse of the acquisition series was observed.

The enable delay of the NI-DAQ card is on the order of 100 ns which is negligible with respect to the audio sample time of 50 μs at 20 kHz. Therefore, the audio recording begins almost exactly when the MRI acquisitions start. And since the ADC sample clock is directly derived from the MRI scanner's 10 MHz clock signal, which governs the image acquisition, the audio and the MRI images are always exactly synchronized.

SOFTWARE COMPONENTS

Data acquisition and sample rate conversion

The real-time data acquisition routine was written in MATLAB[9] and it uses the Data Acquisition Toolbox. In the first postprocessing step, low-pass filtering and decimation of the audio data to a sampling frequency of 20 kHz is carried out. Finally, the processed audio is merged with the reconstructed MRI image sequence using the VirtualDub software.[10]

Noise cancellation

The proposed hardware setup allows for a variety of noise canceling solutions. We describe two noise cancellation methods that we developed: a direct adaptive cancellation method using the well-known normalized least mean square (NLMS) algorithm, and a novel, model-based adaptive cancellation procedure, which yielded the best results in our speech and singing production experiments.

Figure 3 illustrates the location of the microphones and the main sources of noise in the scan room in the proposed set up, namely the subject, the MRI scanner, and the cryogen pump. The dotted lines symbolize the path of the sound, omitting the reflections on the walls of the scan room: The subject's speech is first of all picked up by the main microphone, but there is also a leakage path to the reference microphone. The MRI gradient noise is picked up by both the main and the reference microphone through different paths and, hence, with different time delays and different filtering, but with similar intensity. Last, the cryogen pump noise affects mainly the reference channel.

The GE Signa scanner also has an integrated cooling fan which produces some air flow through the bore of the mag-

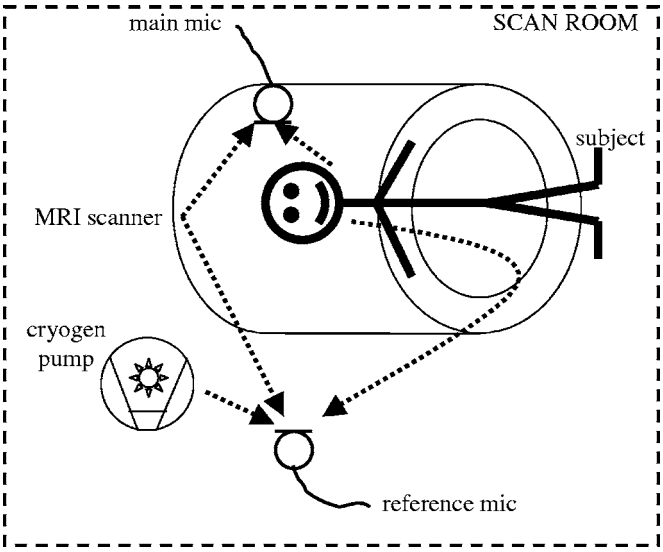


FIG. 3. Noise sources and microphone arrangement.

net. The fan may also produce additional noise but can be turned off during the scan. In our experiments, however, we found the fan noise negligible.

It should be noted that the MRI gradient noise is by far the strongest of all noise sources. But despite its high power, it also has some advantageous characteristics, namely it is stationary, periodic, and directly dependent on the MRI pulse sequence. In our case we used a 13-interleaf spiral gradient-echo sequence with an echo time TE of 0.9 ms, and a repetition time TR of 6.856 ms, which results in a period of 89.12 ms. This means that the scan noise can be thought of as a periodic function with a fundamental frequency of 11.22 Hz. As will be shown below, this characteristic can be exploited to achieve very good noise cancellation results within a modeled-reference framework.

Direct NLMS noise cancellation

In order to overcome the above-mentioned limitations, a noise cancellation procedure was developed which is based on the well-known NLMS algorithm.[11,12] The corresponding system diagram is shown in Fig. 4: The MRI gradient noise is assumed to be filtered by two independent linear systems H_1 and H_2, which represent the acoustic characteristics of the room, before it enters the main and reference channel microphones, respectively. The speech signal on the other hand is captured directly by the main channel microphone.

During the postprocessing, the reference signal is fed into an adaptive finite-impulse response (FIR) filter, and subsequently subtracted from the main channel. The NLMS algorithm continually adjusts the FIR filter coefficients in such

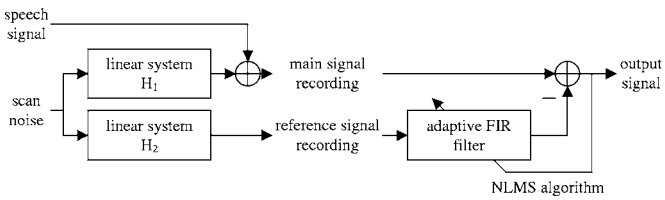


FIG. 4. Adaptive FIR filter using NLMS algorithm for direct cancellation of interference from MRI scanner noise.

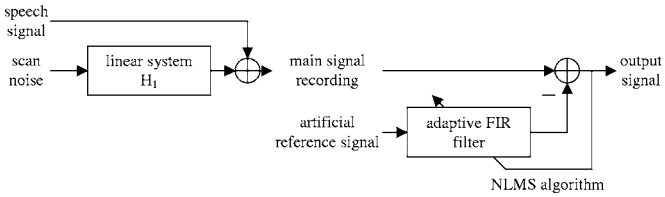


FIG. 5. Adaptive FIR filter using NLMS algorithm for model-based cancellation of interference from MRI scanner noise.

a way that the average output signal power is minimized. Or, in other words, the adaptive FIR filter is continuously adjusted in a way that it best approximates the transfer function H_1/H_2.

Since the noise cancellation is done off-line in our setup, the FIR filter is allowed to be noncausal and the noise cancellation can be achieved regardless whether the time delay between main and reference channel is positive or negative.

The adaptive FIR filter in our case was of order 4000, and the sampling frequency was 20 kHz. The updating coefficient was set to 0.5. The achieved noise reduction was around 17 dB. Further details are provided in the Results section.

Model-based NLMS noise cancellation

A much improved noise reduction was achieved using an artificial reference signal, which is generated based on a pulse-sequence specific model for the MRI gradient noise, rather than the reference signal captured during the scan. The corresponding system diagram is shown in Fig. 5. Hereby we exploit the periodic nature of the gradient noise, and we generate a signal consisting of the sum of unity-amplitude sinusoids of the fundamental frequency of the MRI scan noise, e.g., 11.22 Hz, and all integer multiples up to half the audio sampling frequency: 11.22, 22.44, … ,9996.86 Hz. Hence, this signal contains all spectral components that the periodic gradient noise wave form can possibly have in the audio frequency band. The signal now serves as the reference for an NLMS noise canceller with a FIR filter of order 4000, with an updating coefficient of 0.5. The achieved noise suppression was around 32 dB, and details are provided in the Results section.

The disadvantage of this model-based procedure is that it does not account for other noise sources than the MRI scanner, such as the cryogen pump. The major advantage of this approach, however, is that there is no leakage of the desired signal, i.e., the subject's speech, into the reference channel. Though it might be possible to find a more accurate reference signal model which also includes the cryogen pump, we found that the cancellation of the MRI gradient noise alone provides an output signal with sufficient quality for further analysis.

Another advantage of the model-based procedure is that it lends itself to real-time implementations since even noncausal noise canceling FIR filters are implementable because the modeled reference signal is deterministic.

RESULTS

In order to quantify the effectiveness of the noise cancellation algorithms a 30 s silence recording was obtained,

TABLE I. Noise power suppression for the two presented methods during no speech.

	Noise power suppression in silence recording		
	Unweighted (dB)	A-weighted (dB)	ITU-R468 (1 kHz) (dB)
Direct NLMS	17.1	17.6	16.3
Model-based NLMS	32.8	31.1	32.7

i.e., without any speech activity, and the average output signal power was measured. Table I summarizes the achieved noise suppression for unweighted, A-weighted,[13,14] and ITU-R 468 (1 kHz) weighted output power measurements.[15,16]

The verification of the noise canceller for recordings with speech and/or singing is more difficult since one cannot simply separate the signal and the noise in the recordings and measure their power independently. However, an estimate of the signal to noise ratio (SNR) was obtained by measuring the signal power during speech periods, $P_{\text{speech+noise}}$, and scan noise-only periods, P_{noise}, for a given recording. Due to the stationarity of the noise, and the independence of the noise and speech processes, we can compute the signal power as $P_{\text{speech}}=P_{\text{speech+noise}}-P_{\text{noise}}$. The SNR for the given recording can now be expressed as $\text{SNR}=P_{\text{speech}}/P_{\text{noise}}=(P_{\text{speech+noise}}-P_{\text{noise}})/P_{\text{noise}}$. This computation was carried out for the original main channel recording, the direct noise-cancelled output, and the model-based noise-cancelled output. The improvements in SNR with respect to the original recording are summarized in Table II. The corresponding signal wave forms are shown in Fig. 6. Here we see the main channel recording, the directly noise-cancelled output, the model-based noise-cancelled output, and the voice activity flag of the sample utterance "We look forward to your abstracts by December 19th. Happy holidays! [singing]."

Furthermore, we observed a slight echo-like artifact in the audio output signal most likely believed to result from the following: After convergence (say in a no-speech period), the adaptive noise canceller acts like a comb filter and effectively nulls out all frequencies that are integer multiples of the gradient noise fundamental. If now suddenly a speech signal appears, which generally has energy at those frequencies, the noise-canceling filter will take some time to adapt and again block out these frequencies. When the speech segment is over, the filter again needs a short time to converge back to the no-speech setting. During this time the audio output obviously contains a residue of the reference signal causing a reverberant effect.

TABLE II. Noise power suppression for the two presented methods during speech.

	Noise power suppression in speech recording		
	Unweighted (dB)	A-weighted (dB)	ITU-R468 (1 kHz) (dB)
Direct NLMS	17.2	18.5	13.3
Model-based NLMS	28.4	29.7	26.5

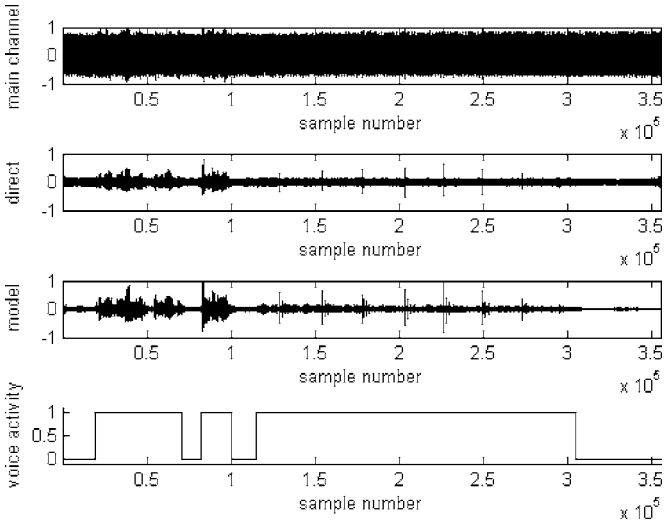


FIG. 6. Sample wave forms for SNR estimation.

As a possible remedy for this effect, one can make the adaptation of the filter dependent on voice activity, such that during the no-speech phases the filter adapts fast, whereas during speech phases the adaptation is slow, or even turned off completely.

Various video and audio clips are provided on the web at ⟨http://sail.usc.edu/span/jasa_letter/index.php⟩ to demonstrate the achieved signal quality. Overall, our model-based NLMS procedure appears to be most attractive since it achieves good noise suppression, requires only a single microphone and is easily implementable. Future improvements may include an improvement of the model to include the cryogen pump.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. R01 DC007124-01. The authors thank their team members especially D. Byrd and S. Lee, as well as USC's Imaging Science Center.

[1]M. Stone, E. Davis, A. Douglas, M. NessAiver, R. Gullapalli, W. Levine, and A. Lundberg, "Modelling the motion of the internal tongue from tagged cine-MRI images," J. Acoust. Soc. Am. **109**, 2974–2982 (2001).
[2]S. Narayanan, K. Nayak, A. Sethy, and D. Byrd, "An approach to real-time magnetic resonance imaging for speech production," J. Acoust. Soc. Am. **115**, 1771–1776 (2004).
[3]M. NessAiver, M. Stone, V. Parthasarathy, Y. Kahana, and A. Paritsky, "Recording high quality speech during tagged Cine MRI studies using a fiber optic microphone," J. Magn. Reson Imaging **23**, 92–97 (2006).
[4]http://phone-or.com, last seen 02/28/2006.
[5]Y. Kahana, A. Paritsky, A. Kots, and S. Mican, "Recent advances in optical microphone technology," Proc. of the 32nd International Congress and Exposition on Noise Control Engineering 2003.
[6]http://www.ni.com, last seen 02/28/2006.
[7]http://www.digilentinc.com, last seen 02/28/2006.
[8]http://www.xilinx.com, last seen 02/28/2006.
[9]http://www.themathworks.com, last seen 02/28/2006.
[10]http://www.virtualdub.org, last seen 02/28/2006.
[11]S. Haykin, *Adaptive Filter Theory* (Prentice Hall, Upper Saddle River, NJ, 2001).
[12]D. Jones, "Normalized LMS," http://cnx.rice.edu/content/m11915/latest/, last seen 02/28/2006.
[13]http://en.wikipedia.org/wiki/A-weighting, last seen 02/28/2006.
[14]IEC 179 standard available at http://www.iec.ch/.
[15]http://en.wikipedia.org/wiki/Standard:ITU-R-468, last seen 06/12/06.
[16]ITU-R 468 standard available at http://www.itu.int/.

The 0/0 problem in the fuzzy-logical model of perception (L)

Jean-Luc Schwartz[a)]
Institut de la Communication Parlée, CNRS UMR 5009, INPG-Université Stendhal INPG, 46 Av. Félix Viallet, 38031 Grenoble Cedex 1, France

(Received 25 April 2005; revised 6 July 2006; accepted 6 July 2006)

The "Fuzzy-Logical Model of Perception" (FLMP) has often been questioned for its presumed ability to fit any data, but no clear-cut evidence has been presented yet. This paper demonstrates the ability of the FLMP to fit random data in the "McGurk region," that is in conditions involving conflicting stimuli. This is due to the so-called "0/0 problem," consisting in the fact that any audio-visual response can be fitted by the FLMP if the audio and visual stimuli provide at least one null probability in each possible category. The consequence is a high instability of the root mean square error in the region of the best fit. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2258814]

PACS number(s): 43.71.An [KWG] Pages: 1795–1798

I. INTRODUCTION

Since the middle of the 1970s, Massaro and his colleagues have extensively studied the Fuzzy-Logical Model of Perception (FLMP) (Massaro, 1987), first in the context of categorical perception and then, since the emergence of the "McGurk" paradigm (McGurk and MacDonald, 1976), in the context of audio-visual (*AV*) interactions and fusion in speech perception (Massaro, 1998). A systematic assessment campaign of quantitative models, comparing the FLMP with various other competitors, led some researchers to suspect a "high flexibility" of the FLMP, enabling it to fit a too large set of data (e.g., Cutting *et al.*, 1992; Grant and Seitz, 1998). However, the demonstration has never been really conclusive (Massaro and Cohen, 1993, 2000; Dunn, 2000). More recently, the debate switched toward tools for comparing models, in the "Bayesian Model Selection" (BMS) framework (Myung and Pitt, 1997; Massaro *et al.*, 2001; Pitt *et al.*, 2003).

This paper discusses a technical problem with the FLMP, particularly acute in the context of a major experimental paradigm for multisensory interactions, that is the perception of conflicting stimuli—including the most well-known situation provided by the McGurk effect. This may lead to questioning some of the model comparison benchmarks using fit as a basic model assessment tool, rather than BMS criteria available in the literature.

II. A TECHNICAL ANALYSIS OF THE FLMP IN THE MCGURK PARADIGM

A. FLMP and the 0/0 problem

In a speech perception task involving the categorization of auditory, visual, and audio-visual stimuli, the FLMP may be defined as a Bayesian fusion model with independence between modalities, and the basic FLMP equation is

$$P_{AV}(C_i) = P_A(C_i)P_V(C_i)/\Sigma_j P_A(C_j)P_V(C_j). \quad (1)$$

C_i and C_j are phonetic categories involved in the experiment, and P_A, P_V, and P_{AV} are the model probability of responses, respectively, in the *A*, *V*, and *AV* conditions (observed probabilities are in lower case and simulated probabilities in upper case throughout this paper). In most papers comparing models in the field of speech perception, the tool used to compare models is the "fit" estimated by the "root mean square error" (RMSE), computed by taking the squared distances between observed and predicted probabilities of responses, averaging them over all categories C_i (total number n_C) and all experimental conditions E_j (total number n_E), and taking the square root of the result:

$$\mathrm{RMSE} = \left[\left(\sum_{E_j, C_i} (P_{E_j}(C_i) - p_{E_j}(C_i))^2 \right) \Big/ (n_E n_C) \right]^{1/2}. \quad (2)$$

In a typical McGurk situation with audio [*b*] plus video [*g*], the unimodal responses are almost incompatible, and hence all phonetic categories involved in the pattern of responses display at least one very low value, either in the *A* modality, or in the *V* modality, or in both. The consequence is that all terms $P_A(C_i)P_V(C_i)$ are likely to be close to zero for all involved categories. To take an example, consider what would happen in an extreme situation with two phonetic classes C1 and C2, and a pair of *A* and *V* stimuli perfectly conflicting, that is with 100% of C1 responses with the *A* stimulus, and 100% of C2 responses with the *V* stimulus (see Table I). Then, it is easy to show that any response in the AV modality, with a probability of C1 response equal to x and a probability of C2 response equal to $(1-x)$ (x being any value between 0 and 1) can be fitted by the FLMP with a RMSE value exactly equal to 0. Indeed, suppose that the corresponding FLMP parameters for C1 and C2 are, respectively, set to $(1-\varepsilon_A)$ and ε_A in the *A* modality, and to ε_V and $(1-\varepsilon_V)$ in the *V* modality, with ε_A and ε_V two small values. Then the probability of the C1 response in the *AV* condition is given, according to Eq. (1), by

a)Electronic mail: schwartz@icp.inpg.fr

TABLE I. Fitting "perfect" two-category McGurk data with FLMP. Fit is perfect for any value of x, provided that ε_A and ε_V follow Eq. (4), with arbitrary low values.

	C1 responses		C2 responses	
	Data	FLMP	Data	FLMP
A cond	1	$1-\varepsilon_A$	0	ε_A
V cond	0	ε_V	1	$1-\varepsilon_V$
AV cond	x	x	$1-x$	$1-x$

$$\varepsilon_V(1-\varepsilon_A)/[\varepsilon_V(1-\varepsilon_A)+\varepsilon_A(1-\varepsilon_V)] \cong \varepsilon_V/(\varepsilon_V+\varepsilon_A). \quad (3)$$

Hence, x may be perfectly fitted by choosing an $(\varepsilon_A, \varepsilon_V)$ pair such that

$$\varepsilon_V/[\varepsilon_V+\varepsilon_A] = x \Leftrightarrow \varepsilon_V = \varepsilon_A x/(1-x). \quad (4)$$

Then, setting ε_A and ε_V at an arbitrarily low value, provided that they respect Eq. (4), allows a perfect fit (RMSE equal to 0) to the pattern of experimental data in Table I, whatever x.

Exactly the same can be done with a three-class situation more similar to the McGurk effect (both for fusions and combinations). This corresponds to a configuration such as audio [b] (perceived as [b]) and video [g] (perceived as [d] or [g]), for which any response pattern can be perfectly fitted by the FLMP, with a RMSE value equal to 0.

This is due to a simple and well-known mathematical fact: 0/0 is an arbitrary value, or, to state this more precisely, $\lim(x/y)$ when $x \to 0$ and $y \to 0$, if it exists, may be any real value.

B. Application to real McGurk data

Of course, it could be argued that the previous section was not about real data, which seldom produce perfect zero values. However, the McGurk paradigm typically leads to quite similar situations, since it deals with conflicting A and V stimuli. In a study of the McGurk effect in French (Cathiard *et al.*, 2001), the pattern of responses to [b_A], [d_V], [g_V], [$b_A d_V$] and [$b_A g_V$] for 126 French subjects, provided in Table II(a), surprisingly showed that there were less [d] and more [b] responses to [$b_A d_V$] than to [$b_A g_V$]. This pattern, coherent with many other published data, seems difficult to understand on the classical view that "[b_A] is similar to [d_A] and [g_V] is similar to [d_V]." Indeed, replacing [g_V] by [d_V] should obviously not result in a decrease of the [d] score in this reasoning. However, the FLMP performed very well on these data, with a RMSE of 0.0062.[1] But actually, with the same A and V responses, any pattern of AV response can be fitted as well by the FLMP. This is displayed in Table II(b), providing the FLMP fits to hypothetical patterns of AV responses with [$b_A d_V$] and [$b_A g_V$] both perceived as mostly [b] (Response 1), mostly [d] (Response 2), mostly [g] (Response 3), [$b_A d_V$] perceived as [b] and [$b_A g_V$] as [d] (Response 4) or the inverse, [$b_A d_V$] as [d] and [$b_A g_V$] as [b] (Response 5). All the fits are equally good, and as good as the fit of true data: in this McGurk context, FLMP is able to fit everything, even a random pattern of response.

While *fitting both unimodal and multimodal data* by the FLMP in Table II(a) provides excellent results (with RMSE as low as 0.0062), *predicting audiovisual from auditory and visual data* provides a dramatic RMSE increase up to 0.116, hence 20 times more. This is due to the fact that the FLMP ability to fit any pattern in this region has a severe drawback: the fit is highly unstable, hence the difference between the "fitting all" and the "prediction" strategies. Therefore, very small variations (±0.01) applied to each FLMP parameter around the best fit to the experimental data in Table II(a) may lead to dramatic changes from the almost perfect value RMSE=0.0062 to values as high as 0.25. The difference between a "fitting all" and a "prediction" strategy has been discussed in detail by Massaro (1998, Chap. 10), and the "fitting all" technique, called "variable FLMP," is sometimes discarded in benchmarks, because of its presumed overfitting ability (e.g., Grant *et al.*, 1998), though Massaro considers it to be the only valid one.[2]

TABLE II. Fitting experimental McGurk data by the FLMP. (a) McGurk data for 126 French subjects obtained by Cathiard *et al.* (2001); (b) RMSE obtained when using the FLMP to fit the data in (a), or arbitrary AV responses 1, 2, 3, 4, 5 with the same A and V unimodal data (see the text).

(a)	Responses	[b]	[d]	[g]	Other
	[b_A]	0.98	0	0	0.02
	[d_V]	0.005	0.88	0.06	0.055
	[g_V]	0	0.125	0.845	0.03
	[$b_A d_V$]	0.835	0.095	0	0.07
	[$b_A g_V$]	0.68	0.23	0.02	0.07

(b)		Answer [b]	Answer [d]	Answer [g]	RMSE
True response	[$b_A d_V$]	0.835	0.095	0	0.0062
	[$b_A g_V$]	0.68	0.23	0.02	
Response 1	[$b_A d_V$]	0.9	0.1	0	0.0049
	[$b_A g_V$]	0.9	0.1	0	
Response 2	[$b_A d_V$]	0.1	0.9	0	0.0053
	[$b_A g_V$]	0.1	0.9	0	
Response 3	[$b_A d_V$]	0.1	0	0.9	0.0061
	[$b_A g_V$]	0.1	0	0.9	
Response 4	[$b_A d_V$]	0.9	0.1	0	0.0082
	[$b_A g_V$]	0.1	0.9	0	
Response 5	[$b_A d_V$]	0.1	0.9	0	0.0047
	[$b_A g_V$]	0.9	0.1	0	

III. A BAYESIAN-MODEL-SELECTION SOLUTION TO THE 0/0 PROBLEM

The 0/0 problem may result in a difficulty in testing FLMP in light of McGurk data (e.g., Cathiard *et al.*, 2001; Tiippana *et al.*, 2004). A possible solution could be to discard McGurk data from model assessment studies, but this seems odd in light of the many facts discovered about audio-visual speech perception by studying this paradigm through both behavioral (e.g., Green, 1998) and, more recently, neurophysiological (e.g., Jones and Callan, 2003; Sekiyama *et al.*, 2003) paradigms.

The solution recommended by Massaro (1998) is to use large data sets rather than restricted McGurk data to assess and compare models. However, the benchmark set he used in a number of studies (e.g., Massaro, 1998), crossing a synthetic five-level audio [*ba*]-[*da*] continuum with a synthetic

video similar continuum (http://mambo.ucsc.edu/ps1/8236/), does contain typical conflicting configurations involving the 0/0 problem. Therefore, this problem actually interferes with the results of the performed benchmark studies. The question is to know how much it interferes, and the answer is not at all obvious. Hence, our interest in another tool for comparing models, namely Bayesian Model Selection.

A. The Bayesian framework for model assessment

The fit may be derived from the logarithm of the *maximum likelihood of a model*, considering a data set. If $\mathbf{D}$ is a set of k data d_i, and M a model with parameters $\mathbf{\Theta}$, $L(\mathbf{\Theta}|M)$ is the likelihood of parameters $\mathbf{\Theta}$ for the model, considering the data:

$$\mathbf{L}(\mathbf{\Theta}|M) = p(\mathbf{D}|\mathbf{\Theta},M). \tag{5}$$

The $\boldsymbol{\theta}$ parameters maximizing the likelihood of M are provided by[3]

$$\boldsymbol{\theta} = \text{argmax}\, L(\mathbf{\Theta}|M) \tag{6}$$

and it is possible to show that maximizing likelihood is not very different from minimizing RMSE, that is searching the best fit to the experimental data. However, comparing two models by comparing their best fits means that there is a first step of estimation of these best fits, and it must be acknowledged that the estimation process is not error-free. Therefore, the comparison must account for this error-prone process, which is done in BMS by computing the total likelihood of the model knowing the data. This results in integrating likelihood over all model parameter values:

$$p(\mathbf{D}|M) = \int p(\mathbf{D},\mathbf{\Theta}|M)d\mathbf{\Theta} = \int p(\mathbf{D}|\mathbf{\Theta},M)p(\mathbf{\Theta}|M)d\mathbf{\Theta}$$

$$= \int L(\mathbf{\Theta}|M)p(\mathbf{\Theta}|M)d\mathbf{\Theta}. \tag{7}$$

Taking the opposite of the logarithm of total likelihood leads to the so-called "Bayesian Model Selection" (BMS) criterion that should be minimized for model evaluation (MacKay, 1992; Pitt and Myung, 2002):

$$\text{BMS} = -\log \int L(\mathbf{\Theta}|M)p(\mathbf{\Theta}|M)d\mathbf{\Theta}. \tag{8}$$

B. How BMS integrates fit and stability

The integral in Eq. (7) means that the total likelihood of a model knowing the data evaluates the volume of $\mathbf{\Theta}$ values providing an "acceptable" fit (not too far from the best one) relative to the whole volume of possible $\mathbf{\Theta}$ values. This relative volume decreases if the function $L(\mathbf{\Theta}|M)$ decreases too quickly around its maximum value $L(\boldsymbol{\theta}|M)$: this is what happens if the model is too sensitive, as is the FLMP around its best fit in the McGurk region.

The difference between best fit and global likelihood is illustrated in Fig. 1. This figure deals with a two-category audio-visual experiment with one audio condition, one visual condition, and one audio-visual condition combining the audio and the visual stimuli. A FLMP simulation of this experiment needs two free parameters, that is the audio probability P_A and the visual probability P_V of the first category, the probabilities of the second category being $1-P_A$ and $1-P_V$. Three experimental configurations are considered in the figure.

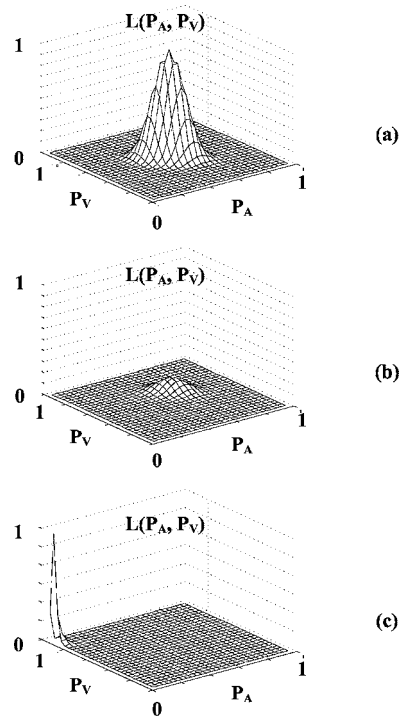


FIG. 1. Likelihood distributions in two- category three-condition configurations (a) $p_A=p_V=p_{AV}=0.5$. (b) $p_A=p_V=0.5$, $p_{AV}=0.8$. (c) $p_A\approx 0$, $p_V\approx 1$, p_{AV} arbitrary (any value between 0 and 1).

In the first one, the data provide an ambiguous perception both in the A, V, and AV modalities ($p_A=p_V=p_{AV}=0.5$). In Fig. 1(a), the distribution $L((P_A,P_V)|\text{FLMP}) = p(\mathbf{D}|(P_A,P_V),\text{FLMP})$ is a smooth curve peaking at the point (0.5, 0.5) in the (P_A,P_V) plane. This peak provides the maximum likelihood, which is high, since the data are compatible with the FLMP prediction. The smooth behavior of the likelihood function leads to a large value of the total integral below the surface, which is precisely the global FLMP likelihood for these data.

In the second configuration, the A and V percepts are still ambiguous ($p_A=p_V=0.5$), but the AV percept is not ($p_{AV}=0.8$). This is of course contradictory to the FLMP prediction, hence the peak of the likelihood distribution in Fig. 1(b) is very low, and both best likelihood and global likelihood are small. Altogether, Figs. 1(a) and 1(b) illustrate the case of a strong FLMP prediction in which fit RMSE and global likelihood BMS provide the same kind of behavior (both good or both poor depending on the coherence of A, V, and AV data).

The third configuration is typical of the McGurk effect, with conflicting A and V stimuli and any AV response (that is, the AV probability of the first category p_{AV} can be any value between 0 and 1). Figure 1(c) shows that here, the FLMP maximum likelihood is quite high, but the likelihood distribution is not smooth at all. The reason is that to be able to fit everything in this case, the region around $(P_A=0,P_V$

$=1$) must be divided into an infinity of small subregions able to predict any value of p_{AV}. As a consequence, the integral below the likelihood curve is small. A high maximum likelihood (or a small RMSE) together with a small global likelihood: this is typically an overfitting configuration, with no prediction ability at all. In such kinds of configurations, FLMP provides a better fit than almost any other model while global likelihood naturally combines fit and stability into an integrated measure, making model assessment sounder. A detailed implementation of the so-called Laplace approximation of BMS together with its conditions of use is provided in http://www.icp.inpg.fr.

IV. CONCLUSION

The 0/0 problem raises a difficulty in model fitting based on RMSE criteria for comparing FLMP with other models on conflicting stimuli. BMS appear as a better model comparison technique in this case, though RMSE may remain an interesting additional criterion enabling one to assess the quality of the fit, apart from model comparison per se. We suggest that BMS could be of great interest in future model comparison studies in the AV speech perception domain.

[1]RMSE in Cathiard *et al.* (2001) was slightly larger, because of the use of a threshold on the minimal acceptable probabilities, not used here, apart from the classical constraint that a probability is within [0-1].

[2]Massaro (1998, Chap. 10) discusses in detail the relationship between "prediction" and what he calls "post-diction," that is the "fitting all" procedure. The proposal he makes for connecting these is based on so-called "benchmark goodness of fit." However, this technique, in which the data are varied according to the observed data statistics, suffers from exactly the same problem: in McGurk cases, any variation of the data can be perfectly fit by FLMP.

[3]In the following, bold symbols deal with vectors or matrices, and all maximizations are computed on the model parameter set $\boldsymbol{\Theta}$.

Cathiard, M. A., Schwartz, J. L., and Abry, C. (**2001**). "Asking a naive question to the McGurk effect: Why does audio [*b*] give more [*d*] percepts with visual [*g*] than with visual [*d*]?," Proceedings of AVSP'2001, pp. 138–142.

Cutting, J. E., Brady, N. P., Bruno, N., and Moore, C. (**1992**). "Selectivity, scope, and simplicity of models: A lesson from fitting judgments of perceived depth," J. Exp. Psychol. **121**, 364–381.

Dunn, J. C. (**2000**). "Model complexity: The fit to random data reconsidered," Psychol. Res. **63**, 174–182.

Grant, K. W., and Seitz, P. F. (**1998**). "Measures of auditory-visual integration in nonsense syllables and sentences," J. Acoust. Soc. Am. **104**, 2438–2450.

Grant, K. W., Walden, B. E., and Seitz, P. F. (**1998**). "Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition and auditory-visual integration," J. Acoust. Soc. Am. **103**, 2677–2690.

Green, K. P. (**1998**). "The use of auditory and visual information during phonetic processing: Implications for theories of speech perception," in *Hearing by Eye II*, edited by R. Campbell, B. Dodd, and D. Burnham (Psychology Press, Hove, UK), pp. 3–25.

Jones, J. A., and Callan, D. E. (**2003**). "Brain activation during an audiovisual speech perception task: An fMRI study of the McGurk effect," NeuroReport **14**, 1129–1133.

MacKay, D. J. C. (**1992**). "Bayesian interpolation," Neural Comput. **4**, 415–447.

Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Laurence Erlbaum Associates, London).

Massaro, D. W. (**1998**). *Perceiving Talking Faces* (MIT, Cambridge).

Massaro, D. W., and Cohen, M. M. (**1993**). "The Paradigm and the Fuzzy Logical Model of Perception are alive and well," J. Exp. Psychol. **122**, 115–124.

Massaro, D. W., and Cohen, M. M. (**2000**). "Tests of auditory-visual integration efficiency within the framework of the fuzzy-logical model of perception," J. Acoust. Soc. Am. **108**, 784–789.

Massaro, D. W., Cohen, M. M., Campbell, C. S., and Rodriguez, T. (**2001**). "Bayes factor of model selection validates FLMP," Psychonomic Bulletin & Review **8**, 1–17.

McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," Nature (London) **264**, 746–748.

Myung, I. J., and Pitt, M. A. (**1997**). "Applying Occam's razor in modeling cognition: A Bayesian approach," Psychonomic Bulletin & Review **4**, 79–95.

Pitt, M. A., Kim, W., and Myung, I. J. (**2003**). "Flexibility versus generalizability in model selection," Psychonomic Bulletin & Review **10**, 29–44.

Pitt, M. A., and Myung, I. J. (**2002**). "When a good fit can be bad," Trends in Cognitive Science **6**, 421–425.

Sekiyama, K., Kanno, I., Miura, S., and Sugita, Y. (**2003**). "Auditory-visual speech perception examined by fMRI and PET," Neurosci. Res. (N Y) **47**, 277–287.

Tiippana, K., Andersen, T. S., and Sams, M. (**2004**). "Visual attention modulates audiovisual speech perception," European Journal of Cognitive Psychology **16**, 457–472.

Children hear the forest (L)

Susan Nittrouer[a)]
Ohio State University

(Received 21 March 2006; revised 22 June 2006; accepted 8 July 2006)

How do children develop the ability to recognize phonetic structure in their native language with the accuracy and efficiency of adults? In particular, how do children learn what information in speech signals is relevant to linguistic structure in their native language, and what information is not? These questions are the focus of considerable investigation, including several studies by Catherine Mayo and Alice Turk. In a proposed Letter by Mayo and Turk, the comparative role of the isolated consonant-vowel formant transition in children's and adults' speech perception was questioned. Although Mayo and Turk ultimately decided to withdraw their letter, this note, originally written as a reply to their letter, was retained. It highlights the fact that the isolated formant transition must be viewed as part of a more global aspect of structure in the acoustic speech stream, one that arises from the rather slowly changing adjustments made in vocal-tract geometry. Only by maintaining this perspective of acoustic speech structure can we ensure that we design stimuli that provide valid tests of our hypotheses and interpret results in a meaningful way.
[DOI: 10.1121/1.2335273]

PACS number(s): 43.71.Ft [MSS] Pages: 1799–1802

Once upon a time, the belief was common that phonetic segments line up in the acoustic speech stream like so many trees planted neatly in a row. Accordingly, great effort was put forth to describe these discrete elements in linguistic and acoustic terms. Several notions that continue to be cornerstones of our collective worldview were developed, such as the idea that phonetic segments can be described using linguistic features. So, we all agreed, a segment can be distinctively plus or minus voiced, tense or lax, rounded or unrounded, and so on (e.g., Chomsky and Halle, 1968). Many believed that the phonetic element and/or linguistic feature are appropriately described as collections of acoustic properties. Therefore, investigators searched for the acoustic correlates of segments or features, as indicated in this passage from Blumstein and Stevens (1981):

> "The major claim of a theory of acoustic invariance is that invariant acoustic properties can be derived directly from the acoustic signal, and these properties correspond to the phonetic dimensions which ultimately form the inventory of speech sounds used in natural language. Such a view provides an explicit characterization of the universal set of features, and, in particular, of the phonetic dimensions delineating natural classes. If it is the case that invariant properties structure phonetic dimensions, then such invariance provides an instantiation of particular feature systems." (pp. 27–28)

This view of the acoustic structure of speech and its relation to linguistic structure has enjoyed popular support for a long time.

Relatedly, the notion of categorical perception (e.g., Studdert-Kennedy, Liberman, Harris, and Cooper, 1970) fit the earlier worldview of what happens when a listener hears a sequence of these presumably discrete elements. Categorical perception was based on the idea that a phonemic category can be defined as a circumscribed range of settings for any one unique acoustic property.[1] According to this view, listeners reliably hear sounds with one of these settings as instances of that phonemic category, and similarly reject as instances of that category sounds with different settings. The effect is so strong, so it was hypothesized, that listeners fail to notice acoustic variability across the range of settings on the property that correspond to any one category. Thus, phonemes could effectively be defined as ranges of settings on specific acoustic dimensions. Even the concept of coarticulation rests upon the concept of linear representation of phonetic segments: The acoustic structure of any one segment can be influenced by surrounding phonetic elements, according to traditional accounts of coarticulation. This idea is based on the premise that there must be canonical settings for each phonetic segment. Overall, these notions of categoricalness and linear representation in the acoustic speech signal give us solace and form the framework of how speech is generally thought to be produced and perceived. These ideas have served as the basis of experimental design and theory building for a large body of research, including everything from speech recognition to dyslexia. Attempts to identify the acoustic properties that define phonemic categories (frequently referred to as "cues") have played central roles in the fields of psychology and linguistics for decades, and continue to do so.

But several events over the last 25 years have shaken the foundations of this worldview for some of us. In 1981, four scientists published a paper that would cause us to question our most basic beliefs. Remez, Rubin, Pisoni, and Carrell (1981) showed us that listeners can recover linguistic structure when all of the traditional cues to phonemic identity that we held so dear were eliminated from the signal. These investigators synthesized signals consisting only of the dy-

[a)]Electronic mail: nittrouer.1@osu.edu

namic spectral patterns of speech, and yet listeners were able to understand those signals. At the same time, engineers were designing devices that could be implanted in the cochleas of individuals with profound hearing loss to stimulate the auditory nerve directly. Of course, without the hair cells to provide frequency resolution the best these devices could do was to provide a single channel of information about the acoustic wave form of speech, and many of us predicted unmitigated failure for these devices. But in sharp contrast to our predictions, implant users somehow managed to use that horribly impoverished signal to recover linguistic structure. Since the days of those early implants, cochlear implants have improved, but still implant users have access to only a few channels of information. Even at that, the amplitude envelopes of those few channels is primarily what is received, without even the spectral skeleton provided by sine wave replicas of speech. Nonetheless, experiments with deaf and hearing listeners alike demonstrate that these envelopes serve speech recognition well (Smith, Delgutte, and Oxenham, 2002; Zeng *et al.*, 2004). Clearly both the dynamic changes in spectral resonances and the amplitude envelopes of speech can provide sufficient information to support recognition of linguistic units without traditional speech cues.

Simultaneously with the occurrence of these changes in our understanding of what speech perception entails, investigators interested in production were modifying their worldview. In particular, the idea that there are discrete motor commands for individual phonemes was questioned, and instead the notion was hatched that task-specific goals reduce the multiple degrees of freedom inherent in vocal-tract mechanics. According to this perspective, relations among movements of various articulators in time and space instantiate linguistic structures such as stress, syllable cohesion of intervocalic consonants, and even phonetic identity in the resulting signal (e.g., Kelso, Saltzman, and Tuller, 1986). It became clear that the acoustic speech signal was not shaped by discrete actions of individual articulators. Instead, relatively slow movements such as the rise and fall of the jaw are punctuated by the actions of more rapidly moving articulators, such as the lips, but all these actions are coordinated to impose linguistic structure across the length of an utterance.

It was against these theoretical backdrops that the idea of the developmental weighting shift (DWS) for speech perception emerged. Specifically this idea began with the thought to examine in children's perception what was then termed "trading relations." The premise of trading relations was that settings for one acoustic property could trade with settings for another acoustic property in the phonetic judgments of listeners. For example, Mann and Repp (1980) found that the frequency of a fricative noise that supported a [s] judgment could be lower if the vowel formant frequencies were appropriate for a more apical, rather than velar, place of constriction. The original question posed by Michael Studdert-Kennedy and this author (Nittrouer and Studdert-Kennedy, 1987) was "Would children show this same shift in acceptable settings for one property, based on settings of another property?" Fricative-vowel stimuli similar to those of Mann and Repp were used in this initial exploration of the question, and the results were dramatic. Not only did children show the trading relation described by Mann and Repp, but their responses seemed to be even more strongly influenced by formant transitions than those of adults.

At the time and under the rubric of "trading relations," fricative noise frequency and dynamic formant transitions were viewed as being equal in kind and similar in nature. Each property was one separate bit of the acoustic fiber, a discrete clue that could tell the listener what that temporally discrete element, the fricative, was. Over time, however, the significance of the changes in perspective of speech perception and production described above became apparent: The consonant-vowel (CV) formant transition is not equal in kind to the other acoustic properties that we, as a field, have generally termed "cues." Instead, the CV (or VC) formant transition is one short piece of the larger, continuously changing spectral pattern that arises from the relatively slow modifications made to overall vocal-tract posture. It is these slow changes that children first notice in the speech around them. Gradually, through experience with a native language, children discover the other acoustic properties that are relevant for phonetic identity in their native language. It is not that adults cannot and do not use the dynamic components of the speech signal to recover linguistic structure. Clearly they do. Results with sine wave speech (e.g., Remez *et al.*, 1981) illustrate that fact. Thus, adults are sensitive to both the global spectral structure that arises from the relatively slow modulations of the vocal tract, as well as to the acoustic details imposed by articulatory factors such as the precise shape of a fricative constriction or the exact timing between two gestures. Children, on the other hand, attend primarily to acoustic changes that arise from the slow modulations of the vocal tract, learning about the phonetic significance of details of the signal only as they gain experience with their native language.

This suggestion regarding the development of speech perception skills is supported by findings for perceptual development in general: There is evidence that children glean the global structure of sensory input before discovering the details. For example, Kimchi, Hadad, Behrmann, and Palmer (2005) presented sets of visual patterns that either matched on global structure, but not on local structure, or vice versa to adults and children (ages 5 to 14 years). Participants were alternately asked to judge similarity based on global or local structure. Results showed that children performed as well as adults when asked to judge similarity based on global structure, but errors increased with diminishing age when the task was to judge local structure.

The suggestion that children initially focus their perceptual attention on the acoustic consequences of relatively slow vocal-tract movements finds support from studies of the development of speech production, as well. These global signal components most consistently provide information about places of constriction. In reviewing data from a report by Vihman (1996), Studdert-Kennedy (2000) shows that young children rarely make errors of place in their productions. One of children's earliest accomplishments involves learning to produce the more general movements of the vocal tract, and that includes being able to get from one constriction that is appropriate in their native language to another. Only later do

they learn to refine shapes for consonant constrictions and to precisely coordinate timing among various gestures. Work by de Boysson-Bardies, Sagart, Halle and Durand (1986) provides another, elegant illustration of the fact that children initially attend to the overall changes in vocal-tract geometry of those around them. These investigators computed the long-term spectra of adults and 10-month-olds whose native languages were French, Cantonese, or Algerian. The resulting spectra clearly showed that the infants' spectra already resembled those of the adults they hear speaking everyday. What this means is that these children were learning about the large postural adjustments that are typical of the languages they would come to speak, adjustments having to do with the larynx, the pharynx, and the velum. Again, this perspective of speech development has correlates in another developmental literature: Thelen (e.g., 1985) has shown that children initially acquire a general leg motion for walking, without differentiation of the individual joints. Only later do they acquire independent control over the actions of the hips, knees, and ankles.

In summary, the theoretical perspective being offered here is that children initially attend strongly to the global changes in the acoustic speech signal arising from relatively slow modifications in vocal tract postures. This perspective has explanatory power for many earlier findings. For example, we would not expect, and have not found, that adults and children differ in their labeling of stimuli when dynamic signal components are pretty much all that are available. An example of this situation is provided by the English weak fricatives [θ] and [f]. These fricatives differ very little from each other in noise spectra; accordingly, Harris (1958) found that adults weight formant transitions strongly in decisions regarding their identity. Nittrouer (2002) hypothesized that under these conditions adults and children would weight similarly fricative noise spectra and formant transitions, and results supported that position: Both adults and children depended largely on formant transitions for their fricative decisions. Another example of a situation in which we would not expect weighting of formant transitions to differ for adults and children is when formant transitions do not differ across the stimuli listeners are being asked to label. Mayo and Turk (2004) constructed continua of stop-vowel stimuli with the same place of constriction across each continuum, but with different voicing characteristics for the stops. Thus, formant trajectories were the same across stimuli. As expected, the primary difference among stimuli was when those formants switched from being excited by an aperiodic to a periodic source. Given the constraint on information conveyed by the dynamic character of formant transitions in this contrast, adults and children did not differ in the extent to which formant transitions contributed to their voicing decisions.

None of what has been written here is meant to support or challenge specific theories of phonology, such as whether phonology is word or phoneme based. Human speech is structured at many levels, and presumably each level has significance in communication. Furthermore, children ultimately need to know about structure at each of these levels for their native language in order to attain adult levels of language competency. Many intriguing questions are left to be explored, such as if children explicitly need to learn about each kind of structure or if some structure is automatically recoverable. The purpose of this Letter has merely been to illustrate that we have only recently recognized the significance of global structure in language processing. From that perspective we realize that the formant transition is a specific instance of general vocal-tract dynamics. In the conduct of speech perception experiments, particularly those involving children, we must be mindful of how stimulus design relates to both global and local signal structure. Such attention will help us avoid designing stimuli that violate natural constraints, as well as help us interpret results more appropriately. In general it is fair to say that while we have been busy arranging and rearranging the details of the speech signal in our experiments, it has been the global structure that children have been noticing. Children initially hear the forest, not the trees.

This Letter was originally composed as a Reply to a Comment written by Catherine Mayo and Alice Turk. Although they withdrew their Comment, the decision was made to let this Letter stand on its own. The author thanks Dr. Mayo and Dr. Turk for their stimulating exchange, and several reviewers, including Susan Rvachew, for their comments. This work was supported by research Grant No. R01 DC000633 from the National Institute on Deafness and Other Communication Disorders.

[1]The terms "phoneme" and "phonetic segment" are generally used to refer to the abstract unit and the physical segment as instantiated in the acoustic signal, respectively. According to the view of speech to be presented here, the distinction becomes rather blurry.

Blumstein, S. E., and Stevens, K. N. (**1981**). "Phonetic features and acoustic invariance in speech," Cognition **10**, 25–32.

Chomsky, N., and Halle, M. (**1968**). *The Sound Pattern of English* (MIT Press, Cambridge, MA).

de Boysson-Bardies, B., Sagart, L., Halle, P., and Durand, C. (**1986**). "Acoustic investigations of cross-linguistic variability in babbling," in *Precursors of Early Speech*, edited by B. Lindblom and R. Zetterström (Stockton Press, New York), pp. 113–126.

Harris, K. S. (**1958**). "Cues for the discrimination of American English fricatives in spoken syllables," Lang Speech **1**, 1–7.

Kelso, J. A. S., Saltzman, E. L., and Tuller, B. (**1986**). "The dynamical perspective on speech production: Data and theory," J. Phonetics **14**, 29–59.

Kimchi, R., Hadad, B., Behrmann, M., and Palmer, S. E. (**2005**). "Microgenesis and ontogenesis of perceptual organization—Evidence from global and local processing of hierarchical patterns," Psychol. Sci. **16**, 282–290.

Mann, V. A., and Repp, B. H. (**1980**). "Influence of vocalic context on perception of the /ʃ/-/s/ distinction," Percept. Psychophys. **28**, 213–228.

Mayo, C., and Turk, A. (**2004**). "Adult-child differences in acoustic cue weighting are influenced by segmental context: children are not always perceptually biased toward transitions," J. Acoust. Soc. Am. **115**, 3184–3194.

Nittrouer, S. (**2002**). "Learning to perceive speech: How fricative perception changes, and how it stays the same," J. Acoust. Soc. Am. **112**, 711–719.

Nittrouer, S., and Studdert-Kennedy, M. (**1987**). "The role of coarticulatory effects in the perception of fricatives by children and adults," J. Speech Hear. Res. **30**, 319–329.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947–949.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.

Studdert-Kennedy, M. (**2000**). "Imitation and emergence of segments," Phonetica **57**, 275–283.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., and Cooper, F. S. (**1970**). "Theoretical notes. Motor theory of speech perception: A reply to Lane's critical review," Psychol. Rev. **77**, 234–249.
Thelen, E. (**1985**). Developmental origins of motor coordination: Leg movements in human infants," Dev. Psychobiol. **18**, 1–22.
Vihman, M. M. (**1996**). *Phonological Development* (Blackwell, Oxford, UK).
Zeng, F. G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y. Y. *et al.* (**2004**). "On the dichotomy in auditory perception between temporal envelope and fine structure cues," J. Acoust. Soc. Am. **116**, 1351–1354.

Sonar gain control in echolocating finless porpoises (*Neophocaena phocaenoides*) in an open water (L)

Songhai Li
Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan, 430072, People's Republic of China and Graduate School of the Chinese Academy of Sciences, Beijing, 100039, People's Republic of China

Ding Wang[a] and Kexiong Wang
Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan, 430072, People's Republic of China

Tomonari Akamatsu
National Research Institute of Fisheries Engineering, Fisheries Research Agency, Hasaki, Kamisu, Ibaraki 314-0408, Japan

(Received 30 November 2005; revised 9 July 2006; accepted 18 July 2006)

Source levels of echolocating free-ranging Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*) were calculated using a range estimated by measuring the time delays of the signals via the surface and bottom reflection paths to the hydrophone, relative to the direct signal. Peak-to-peak source levels for finless porpoise were from 163.7 to 185.6 dB *re*:1 μPa. The source levels are highly range dependent and varied approximately as a function of the one-way transmission loss for signals traveling from the animals to the hydrophone. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2335674]

PACS number(s): 43.80.Ka, 43.80.Lb, 43.80.Nd [WWA] Pages: 1803–1806

I. INTRODUCTION

Odontocetes use echolocation to locate prey and detect their surroundings for navigation and orientation. Dolphins and porpoises have been shown to use echolocation in captivity (Au, 1993). Recently, sonar signals and capabilities were explored in wild dolphins and porpoises (Akamatsu *et al.*, 1998; Lammers *et al.*, 2003; Au *et al.*, 2004; Au and Würsig, 2004; Rasmussen *et al.*, 2002; Li *et al.*, 2005a). The finless porpoise, another odontocete species, has well-developed echolocation skills (Akamatsu *et al.*, 2005a). The species produces a high peak frequency (~125 kHz), narrow 3-dB bandwidth (~20 kHz), and short duration (~68 μs) echolocation signals (Akamatsu *et al.*, 1998; Kamminga *et al.*, 1986; Goold and Jefferson, 2002; Li *et al.*, 2005a).

However, the source level (sound pressure level at 1 m from the source recorded on the acoustic axis) of echolocation signals and its adaptive modulation in finless porpoise was not studied extensively, except for the off-axis signals (Akamatsu *et al.*, 2005b), received sound pressure levels with visually estimated ranges and then the calculated source levels according to those ranges (Akamatsu *et al.*, 2001; Wang *et al.*, 2005). The source level of emitted signals is a key parameter for gaining insight into porpoise sonar capabilities. A good range determination is necessary to calculate the source levels of the signals and evaluate its adaptive modulation; however, it is extremely difficult to obtain an accurate measurement of the range between a recording hydrophone and a moving phonating animal in the wild.

The echolocation signals of finless porpoises often exhibit a multipath reflection structure in waters with shallow, flat bottoms and flat surfaces (Li *et al.*, 2005b). Whenever bottom and surface reflections from a sonar click can be isolated in time from direct signals, the phonating porpoise's range can be determined by using the arrival time difference of the direct, surface, and bottom reflected signals (Li *et al.*, 2005b). Using this method, researchers could estimate the range of the phonating animal fairly accurately (Aubauer *et al.*, 2000; Li *et al.*, 2005b). Aubauer *et al.* (2000) and Thode *et al.* (2002) had used this method successfully to track and localize marine animals in the wild.

In this study, we estimated the animal distance to the hydrophone by using the above mentioned one-hydrophone method (Li *et al.*, 2005b), then calculated the source levels and measured the interclick intervals of echolocation signals from free-ranging Yangtze finless porpoise. The data were used to evaluate the sonar gain control in this species.

II. MATERIALS AND METHODS

A. Study site and subjects

Recordings were made on 22 and 23 June 2004 in Tongling Baiji Seminatural Reserve, on a closed channel between Heyue Islet and Tieban Islet, a narrow old lane of the Yangtze River (Fig. 1) in Tongling, Anhui, China. The channel is 1600 m long and 80 to 220 m wide, with a flat, sandy bottom structure. The depth in the channel ranges from 4 to 8 m, and is about 4 m deep in the study area between C and D (Fig. 1). Throughout the experiment, the water surface of the channel was flat without waves. At the time of this study, there were five porpoises living in the channel supported by a combination of artificial feeding and natural predation.

[a]Author to whom correspondence should be addressed. Electronic mail: wangd@ihb.ac.cn

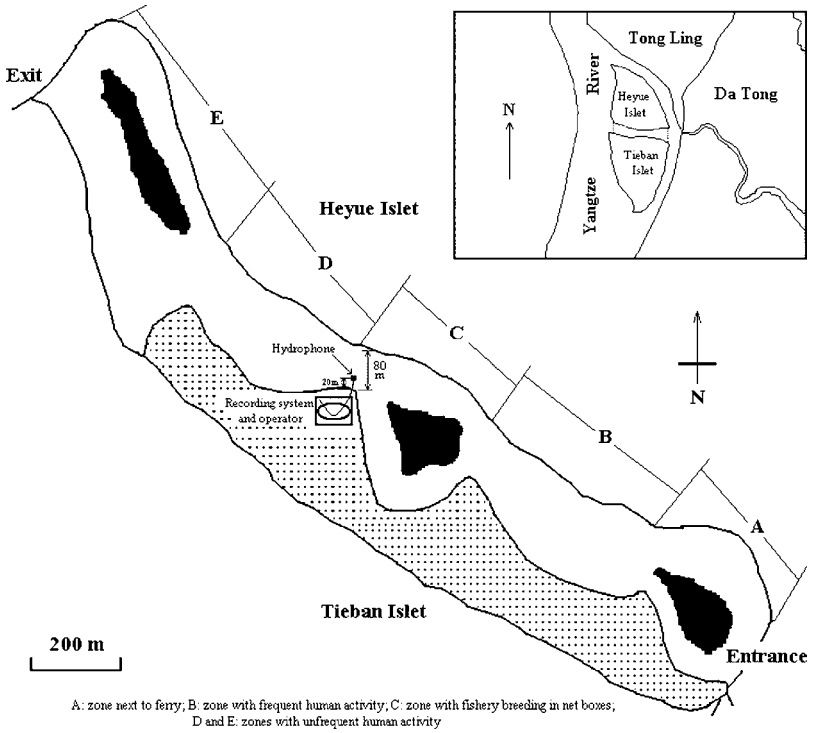


FIG. 1. The study site, Tongling Baiji Semi-Natural Reserve, which is a closed channel between Heyue Islet and Tieban Islet.

B. Experimental setup and data recording

A hydrophone (ST1020, Oki Electric Co. Ltd., Japan) with a built-in preamplifier, was deployed 0.5 m under the water surface in the narrowest section (about 80 m) of the channel (Fig. 1). A 30 m cable connected the hydrophone to an underwater sound level meter (SW1020, Oki) located on the bank of the channel, along with a data recorder (Sony PCHB244). Both upstream and downstream, two sites with linear distances of 25 and 50 m to the hydrophone, respectively, were marked using white-color floats. The floats made it possible to make a visual estimation of animal distances. The visual-estimated distances were used to verify the acoustic-calculated distances. Distances between sites were measured using a flexible measuring tape, and depth was measured using a handheld depth measure (PS-7 LCD DIGITAL SOUNDER; Hondex, Japan).

The hydrophone has a sensitivity of -180 dB *re*: 1 V μPa^{-1} $+3/-12$ dB, up to 150 kHz. The sensitivity declines monotonously with an increase in frequency from 100 to 150 kHz. Recordings were made using a Sony PCHB244 digital data recorder, with a flat frequency response from dc to 147 kHz within 3 dB (Akamatsu *et al.*, 1998). The sound recording system is capable of recording signals with frequencies up to 147 kHz, which is sufficient to receive and store the echolocation signals of the porpoise (see Akamatsu *et al.*, 1998; Li *et al.*, 2005a). A high-pass filter of 10 kHz was incorporated to eliminate low-frequency environmental noise.

An operator on the channel bank manually activated signal recording when finless porpoises were visually spotted. The animals' surface behaviors, abundance, estimated swimming speed, and visually estimated distances to the hydrophone were annotated by the operator using a supplied microphone to the voice channel of the digital data recorder and digitally recorded synchronously with the high-frequency click signals.

C. Data analysis

An analysis of echolocation signals was done using PC-based SIGNAL/RTS™ software (Version 3.0, July 1996; American Engineering Design, USA). The recorded signals were replayed from the data recorder at half-speed and were digitized using a 12-bit Data Translation-2821G A/D board with a sampling frequency of 200 kHz (i.e., an equivalent real time sample rate of 400 kHz). Raw click trains were first reviewed to assess the signal-to-noise ratio. To assure the porpoises detecting the hydrophone, even at the long distances, we always selected the click trains with a high signal-to-noise ratio without regard to the distances between echolocating animals and hydrophone to be analyzed. Then individual clicks acquired manually using the built-in cursor option in these click trains were examined. Only click trains with clicks having very distinct surface and bottom reflections [see Fig. 3(b) in Li *et al.* (2005b)] were included in the subsequent acoustic analysis. Since porpoises emit echolocation clicks directionally and the orientation of a phonating animal cannot be determined with the method presented here, the collected clicks were very likely acquired from both directly on and off the axis of its sonar transmission beam. However, echolocation signals acquired from off the beam axis are distorted, with lower peak frequencies and lower amplitude levels, relative to the source signals (Au, 1993). As a result, only the three clicks with the highest recorded amplitude, highest peak frequency, and smoothest envelope per click train (indicative of distinct surface and bottom reflections, and high signal-to-noise ratio) were analyzed.

An analysis of the selected clicks was performed using both the waveform and spectrum whenever necessary. The received sound pressure level (SPL) of the click and interclick interval (ICI) between the analyzed click and the next click were determined. SPL was calculated as the dB reference to 1 μPa peak to peak, by comparison with a calibration signal of 1 kHz and compensation with +5 dB, which is due to the fact that the sensitivity of the hydrophone near the peak frequency (about 125 kHz; see Li *et al.*, 2005a) of clicks of finless porpoise is about +5 dB lower than that in frequency of 1 kHz. Range (R) of a phonating animal was estimated by using arrival-time differences between the direct pulse, the 180° phase-shifted surface reflection, and the nonphase-shifted bottom reflection (Aubauer *et al.*, 2000; Lammers *et al.*, 2003; Li *et al.*, 2005b). Source level (SL) is defined as the sound pressure level at 1 m from the source. It can be calculated as:

$$\mathrm{SL} = \mathrm{SPL} + \mathrm{TL}, \tag{1}$$

where SPL is the sound pressure level reference to 1 μPa peak to peak of the recorded signal, TL is transmission loss, which can be calculated from the range between the phonating animal and the hydrophone, assuming spherical spreading [typical of spreading observed in dolphin and porpoise sonar (Au, 1993)]:

$$\mathrm{TL} = 20\log(R) + \alpha R, \tag{2}$$

where α is the absorption coefficient of the water measured in dB/m [$\sim$0.035 dB/m at 120–130 kHz and 30 °C (Au, 1993)].

III. RESULTS

During this experiment, finless porpoises were frequently observed passing up and down on both sides of the hydrophone, at speeds of about 1 m/s. A total of 6 h of underwater signal recordings were obtained, and 167 click trains with high signal-to-noise ratios were examined. Of these, 53 (about 32%) were suitable (i.e., had clear, distinguishable surface and bottom reflections) for estimating ranges and calculating source levels of phonating animals. All of the acoustically estimated ranges were comparable to the visually estimated ranges. A total of 159 SLs were calculated. The maximum SL was 185.6 dB calculated at a range of 47.5 m, and the minimum SL was 163.7 dB calculated at a range of 3.8 m from a phonating porpoise.

By assuming the porpoise was scanning the hydrophone by sonar when ICIs of the clicks were longer than the two-way transit times, which is defined as the time needed for an echolocation click to travel from the phonating porpoise to the hydrophone and back to the porpoise, the distance to the hydrophone can be used as the index of the target range. Figure 2 shows that the SLs of analyzed clicks with ICIs longer than the two-way transmit times increase with the log of R, according to the equation $\mathrm{SL}=19.37\log(R)+151.59$. A high correlationd coefficient ($r^2=0.93$) was determined, which is significant ($n=131$, $p<0.01$).

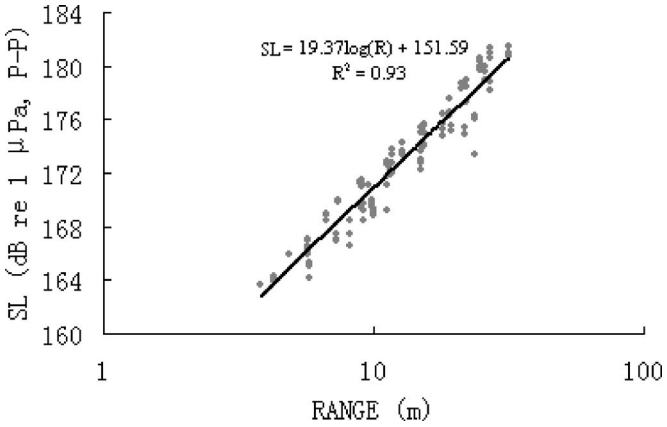


FIG. 2. Source levels SLs in dB (p-p) re: 1 μPa as a function of range R (m). The SLs increase with the log of ranges [$\mathrm{SL}=19.37\log(R)+151.59$, $r^2=0.93$], which is approximate to the one-way transmission loss [$\mathrm{SL}=20\log(R)$].

IV. DISCUSSION AND CONCLUSIONS

The experiment was performed under conditions where the directionality and range of a freely swimming phonating porpoise were uncertain. For instance, we presumed the click dataset included in the analysis is the most representative description of the on-axis or near to on-axis sonar clicks emitted by the porpoises. Actually, it is difficult to determine which clicks were recorded on-axis and which were not. However, by selecting click trains with a high signal-to-noise ratio and including only the three clicks with the highest amplitude and highest peak frequency in one click train, we can surmise the analyzed clicks in the present study were approximately on the transmission beam. The SL was calculated using the range estimated by the one-hydrophone acoustic method (Li *et al.*, 2005b) under the condition of assuming the bottom of the experiment area to be flat with a consistent depth of 4 m, which might introduce some errors for the distance estimates. Nevertheless, the potential problem was minimized by only using click trains with clicks having very distinct surface and bottom reflections to estimate ranges. Additionally, a comparison to visual estimates showed all the acoustic estimates of distances are approximately correct. Even considering measurement errors of 0.5 m for the bottom depth, 4 μs for the time delays of the signals via the surface, and bottom reflection paths to the hydrophone, relative to the direct signal, the acoustically calculation error for the range is less than 40% (Li *et al.*, 2005b), which can be considered to be fairly accurate for the range estimates (Aubauer *et al.*, 2000). In addition, the frequency response of our hydrophone was not so flat, being +3/−12 dB with a frequency between 100 and 150 kHz; however, the frequency response of the hydrophone near the peak frequency of clicks in a finless porpoise had been compensated in the data analysis. As a result, we believe the calculated source levels in the present study were approximate to the real source levels of on-axis signals in finless porpoise.

When clicks with ICIs longer than the two-way transit times were received by hydrophone, the phonating animals were assumed to be detecting the hydrophone as a target.

Figure 2 shows the SLs of the clicks with ICIs longer than the two-way transit times increased linearly with the log of the range to the hydrophone, with a high correlation coefficient ($r^2=0.93$). The high value of r^2 might have been caused by the selection of click trains with a high signal to noise ratio without regard to the distances between phonating animals and hydrophone, and clicks with ICIs longer than the two-way transit times, both of which are to assure that the porpoises were detecting the hydrophone, even at the long distances. The regression line [$19.37\log(R)$ in Fig. 2] was close to the one-way transmission loss of the echolocation signals for a given distance R [$20\log(R)$]. This indicates that as the porpoises close in on a target, the SL of the echolocation signal decreases continuously by 6 dB for each halving of the distance. The reduction in SL with decreasing distance to the hydrophone is a form of dynamic time-varying gain control previously described in the dolphin sonar system (Au and Benoit-Bird, 2003; Rasmussen *et al.*, 2002). This type of gain control system adjusts the level of the transmitted signal with a coefficient of about $20\log(\mathrm{R})$. This amplitude is a function of the one-way transmission loss for signals traveling from the animals to the hydrophone, and is different from that of bats, who control their hearing sensitivity by contracting their middle-ear muscles synchronized to transmissions with a coefficient of $40\log(R)$ (Simmons *et al.*, 1992). The echo gain control in bats keeps echoes from point targets at a fixed sensation level (Simmons *et al.*, 1992). Hartley (1992)found that *Eptesicus fuscus* also reduces emitted intensity by $20\log(R)$ under some conditions. In this case, the decrease in emitted intensity is synchronized to a reduction in auditory sensitivity of 6 to 7 dB [about $20\log(\mathrm{R})$] per halving of range, to stabilize perceived echo amplitudes during target approach (Hartley, 1992). However, as the porpoises and dolphins cannot contract the middle-ear muscles, they are not known as having the capability to attenuate audition as bats (Ketten, 2000), when detecting the point targets with no change in target reflectivity (such as the hydrophone in the present study) by only $20\log(R)$ SL adjustment, the dolphins and porpoises perceive echo level increasing by 6 dB for each halving of the target range.

ACKNOWLEDGMENTS

We especially acknowledge the staff at Baiji Aquarium and Tongling Baiji Semi-Natural Reserve for sharing time and resources, and for providing endless support. We would also like to express our appreciation to Dr. Isaac Bao, Prof. Whitlow Au, and three anonymous reviewers for their valuable comments on the earlier version of this manuscript. This research was supported by grants from the Institute of Hydrobiology, Chinese Academy of Sciences (220103) and the Program for Promotion of Basic Research Activities for Innovative Biosciences of Japan.

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 truncatus*)," J. Acoust. Soc. Am. **104**, 2511–2516.

Akamatsu, T., Wang, D., Wang, K., and Naito, Y. (**2005a**). "Biosonar behaviour of free-ranging porpoises," Proc. R. Soc. London **272**, 797–801.

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., 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.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).

Au, W. W. L. and Benoit-Bird, K. J. (**2003**). "Automatic gain control in the echolocation system of dolphins," Nature **423**, 861–863.

Au, W. W. L., Ford, J. K. B., Horne, J. K., and Newman Allman, K. A. (**2004**). "Echolocation signals of free-ranging killer whales (*Orcinus orca*) and modeling of foraging for Chinook salmon (*Oncorhynchus tshawytscha*)," J. Acoust. Soc. Am. **115**, 901–909.

Au, W. W. L. and Würsig, B. (**2004**). "Echolocation signals of dusky dolphins (*Lagenorhynchus obscurus*) in Kaikoura, New Zealand," J. Acoust. Soc. Am. **115**, 2307–2313.

Aubauer, R., Lammers, M. O., and Au, W. W. L. (**2000**). "One-hydrophone method of estimating distance and depth of phonating dolphins in shallow water," J. Acoust. Soc. Am. **107**, 2744–2749.

Goold, J. C. and Jefferson, T. A. (**2002**). "Acoustic signals from free-ranging finless porpoises (*Neophocaena phocaenoides*) in the waters around Hong Kong," Raffles Bull. Zool., Suppl. **10**, 131–139.

Hartley, D. J. (**1992**). "Stabilization of perceived echo amplitudes in echolocating bats. II. The acoustic behavior of the big brown bat, *Eptesicus fuscus*, when tracking moving prey," J. Acoust. Soc. Am. **91**, 1133–1149.

Kamminga, C., Kataoka, T., and Engelsma, F. J. (**1986**). "Investigations on cetacean sonar VII: Underwater sounds of *Neophocaena phocaenoides* of the Japanese coastal population," Aquat. Mamm. **12**, 52–60.

Ketten, D. R. (**2000**). in Hearing by *Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 43–108.

Lammers, M. O., Au, W. W. L., Aubauer, R., and Nachtigall, P. E. (**2003**). "A comparative analysis of the pulsed emissions of free-ranging Hawaiian spinner dolphins (*Stenella longirostris*)," in *Echolocation in bats and dolphins*, edited by J. A. Thomas, C. F. Moss, and M. Vater (University of Chicago, Chicago), pp. 414–419.

Li, S., Wang, K., Wang, D., and Akamatsu, T. (**2005a**). "Echolocation signals of the free-ranging Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientialis*)," 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 porpoise (*Neophocaena phocaenoides asiaeorientialis*)," J. Acoust. Soc. Am. **118**, 3934–3940.

Rasmussen, M. H., Miller, L. A., and Au, W.W. L. (**2002**). "Source levels of clicks from free-ranging white-beaked dolphins (*Lagenorhynchus albirostris* Gray 1846) recorded in Icelandic waters," J. Acoust. Soc. Am. **111**, 1122–1125.

Simmons, J. A., Moffat, A. J. M., and Masters, W. M. (**1992**). "Sonar gain control and echo detection thresholds in the echolocating bat, *Eptesicus fuscus*," J. Acoust. Soc. Am. **91**, 1150–1163.

Thode, A., Mellinger, D. K., Stienessen, S., Martinez, A., and Mullin, K. (**2002**). "Depth-dependent acoustic features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico," J. Acoust. Soc. Am. **112**, 308–321.

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.

A short history of bad acoustics

M. C. M. Wright[a)]
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom

(Received 12 July 2006; accepted 14 July 2006)

Every branch of science attracts its share of cranks and pseudoscientists, and acoustics has been no exception. A brief survey of those who touched on acoustics is given with quotations from the more interesting or egregious examples. A contrast is drawn between the nineteenth century contrarian's quarrel with particular theories and the modern new age wholesale rejection of theory. This world-view is traced back to the later scientific writings of Goethe. Examples of pseudoscience applied to biomedical acoustics, architectural acoustics, and audio reproduction are given. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2336746]

PACS number(s): 43.10.Gi [ADP] Pages: 1807–1815

I. INTRODUCTION

"Bad acoustics" is not the same as "wrong acoustics." Acoustical science has progressed to its present, well-attested status by continuous debate, and it would be quite wrong to fault those who, for honest reasons, found themselves on the losing side of an argument. After all, as pointed out by Jeng,[1] when Laplace calculated his correction to Newton's famous underprediction of sound speed he made two errors that luckily canceled each other out.

Nor is "bad acoustics" the same as "odd acoustics." The story of Darwin playing the bassoon to worms is often given as an example of scientific eccentricity. In fact it was part of an entirely sensible test to ascertain whether they could hear, or sense vibrations directly through their bodies (he also tried the piano and a metal whistle[2]).

There is, however, a distinct strand of acoustical pseudoscience, which has produced some particularly strange and, in some cases, remarkable publications, and it is these that will be reviewed in this article.

II. DOWN WITH WAVES!

A. Alexander Wilford Hall (1819–1902)

In 1877 a Methodist clergyman from New York named A. Wilford Hall published a book called "The Problem of Human Life: Here and Hereafter," [3] Its second edition was issued in 1883 and is available for purchase today in a print-on-demand facsimile edition. His aim was to present scientific arguments against Darwinian evolution which he felt to be incompatible with religion. He was neither the first nor the last to do so, and if that were all he did his book would be of little interest to us. What is of interest is this promise from the preface:

> Prior, however, to undertaking the task of breaking through the entrenched works of the evolutionist, and in order to prepare the reader for placing the proper estimate upon these so-called scientific theories which assume to overthrow religion […] I resolved, as an example of what might be expected in the future, to attempt the overthrow of one of the universally accepted theories of science […] namely, the Wave-Theory of Sound, out of which has been developed the Undulatory Theory of Light and the more recently constructed theory of Heat as a Mode of Motion.

Wilford Hall's claim that he will take on the wave theory of acoustics as a sort of warm-up lap before overthrowing evolution is disingenuous. His true motive is shortly revealed:

> In this seemingly preposterous and hazardous attempt I was necessarily compelled to undertake the additional task of reviewing no less an authority than Professor Tyndall (the ablest and most popular exponent of the sound-theory now living), and of thus demonstrating the complete unreliability and defenselessness of the scientific opinions and statements of one of the most aggressive advocates of modern evolution, even when treating on the simple facts of science and making the most ordinary philosophical deductions.

Tyndall's book "Sound" had, by this time, become a recognized classic of popular science.[4] In his introduction to the third edition of 1875 Tyndall notes with satisfaction that it had been translated into German in an edition supervised by no less an authority than Helmholtz, and also into Chinese. Acoustics was just one of Tyndall's many scientific interests and in his time he was renowned as a passionate scientific educator and proponent of evolution, and it is this last fact that surely roused Wilford Hall. Another reason to attack Tyndall, as opposed to other acousticians, is that his book was eminently readable by those without higher scientific education when compared to that of, say, Rayleigh. Indeed, Wilford Hall gives no sign that he is even aware of Rayleigh's existence. His relish for the assault on Tyndall, as well as his misunderstanding of acoustics can be seen in the following extract (emphasis in the original):

> Whether two unison forks, or other instruments, if sounded half a wave-length apart, with the ear stationed in line, can be heard the same as in any other position, must absolutely settle the whole undula-

a)Electronic mail: mcmw@isvr.soton.ac.uk

> tory problem, now and forever. *If they can be heard the same in that as in any other position, which the whole world knows to be a fact, then the wave-theory falls to pieces, and with it falls Professor Tyndall as a scientist!*

In fact, by being sure to attack Tyndall on every possible point he did manage to score one hit. In "Heat: a Mode of Motion" [5] Tyndall uses the concept of the luminiferous ether to describe the transmission of light, memorably pointing out that

> Here your conceptions must be perfectly clear. It is just as easy to picture a vibrating atom as to picture a vibrating cannonball; and there is no more difficulty in conceiving of this *ether*, as it is called, which fills all space, than in imagining all space filled with jelly.

The fact that Wilford Hall was right to disagree with this is only marginally to his credit; as the saying goes even a broken clock will be right twice a day.

Wilford Hall's style of writing, taken in small doses, has a preacher's eloquence to it; when comparing Christian theists with the free-thinkers Underwood and Ingersoll in his first chapter he rhapsodizes that:

> The one represents the glorious eagle which is never so proud and happy as when facing the sun and soaring toward heaven, while the other is a fit symbol of the buzzard, whose glory is in its shame and whose fondest felicity is in feasting on filth.

His arguments, however, are long-winded in the extreme and he often puts words in his adversary's mouths, in one case going so far as to stage-manage an imaginary debate:

> I wish I could have the opportunity of saying to Mr. Comte, Sir: Your impression of the tree's existence is not a reality at all […] Should he admit this, as he would be forced to do by his own logic, I would then take him a step further […] Thus I might keep him going with this house-that-Jack-built logic […] till he would be totally lost in the labyrinths of his own metaphysical confusion, and be obliged to admit […]

and so on.

So what was his argument with acoustics? He described his own philosophy as Substantialism, and held that all things are material substances. He applied this principle theologically to human souls, and physically to sound waves, which he believes to be corpuscular. Naturally, he shows supreme confidence in his theory:

> Should any physicist a hundred years hence happen to be so illy informed and so far behind the age as to believe in and advocate the preposterous position involved in the current wave-theory of sound, the educated scientist of that epoch in attempting to set him right will then feel about the same indefinable sensation of pity mingled with disgust that the astronomer of to-day feels when hearing some scientific lunatic urge, as is sometimes the case, that the earth can not revolve on its axis, because if it did so it would overturn the water-bucket; or that the writer of this review is compelled to feel while trying to convince Professors Tyndall, Helmholtz, and Mayer that a locust can not, by moving its legs, throw four cubic miles of air into condensations and rarefactions, and thus exert a mechanical pressure of thousands of millions of tons!

He repeatedly returns to the example of a locust or a cricket, which can be heard from a great distance. He calculates the volume of air that would have to be set in motion for this to happen by waves, and then, fatally, assumes this body of air to be moving in unison, that is to say in solid body motion. He then divides the force necessary to achieve this by the area of a locust's leg to achieve his value of pressure. His preferred explanation is that anything that emits sound exudes particles and that sound works like smell. He was prepared for the counterargument that the locust would have to fill four cubic miles with sound particles, pointing out that some substances can be smelled over comparable distances without losing appreciable mass.

He was also unsatisfied that something as small as the human eardrum can respond to sound whose wavelengths in air can be meters long. Never one to leave a point unlabored he scripts the following imaginary exchange between Helmholtz and Steinway.

> HELMHOLTZ. Good morning, Mr Steinway. What in the world are you making there, in which you seem to be so deeply absorbed?
>
> STEINWAY. A grand piano, sir;—an improvement that is going to revolutionize the business, based on late acoustical discoveries which do away with the necessity of such enormous size and expense in construction. I am building, sir, a vest-pocket piano,—one that a musician can carry with him, where he goes, as easily as he can carry his watch. There are millions in it!
>
> HELMHOLTZ. What length, Mr. Steinway, do you propose to have the strings?
>
> STEINWAY. The longest string, or those producing the lowest notes of the bass, according to my improved scale, which I have just completed, will be exactly one inch in length, while, for the highest notes, seven octaves above, the strings will be just half that length.
>
> HELMHOLTZ. Mr. Steinway, you are a practical joker. But come, now, be serious. We Germans do not deal in jokes when we come to mechanical improvements, involving, as yours does, the established laws of acoustics […]

This script continues over three large pages.

All the quotes so far are taken from the second edition, the preface of which contains the following tantalizing information about its predecessor:

> Since the early edition of the book was published, partly in meter, the author has had an abundant reason to become satisfied that the metrical form of the argument was a mistake, so far, at least, as the general reading public is concerned.

Sure enough, the considerably scarcer first edition was largely written in nonrhyming verse, with copious prose footnotes to amplify concepts that did not easily fit within

the scansion. Typical examples of the metrical form of argument are:

> Sound I now proclaim as substance
> Real as the ear which hears it
> Or the objects which produce it,
> Notwithstanding all the reasons
> And phenomena so numerous
> Drawn from vibratory motion
> Which appear to contradict it,
> Which the reader will remember,
> As I have distinctly hinted,
> Are in harmony completely
> With the view as here foreshadowed,
> When we come to analyze them;
> And so infinitely simple,
> When compared to explanations
> Given by the current doctrine,
> That the mind at once accepts them
> As the only true solutions.

and

> And though I may speak of sound-waves
> In the course of this discussion,
> I shall do so under protest,
> With this frank asseveration
> That sonorous undulations
> Are a work of pure invention,
> Brilliantly imaginary,
> Having not the least foundation
> Either in the laws of nature
> Or the principles of science.

As a final curious footnote it turns out that in 1879 a US Patent was awarded to one A. Wilford Hall of New York for an improvement to Edison's phonograph,[6] although I have not been able to establish that this is definitely the same person.

B. Joseph Battell (1839–1915)

Colonel Joseph Battell was a wealthy Vermont landowner who donated over 30 000 acres of land to the state on condition that it be kept as a public wilderness, an action that has rightly made him something of a hero to the environmental movement. He was also a Republican member of the Vermont Legislature eight times (once in the Senate) and wrote a book about the breed of horse known as the Morgan. In 1901 he also wrote and published a book called "Ellen or Whisperings of an Old Pine,"[7] initially in one volume, though a second and a third were added in subsequent editions. It is lavishly produced with gold embossed lettering on the cover (see Fig. 1) and a profusion of photographs of Vermont scenery. It is summed up by Martin Gardner, in his classic study of pseudoscience "Fads and Fallacies,"[8] as follows:

> Few odder works than Ellen have appeared in the United States. All three volumes are in the form of a Platonic dialogue between a sixteen-year-old girl named Ellen and the narrator who happens to be an old Vermont Pine tree.

Ellen and the Pine (who is the nominal narrator) refer to each other and themselves in the third person throughout, which lends the prose a particular sluggishness. Neither expresses any surprise at the situation they find themselves in. Indeed at one point Ellen admonishes the Pine to remember when he first learned algebra, although no explanation of when or why these lessons were conducted is given. The relationship between the two principal characters is highly dysfunctional. The Pine is pathetically grateful for Ellen's attention, which she is constantly threatening to withdraw, often in favor of other trees. Although they pretend to engage in free philosophical discussions Ellen soon makes it clear who decides whether a particular topic is legitimate (the quotes are from the second edition):

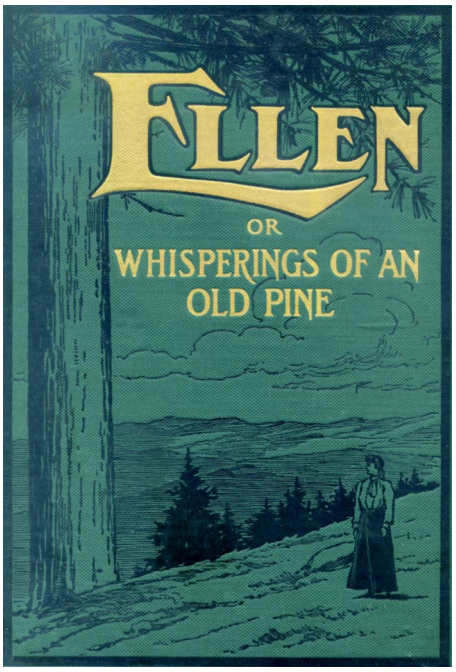


FIG. 1. (Color online) The cover of "Ellen" (second edition)—Ref. 7.

> "[…] does the old Pine suppose that the intelligence which he suggests comes from life lives after the life is gone?" […]
> "The old Pine doesn't think at all," I said. "He is only asking questions, and letting Ellen do all the thinking."
> "And Ellen thinks he is getting crazy," she replied, "with such foolish questions."
> "But the scientists ask such questions," I said, "as though they could not be answered, and they are very great men."
> "Very ignorant men," she said […]
> "Well," I said, "the old Pine was striving to get the facts."
> "Asked Ellen lots of foolish questions," she said. "Ellen got awfully scared about him. Afraid he was losing his wits."
> "But, Ellen, the old Pine doesn't know of any way to get at the truth but by trying. It is the bold mariner only who makes discoveries."
> "And doesn't the old Pine know," she said, "that there are no discoveries possible about things which

> are self-evident? The old Pine was awfully crazy and Ellen was dreadfully frightened; afraid he would never get out of it. He talked just like a scientist."

Ellen takes the Pine through theology, botany, algebra, and trigonometry, in all cases ridiculing experts while defending any weaknesses in her own theories by vehement assertions that they are self-evident and need no proof. Whole chapters are taken up by Ellen quoting learned works, from Plato to Newton to the Edinburgh Philosophical Journal, all seemingly off by heart. By Volume II they have got around to the subject of sound, and Ellen favors (or rather insists upon) a corpuscular theory like Wilford Hall's and has a similar loathing for Tyndall. Among her reasons for dismissing the wave theory is her contention that superposition would be impossible (again this is claimed as self-evident) so that multiple sound waves passing through the same space would be affected by one another. She makes much of Newton's famous mistake of assuming isothermal rather than adiabatic change in sound waves and thus underpredicting the speed of sound. As for Laplace's correction mentioned earlier she has this to say:

> By this hypothesis of Laplace one-half of the air is constantly overheated, and one-half underheated, and this couldn't help being noticeable if it meant any perceptible amount of difference of temperature, even though the two halves should constantly interchange conditions. For the old Pine will remember that some of the hypothetical sound waves are quite long, that of the lower C 28 feet, having 14 feet of condensation and 14 of rarefaction. And this increase of heat which, not in the ordinary way but in some inexplicable manner, is said to add 176 feet per second to the speed of sound, must take place with every sound, even the slightest. […] It makes Ellen pretty sick to discus seriously such intolerable nonsense.

She also harps on the point that the recently invented telephone can still be faintly heard when its diaphragm is removed. This could be explained as magnetostriction in the coils causing a small force on the housing which acts as an inefficient radiator. To Ellen, however, it is proof of nothing less than the complete failure of wave theory.

III. INTO THE NEW AGE

A. John Ernst Worrel Keely (1837–1898)

John Keely claimed to have invented a perpetual motion machine, which he called a "vibratory generator with a hydro-pneumatic pulsating generator," and insisted that sympathetic vibrations were essential to the functioning of his device. The story is taken up in John Sladek's "The New Apocrypha:" [9]

> From time to time investors in the Keely Motor Company began to wonder if they were wasting their money. Keely always persuaded them to waste a bit more. Demonstrations always took place in Keely's home, where the motor tore ropes apart and twisted iron bars, while its gauges showed enormous pressures—all from a pint of water. Committees of scientists and engineers were invited to see his demonstrations, but not to inspect the motor. They did so, however, after his death in 1898, and found in the cellar the compressed-air equipment that really ran it.

In fact, one person invited to witness a demonstration of Keely's motor was none other than the Reverend Dr A. Wilford Hall, who, apparently, was not particularly impressed.

Some, however, have not been shaken by such revelations. The 1996 book "The Physics of Love: The Ultimate Universal Laws" [10] is credited to Dale Pond, Edgar Cayce, John Keely, Rudolf Steiner, and Nikola Tesla, though only the first author was alive at the time of writing. In it, Keely's secrets are supposedly revealed. A brief sample is enough to get the flavor:

> The Law of One: Everything that is in the universe is included in it. Therefore all that is may be considered as one yet each discrete thing is individualized. The common mechanical connection between them all is what they have in common-vibratory motions. Every thing in the universe vibrates according to the laws of harmony. The connecting link between all these seemingly separate things is sympathetic vibrations. When the numbers of the vibrations are the same there is greater action/reaction or commoness [*sic*] of experience […]

Note the contrast between the earlier attacks on establishment science by Wilford Hall and Battell, in which an accepted theory is attacked and evidence (however spurious or badly interpreted) is offered to convince the reader, and the new age mystical literature in which anything can be conjectured and whether or not it agrees with scientific theory or observation is not even considered relevant.

B. Hans Jenny (1904–1972)

The transition between the old and new styles can be detected in the writings of Hans Jenny, a friend and follower of the occultist Rudolf Steiner (1861–1925). A proper survey of Steiner's beliefs and activities would be beyond the scope of this article. Briefly, he broke away from Madame Blavatsky's Theosophy cult to found his own Anthroposophy movement. Today he is best-known for the foundation of Waldorf or Steiner schools, some of which are still in controversial existence. Jenny, a medical doctor by profession, became fascinated with visual manifestations of vibration such as Chladni plates and Lissajous figures and set about studying them. The way he did so emphasizes the difference between the scientific method and Steiner's approach to the world. Jenny coined the term "Cymatics" to mean the study of waves and in 1967 published a book of that name, followed by a second volume in 1972.[11] It contains many photographs, some of them remarkable and beautiful, of powders, liquids, and pastes on vibrating plates, such as those shown in Fig. 2, taken from Chapter 8 of the first volume. These show a suspension of kaolin placed on a vibrating membrane. In some of the pictures the suspension is cooling and solidifying, in others it is a viscous mixture. It is inter-

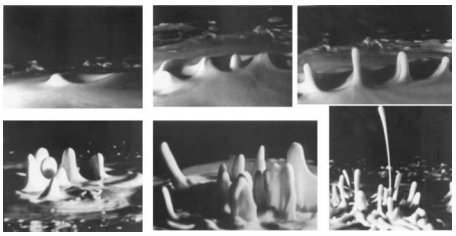


FIG. 2. A montage of pictures of kaolin paste on a vibrating membrane from Jenny's "Cymatics" (Ref. 11). The figures do not appear in this order there, and the properties of the paste are not constant over all figures. © MACROmedia publishing, used by permission.

esting to compare these to Fig. 3, which shows the delocalization of a hole in vertically vibrated cornstarch suspension reported by Merkt *et al.* in 2004.[12] This could be taken to suggest that a self-taught amateur working alone had preempted the work of a number of experts at a world-class research center. But the differences between the reports are as revealing as the similarities. Merkt *et al.* are able to map how the qualitative behavior of the system depends on amplitude and frequency of vibration. In order to do this they take considerable trouble to make sure that the properties of the suspension are fixed, uniform, and repeatable. They are then able to discuss the observed behavior in the context of the shear-thickening property of the cornstarch suspension they used. And, crucially, they were able to observe persistent holes, which Jenny missed.

Jenny, in contrast, gives no frequencies or amplitudes, nor any of the properties of his mixtures. All of this might be forgivable in an amateur investigator who might not possess, for example, the means to measure viscosity, although he makes no reference to the shear-thickening property of kaolin suspensions, which are readily apparent to anyone who tries to stir them. It soon becomes apparent, however, that Jenny didn't just neglect to categorize his results systematically, he actively avoided doing so:

> What it all boils down to is this, we must keep on asking ourselves as Goethe did: "Is it you or is it the object which is speaking here?" If we were to establish rigid definitions and split up the various manifestations into sections, we should be artificially dismembering the phenomenon by applying the analytical instrument of the intellect. If the phenomenon is to remain vital, its spectrum must be grasped as a fluctuating entity. True, there are significant forms there; but what we have to evolve is the concept of moving form and formative movement.

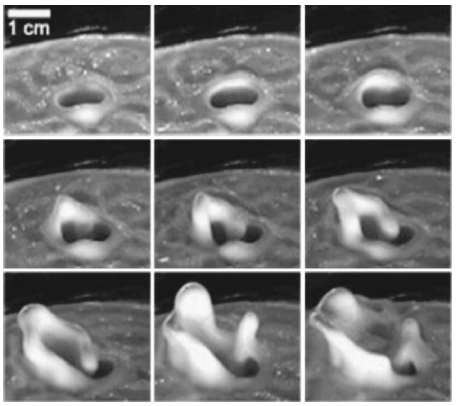


FIG. 3. A persistent hole in a cornstarch suspension being vibrated at 80 Hz at a peak acceleration of 25 g. Photos are taken every 0.9 s. Reprinted figure with permission from F. S. Merkt *et al.*, Phys. Rev. Lett. Vol. 92, 184501, 2004. Copyright 2004 by the American Physical Society.

This avoidance of any attempt at categorization or systematic investigation in favor of what would now be called a holistic approach is made explicit in the opening chapter of the second volume:

> Here again methods have been employed in which the phenomenon is treated as a whole and not dissected. Why is this? When we observe a phenomenon, it is natural to concentrate on one single factor and make it the focus of our attention. Now, if such a factor is abstracted from its context and allowed to dictate our procedure, the investigation tends to become biased and other characteristics of the object under study are easily missed. This is clearly reflected in the history of science in the way interest has alternated between opposed theories.
>
> […]
>
> A very special feature of the study of vibrations is the way in which the observer penetrates to the genetic element. Before our eyes we have the creative and the created, the vibrating and the sounding, and also what is produced by vibration and sound. Now none of this can be simply and harmlessly dissected for our examination. The events of the wave sequence transpire under complex conditions, in interferences, resonances, turbulences, in harmony, consonance, in disharmony, in dissonance, in frequency spectra, amplitude relations, etc. It is in this sphere of multiple creation that the investigator must carry out his observations. He must find out whether amidst all this tumultuous activity there are basic or ultimate phenomena in term of which "everything else" can be comprehended. It happens often enough that we have the parts in our hands but unfortunately lack the "mental ribbon" (Goethe) with which to bind them together. What is the status of the parts, the details, the single pieces, the fragments? In the vibrational field it can be shown that every part is, in the true sense, implicated in the whole.

The result is, for any physicist, a series of missed opportunities. Strange and rich behaviors, which would provide several interesting challenges to model, rub shoulders with Lissajous figures created with sine wave generators and an oscilloscope, which can be completely explained with high school physics; both are treated with the same mystic wonder:

> We have, among other things, been able to generate the formal vocabulary of Gothic tracery. It would therefore be correct to say that these architectural forms actually embody intervals as figures, thus verifying Geothe's dictum that "architecture is frozen music."

Unfortunately, although extensive examples and photographs are given through the book, none are given that show this architectural correspondence. We are, however, shown the result of speaking vowels into his "tonoscope," a stretched membrane with powder, sand, or liquid used to show the resulting nodal pattern, and of playing the music of Bach and Mozart into an electroacoustic version.

C. Johann Wolfgang von Goethe (1749–1832)

Jenny's reference to Goethe (there are several throughout the book) is apposite; the author of Faust had a parallel career as an amateur scientist and wrote "Zur Farbenlehre," [13] a large volume on the theory of color, which he regarded as his most important work; as Gardner[8] puts it:

> Since Goethe had no understanding of experimental methods, and even less of mathematics, his attack proved one of the most irrelevant in the history of physics.[14]

Although it diverts us from acoustics to optics it is worth noting the responses to Goethe of two eminent acousticians. Helmholtz, in an 1853 lecture to the German society of Königsberg[15] considered both his useful contributions to osteology (he discovered evidence for the presence of the intermaxillary bone in the human jaw) and to optics. He describes how Goethe looked at a white wall through a borrowed prism expecting (thanks to a misreading of Newton) to see colors. When he saw them only at edges, such as the boundary of a black figure on a white background he leapt to the conclusion that Newton's theory could only be utterly wrong. Helmholtz continues

> It is evident that at the first moment Goethe did not recollect Newton's theory well enough to be able to find out the physical explanation of the facts I have just glanced at. It was afterward laid before him again and again, and that in a thoroughly intelligible form, for he speaks about it several times in terms that show he understood it quite correctly. But he is still so dissatisfied with it that he persists in his assertion that the facts just cited are of a nature to convince any one who observes them of the absolute incorrectness of Newton's theory. Neither here nor in his later controversial writings does he ever clearly state in what he conceives the insufficiency of the explanation to consist. He merely repeats again and again that it is quite absurd.

He also points out Goethe's disdain for experiment, which Jenny inherited:

> […] in his attack on Newton he often sneers at spectra, tortured through a number of narrow slits and glasses, and commends the experiments that can be made in the open air under a bright sun, not merely as particularly easy and particularly enchanting, but also as particularly convincing!

Goethe's later writings also prefigured Wilford Hall in his attacks on his opponent. Helmholtz again:

> To give some idea of the passionate way in which Goethe, usually so temperate and even courtier-like, attacks Newton, I quote from a few pages of the controversial part of his work the following expressions, which he applied to the propositions of this consummate thinker in physical and astronomical science—'incredibly impudent'; 'mere twaddle'; 'ludicrous explanation'; 'admirable for schoolchildren in a go-cart'; 'but I see nothing will do but lying, and plenty of it.'
> […]
> Thus in the theory of colour, Goethe remains faithful to his principle, that Nature must reveal her secrets of her own free will; that she is but the transparent representation of the ideal world.

Of course Newton was equally irascible but generally preferred to use his position to suppress his opponents work rather than to attack them in print.

John Tyndall, in a Friday Evening Discourse at the Royal Institution of 1880,[16] makes similar points, but draws an analogy that modesty would have forbidden Helmholtz to make:

> We frequently hear protests made against the cold mechanical mode of dealing with aesthetic phenomena employed by scientific men. The dissection by Newton of the light to which the world owes all its visible splendour seemed to Goethe a desecration. We find, even in our own day, the endeavour of Helmholtz to arrive at the principles of harmony and discord in music resented as an intrusion of the scientific intellect into a region which ought to be sacred to the human heart. But all this opposition and antagonism has for its essential cause the incompleteness of those with whom it originates. […] There is no fear that the man of science can ever destroy the glory of the lilies of the field; there is no hope that the poet can ever successfully contend against our right to examine, in accordance with scientific method, the agent to which the lily owes its glory.

D. Pyramidology and beyond

The Egyptian Pyramids seem to act as a magnet to unconventional theories. There has apparently been a longstanding debate about how the stones were moved, to which surely the least convincing answer must be that given in "Gods of Eden" (1998) by Andrew Collins,[17] which is that they were moved by acoustic levitation. Of course acoustic levitation is a scientifically recognized phenomenon, described and explained in the pages of this journal among others. The radiation forces emitted by sound waves can only lift small particles, though more substantial weights can be lifted through nonlinear effects.[18] None of this technology is mentioned by Collins, who bases his belief on the 1961 book "Försvunnen Teknik" by Henri Kjellson,[19] which reports the experiences of a Swedish doctor known only as Jarl who claimed, on a journey through Tibet some time in the two decades before the second world war, to have witnessed monks using drums and trumpets to levitate large stone blocks.

In "The Giza Power Plant" [20] (1998) Christopher Dunn

suggests that the pyramids were, in fact, built to generate energy, possibly for a highly advanced civilization. He also claims that artifacts found in the pyramids could only have been machined with ultrasonic tools.

John Reid, in experiments described in "Egyptian Sonics," [21] apparently sought to emulate Jenny's tonoscope by stretching a membrane over the sarcophagus in the King's chamber of the Great Pyramid and insonifying it to apparently reveal hieroglyphics in the resulting vibration patterns.

Pyramids are not mentioned in "Healing Codes for the Biological Apocalypse" by Dr. Leonard G. Horowitz and Dr. Joseph S. Puleo,[22] but almost everything else is. As the dust jacket puts it:

> This book […] offers the greatest hope for humanity to spiritually evolve, and reveals the divine musical notes destined to be sung by the gathering critical mass of "144,000" people required to establish 1,000 years of world peace. Let the singing and the greatest healing of all time begin!

Mad cow disease, freemasonry, Bible codes, numerology, and so forth are brought up in dizzying sequence. As for the secret frequencies (emphasis in the original):

> *These previously secret sound frequencies, or electromagnetic vibrations, are likely the primary ones associated with the matrix of creation and destruction. That is, they were likely the frequencies used by God to form the cosmos in six days, as well as the tones required to shatter Jericho's great wall in six days*. Additional evidence for this assertion come from the fact that the third note Mi for Miracles, or 528 is the exact frequency used by genetic engineers throughout the world to repair the blueprint of life, DNA, the healthy core of which is six-sided crystal hexagonal clustered water.

Readers may also be interested to learn that

> Beethoven, like his Masonic mentors, most likely created his masterpieces transposing the mathematics encoded in the Bible, and elsewhere, into musical scores.

IV. SOUND POWER

A. Health and human factors

Orthodox medical science has found great benefit in sending sound waves into the human body (for purposes such as therapeutic ultrasound, lithotripsy, or excising tumors depending on the wave form and intensity), monitoring the sounds that come from it (heartbeats, otoacoustic emissions, etc.), or doing both at the same time (ultrasound imaging). A "cargo cult" version of this process seems to have developed among some alternative practitioners, variously known as vibration retraining, vibrational medicine, or sound therapy. As Sharry Edwards of Sound Health Inc. explains on her website[23]

> Each person possesses unique harmonics of frequency that can be expressed through the voice. However, when these complex frequencies of the body become unbalanced, the voice primarily reflects this altered state, and the body manifests it as dis-stress or dis-ease at the structural and biochemical levels.
>
> In reality, there are no solids. We exist in a universe that consists entirely of energy. Einstein proved this. Frequency defines it. Frequency, as vibrational medicine, is at the heart of the world of wellness as we know it.

The details of this therapy vary somewhat among practitioners but the principle appears to be that diseases can be diagnosed by spectral analysis of the voice, and then cured by playing back the necessary sounds. The same website says:

> Experiments have been repeated that show that introducing a person to the frequency formula for niacin, a nutritive substance, can cause a niacin-like skin flushing; the same as if the person actually ingested the nutrient.

This is a claim which, if confirmed by a double blind, randomized, placebo-controlled clinical study would be a significant challenge to explain. So far, however, none have been reported in any reputable peer-reviewed journal. Another sound therapist, whom I had the interesting experience of interviewing, informed me that playing the correct sounds through a domestic loudspeaker to a test-tube of blood could change the level of uric acid in it to a measurable extent.

It can be difficult for legitimate scientists to know how to respond to such grandiose but ridiculous claims. To test every such claim would be an unconscionable drain on time and resources, and would in many cases unjustly dignify the claimant's position, but to deny the claim without testing it is open to mischaracterization as closed-mindedness or fear of new ideas. Fortunately there is a third way: the magician[24] and skeptic James Randi's Educational Foundation offers a one million dollar prize for a demonstration under an agreed protocol of any such paranormal power. The progress of applicants can be followed through the foundation's website.[25] Any scientist faced with an unreasonable but in principle testable claim can suggest that they win the million dollar challenge, thus impressing scientists the world over, instead of having to convince them one at a time. Of course, many excuses for not taking the challenge are offered; as one wag put it "if it ducks like a quack, it probably is a quack." Strangely, no practitioner of vibrational healing or any of its variants has been successful, and to the best of my knowledge none have ever applied.

One amusing effect of the prevalence of such ideas can be seen in the warning added to the website of one supplier of scientific tuning forks:[26]

> These tuning forks are intended for science and engineering applications; we cannot guarantee their performance for any other uses or for metaphysical expectations.

B. Snake oil for the ears

The fact that the human auditory system is connected to the human brain makes it a marvellous subject for study, but it also means that we are capable of being fooled about what we are listening to. This, among other factors, has made

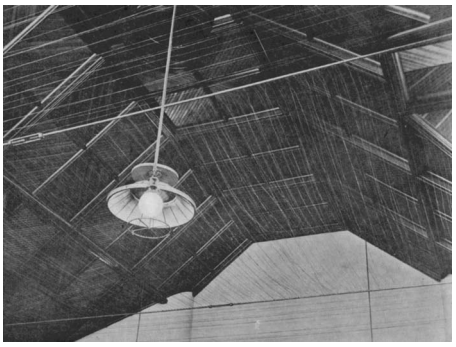


FIG. 4. Stretched wires in the ceiling of a church in San Jose, CA (Ref. 27).

objective assessment of subjective listening experiences very difficult, and easily swayed by suggestion, a fact that is exploited by many purveyors of devices that purport to improve sound.

Architectural acoustics was prey to a form of this delusion as recorded by Sabine:[27]

> The stretching of wires is a method [for remedying acoustical difficulties] which has long been employed, and its disfiguring relics in many churches and court rooms proclaim a difficulty which they are powerless to relieve. Like many other traditions, it has been abandoned but slowly. The fact that it was wholly without either foundation of reason or defense of argument made it difficult to answer or to meet. The device, devoid on the one hand of scientific foundation, and on the other of successful experience, has taken varied forms in its application. Apparently it is a matter of no moment where the wires are stretched or in what amount. There are theatres and churches in Boston and New York in which four or five wires are stretched across the middle of the room; in other auditoriums miles on miles of wire have been stretched; in both it is equally without effect. In no case can one obtain more than a qualified approval, and the most earnest negatives come where the wires have been used in the largest amount. Occasionally the response to inquiries is that "the wires may have done some good but certainly not much," and in general when even that qualified approval is given the installation of the wires was accompanied by some other changes of form or occupancy to which the credit should be given.

A photograph of such an installation, taken from Sabine's collected papers, is shown in Fig. 4.

Many readers will know of the long debate over the importance of Stradivari's varnish to the sound of his violins. Consider, then, the claims made by luthier Dieter Ennemoser for his C37 varnish:[28]

> All attempts by science to explain the secrets of the character of sound have so far been unsuccessful. […] The imperative selection of the right materials (wood and varnish quality) raised the question about the existence of a reference property. I eventually discovered that human bones and tissue to possess similar qualities. A more detailed analysis showed that carbon is the decisive element in sound quality, and since the sound is also coloured by body temperature, I chose to call this property the C37 structure. (Where C=Carbon and 37=body temperature in degrees Centigrade). Further analysis showed that C37 frequencies lie very close together (at least 10 frequencies per octave) and this structure reoccurs in each octave.

This apparently miraculous (but, naturally, expensive) substance apparently improves the sound of anything it is applied to, not just violins but loudspeaker cones and even Hi-Fi volume controls.

Once the sound has left the loudspeaker it still needs to survive transmission through the room.[29]

> Brilliant Pebbles is a unique room & system tuning device for audio systems and satellite TV. Original (Large) Brilliant Pebbles is a 3-inch clear glass bottle containing various minerals/stones. A number of highly-specialized, proprietary techniques are used for preparation/assembly. Brilliant Pebbles acts as both a vibration "node damper" and EMI/RFI absorber via various atomic mechanisms in the crystal structures. On the floor in room corners, Large Brilliant Pebbles reduces comb filter effects caused by very high sound pressure levels that occur in the corners when music is playing. Large Brilliant Pebbles is also effective on tube and solid state amps, on speaker cabinets, on armboards of turntables and on tube traps and Room Lenses.

The doyen of this field is Peter Belt, whose products have a small but devoted band of followers who seem to be convinced beyond doubt that their listening experience has been enhanced by the use of his products. And how could they not, after all:[30]

> Any series of identical species of living objects are linked by Nature irrespective of their location within the world. This applies equally to inanimate objects such as identical Compact Discs, vinyl records, printed objects etc. The energy pattern emanating from such man made objects has similarities to those same living objects to which our senses evolved. The man made objects have, however, some energy patterns which are dissimilar from those emanating from living objects. Placing a strip of the new type 'Real' Foil on these man made objects within the listening room interjects a changed energy pattern which allows the senses to respond as though the man made object had the same energy pattern as a living object.

If the actual effects of these foil strips, jars of stones and varnish might seem negligible the prices charged for them are certainly not; readers are invited to guess what they might be before investigating for themselves.

Of course, the actual value of any such device can be separated from the psychological effect of its presence (and price) on the listener by careful double-blind testing. Sadly this is strongly resisted by a significant proportion of the audiophile community. Until it becomes commonplace it will

be hard for those who seek to improve Hi-Fi systems by legitimate means to distinguish themselves from those who just sell false hope. As a last psychosociological note it is worth pointing out that such devices are given short shrift in the world of professional audio systems, where the audience neither knows nor cares what has been done to the equipment, and is therefore immunized to the power of suggestion.

V. CONCLUSIONS

We are unlikely to see another Wilford Hall or Battell arguing that the fundamental theories of acoustics are mistaken, because acoustics is no longer sufficiently novel. Contrarians have long since moved on the denying the validity of relativity, quantum physics, and cosmology. But fantastical powers are still regularly ascribed to acoustic and vibratory phenomena that can be understood with elementary physics. Furthermore, the language of acoustics forms a significant part of the new-age lexicon, replete as it is with resonance, harmony, vibration, waves (all good), and rays (usually bad). Without evidence I can only offer the conjecture that this is because this mindset saw a great growth in popularity in the 1960s and 1970s, when supersonic flight was front page news—if the TV series Dr Who were starting today I do not think its hero would depend on a sonic screwdriver. In the present day a key goal of scientists and educators is clear: to ensure that our students (in the broadest sense of the word) are equipped with the critical thinking skills necessary to avoid falling into any of the traps listed here, or becoming ensnared in others of their own making.

ACKNOWLEDGMENTS

This article is based on a lecture first prepared for the EPSRC Summer School on Mathematics for Acoustics, and subsequently given to other audiences in Southampton and Cambridge. Thanks are due to all the audience members whose encouragement led to the writing of this article and whose questions and comments helped to improve it. Several examples of Hi-Fi pseudoscience were made known to me by James Randi's online newsletter "SWIFT." Thanks are also due to the following people for assistance, discussions, and suggestions: David Chillingworth, Frank Fahy, Martin Gardner, Martyn Hill, Lars Hinke, Sheilah Mackie, Christopher Morfey, Allan Pierce, James Randi, Christine Shadle, Harry Swinney, Niels Søndergaard, Jeff Volk, and Jim Woodhouse. This article was written while supported by an EPSRC Advanced Research Fellowship.

[1]M. Jeng, "A selected history of expectation bias in physics," Am. J. Phys. **74**, 578–583 (2006).
[2]C. Darwin, *Formation of Vegetable Mould Through the Action of Worms With Observations of Their Habits* (Murray, London, 1904).
[3]A. W. Hall, *The Problem of Human Life: Here and Hereafter*, 2nd ed. (Hall, New York, 1883), 1st ed. published anonymously 1877, facsimile 2nd ed. published by Kessinger, Whitefish, MT.
[4]J. Tyndall, *On Sound*, 3rd ed. (Longmans, Green, London, 1875).
[5]J. Tyndall, *Heat, a Mode of Motion*, 6th ed. (Longmans, Green, London, 1904).
[6]A. W. Hall, "Improvement in phonographs," US Patent 219,939, 23 September 1879.
[7]J. Battell, *Ellen, or Whisperings of an Old Pine*, 2nd ed. (The American Publishing Company, Middlebury, VT, 1901).
[8]M. Gardner, *Fads and Fallacies: In the Name of Science* (Dover, New York, 1957).
[9]J. Sladek, *The New Apocrypha*, 3rd ed. (Panther, London, 1978).
[10]D. Pond *et al.*, *The Physics of Love; The Ultimate Universal Laws* (Message, Santa Fe, NM, 1996).
[11]H. Jenny, *Cymatics* (MACROmedia, Newmarket, NH, 2001), first published Vol. **1**, 1967, Vol. **2**, 1972.
[12]F. S. Merkt, R. D. Deegan, D. I. Goldman, E. C. Rericha, and H. L. Swinney, "Persistent holes in a fluid," Phys. Rev. Lett. **92**, 184501 (2004).
[13]J. W. von Goethe, *Theory of Colours* (MIT, Cambridge, 1969), first published 1810.
[14]This view has recently been contested. (Ref. 31).
[15]H. von Helmholtz, "On Goethe's scientific researches," in *Popular Lectures on Scientific Subjects* (Longmans, Green, London, 1898).
[16]J. Tyndall, "Goethe's 'Farbenlehre,' " in *New Fragments* (Appleton, New York, 1897).
[17]A. Collins, *Gods of Eden* (Bear, Rochester, VT, 1998).
[18]V. Vandaele, P. Lambert, and A. Delchambre, "Non-contact handling in microassembly: Acoustical levitation," Precision Engineering—Journal of the International Societies for Precision Engineering and Nanotechnology **29**, 491–505 (2005).
[19]H. Kjellson, *Forsvunden Teknik* (Nihil, Copenhagen, 1974). (Danish translation, First published in Swedish as *Försvunnen Teknik*, 1961).
[20]C. Dunn, *The Giza Power Plant* (Bear, Rochester, VT, 1998).
[21]J. Reid, *Egyptian Sonics* (Sonic Age, Northumberland, 2001).
[22]L. G. Horowitz and J. S. Puleo, *Healing Codes for the Biological Apocalypse* (Tetrahedron, Sandpoint, ID, 1999).
[23]S. Edwards, "Sound Health Inc.," 2001, URL http://www.sharryedwards.com.
[24]Interestingly the suggestion that magicians, as well as scientists, should be employed to examine supernatural claims was apparently first made by another acoustician, Sir Charles Wheatstone (Ref. 32).
[25]J. Randi, "James Randi Educational Foundation," URL http://www.randi.org.
[26]URL http://www.indigo.com/tuning/scientific-tuning-forks.html.
[27]W. C. Sabine, "Acoustical difficulties," in *Collected Papers on Acoustics* (Dover, New York, 1964), Chap. 6, pp. 132–133, first published in The Architectural Quarterly of Harvard University, March 1912.
[28]D. Ennemoser, "C37 acoustics," URL http://www.ennemoser.com.
[29]"Machina Dynamica Advanced Audio Concepts," URL http://www.machinadynamica.com.
[30]"P.W.B. 'Real' Foil—the entry point into the Real World," URL http://www.belt.demon.co.uk/product/realfoil/realfoil.html.
[31]N. Ribe and F. Steinile, "Exploratory experimentation: Goethe, Land, and color theory," Phys. Today **55**, 43–49 (2002).
[32]B. Bowers, *Sir Charles Wheatstone FRS 1802–1875* (HMSO, London, 1975).

Ultrasonic characterization of human cancellous bone using the Biot theory: Inverse problem

N. Sebaa
Laboratorium voor Akoestiek en Thermische Fysica, Katholieke Universiteit Leuven, Celestijnenlaan 200 D, B-3001 Heverlee, Belgium

Z. E. A. Fellah
Laboratoire de Mécanique et d'Acoustique, CNRS-UPR 7051, 31 chemin Joseph Aiguier, Marseille, 13009, France

M. Fellah
Laboratoire de Physique Théorique, Institut de Physique, USTHB, BP 32 El Alia, Bab Ezzouar 16111, Algeria

E. Ogam and A. Wirgin
Laboratoire de Mécanique et d'Acoustique, CNRS-UPR 7051, 31 chemin Joseph Aiguier, Marseille, 13009, France

F. G. Mitri
Ultrasound Research Laboratory, Department of Physiology and Biomedical Engineering, Mayo Clinic and Foundation, 200 First Street SW, Rochester, Minnesota 55905

C. Depollier
Laboratoire d'Acoustique de l'Université du Maine, UMR-CNRS 6613, Université du Maine, Avenue Olivier Messiaen 72085 Le Mans Cedex 09, France

W. Lauriks
Laboratorium voor Akoestiek en Thermische Fysica, Katholieke Universiteit Leuven, Celestijnenlaan 200 D, B-3001 Heverlee, Belgium

(Received 5 March 2006; revised 11 July 2006; accepted 13 July 2006)

This paper concerns the ultrasonic characterization of human cancellous bone samples by solving the inverse problem using experimental transmitted signals. The ultrasonic propagation in cancellous bone is modeled using the Biot theory modified by the Johnson *et al.* model for viscous exchange between fluid and structure. The sensitivity of the Young modulus and the Poisson ratio of the skeletal frame is studied showing their effect on the fast and slow wave forms. The inverse problem is solved numerically by the least squares method. Five parameters are inverted: the porosity, tortuosity, viscous characteristic length, Young modulus, and Poisson ratio of the skeletal frame. The minimization of the discrepancy between experiment and theory is made in the time domain. The inverse problem is shown to be well posed, and its solution to be unique. Experimental results for slow and fast waves transmitted through human cancellous bone samples are given and compared with theoretical predictions. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2335420]

PACS number(s): 43.20.Bi, 43.20.Hq, 43.20.Jr, 43.80.Cs [TDM] Pages: 1816–1824

I. INTRODUCTION

Osteoporosis is a degenerative bone disease associated with biochemical and hormonal changes in the aging body. These changes[1] perturb the equilibrium between bone apposition and bone removal, resulting in a net decrease in bone mass. This leads to a modification of the structure (porosity, trabecular thickness, connectivity, etc.) and, to a lesser extent, the composition (mineral density) of the bone. These changes result in a decrease of the mechanical strength of bone and in an increase of the risk of fracture. Osteoporosis mainly affects the trabecular bone (located at the hip, vertebrae, or heels, for instance). Early clinical detection of this pathological condition is very important to insure proper treatment.

The primary method currently used for clinical bone assessment is based on x-ray absorptiometry, and measures total bone mass at a particular anatomic site.[2] Because other factors, such as architecture, also appear to have a role in determining an individual's risk of fracture, ultrasound is an alternative to x-rays that has generated much attention.[3,4] In addition to their potential for conveying the architectural aspects of bone, ultrasonic techniques also may have advantages in view of their use of nonionizing radiation and inherently lower costs, compared with x-ray densitometric methods. Although ultrasonic methods[5–15] appear promising for noninvasive bone assessment, they have not yet fulfilled their potential. Unfortunately, a poor understanding of the ultrasound interaction with bone has become one of the ob-

stacles preventing it from being a fully developed diagnostic technique. Despite extensive research on the empirical relationship between ultrasound and the bulk properties of bone, the mechanism of how ultrasound physically interacts with bone is still unclear.

Since trabecular bone is an inhomogeneous porous medium, the interaction between ultrasound and bone is a highly complex phenomenon. Modeling ultrasonic propagation through trabecular tissue has been considered using porous media theories, such as Biot's theory.[16,17] The Biot theory is an established way of predicting ultrasonic propagation in an inhomogeneous material and was originally applied to fluid saturated porous rocks for geophysical testing. The Biot model treats both individual and coupled behavior of the frame and pore fluid. Energy loss is considered to be caused by the viscosity of the pore fluid as it moves relative to the frame. The model predicts that the sound velocity and attenuation in a two phase media will depend on frequency, the elastic properties of the constituting materials, porosity, permeability, tortuosity, and effective stress. This method should allow us to relate the physical parameters of our porous medium to ultrasonic velocity and attenuation. The Biot theory has been applied to trabecular bone with varying degrees of success.[18–24] This theory predicts two compressional waves: a fast wave, whereby the fluid (blood and marrow) and solid (calcified tissue) move in phase, and a slow wave whereby the fluid and solid move out of phase. McKelvie[18,21] predicted qualitatively the dependence of attenuation upon ultrasound frequency in cancellous bone; the attenuation values were of the right order of magnitude, but did not reproduce the full range of experimental values observed in natural tissues. However, McKelvie[18,21] was unable to predict correctly the trends in ultrasound velocity. Hosokawa and Otani[23] obtained better results by comparing the theoretical predictions using Biot theory and experiments for the wave velocities (fast and slow) than for the acoustic attenuation. Williams[19] used a limited formulation of Biot theory to calculate velocities alone and found good agreement for the fast wave velocity to predict experimental values obtained from tibial and femoral bovine cancellous bone samples. An excellent review of the application of Biot theory to ultrasound propagation through cancellous bone is given by Haire and Langton in Ref. 22.

In this paper, the ultrasonic characterization of human cancellous bone is investigated using the modified[25] Biot theory. The inverse problem is solved in the time domain using experimental transmitted signals. Five parameters are inverted: porosity, tortuosity, viscous characteristic length, Young modulus, and Poisson ratio of the skeletal frame. Experimental results are compared with theoretical predictions, giving a good correlation.

II. MODEL

The equations of motion of the frame and fluid are given by the Euler equations applied to the Lagrangian density. Here $\vec{u}$ and $\vec{U}$ are the displacements of the solid and fluid phases. The equations of motion are[17,26]

$$\tilde{\rho}_{11}\frac{\partial^2\vec{u}}{\partial t^2}+\tilde{\rho}_{12}\frac{\partial^2\vec{U}}{\partial t^2}=P\vec{\nabla}\cdot(\vec{\nabla}\cdot\vec{u})+Q\vec{\nabla}(\vec{\nabla}\cdot\vec{U})-N\vec{\nabla}\wedge(\vec{\nabla}\wedge\vec{u}), \tag{1}$$

$$\tilde{\rho}_{12}\frac{\partial^2\vec{u}}{\partial t^2}+\tilde{\rho}_{22}\frac{\partial^2\vec{U}}{\partial t^2}=Q\vec{\nabla}(\vec{\nabla}\cdot\vec{u})+R\vec{\nabla}(\vec{\nabla}\cdot\vec{U}), \tag{2}$$

wherein P, Q, and R are generalized elastic constants which are related, via Gedanken experiments, to other, measurable quantities, namely ϕ (porosity), K_f (bulk modulus of the pore fluid), K_s (bulk modulus of the elastic solid), and K_b (bulk modulus of the porous skeletal frame). N is the shear modulus of the composite as well as that of the skeletal frame. The equations which explicitly relate P, Q, and R to ϕ, K_f, K_s, K_b, and N are given by

$$P=\frac{(1-\phi)\left(1-\phi-\frac{K_b}{K_s}\right)K_s+\phi\frac{K_s}{K_f}K_b}{1-\phi-\frac{K_b}{K_s}+\phi\frac{K_s}{K_f}}+\frac{4}{3}N,$$

$$Q=\left(1-\phi-\frac{K_b}{K_s}\right)\phi K_s\Big/1-\phi-\frac{K_b}{K_s}+\phi\frac{K_s}{K_f},\quad R=\phi^2K_s\Big/1-\phi-\frac{K_b}{K_s}+\phi\frac{K_s}{K_f}.$$

The Young modulus and the Poisson ratio of the solid E_s, ν_s and of the skeletal frame E_b, ν_b depend on the generalized elastic constant P, Q, and R via the relations

$$K_s=\frac{E_s}{3(1-2\nu_s)},\quad K_b=\frac{E_b}{3(1-2\nu_b)},\quad N=\frac{E_b}{2(1+\nu_b)}. \tag{3}$$

ρ_{mn} are the "mass coefficients" which are related to the densities of solid (ρ_s) and fluid (ρ_f) phases by $\rho_{11}+\rho_{12}=(1-\phi)\rho_s$ and $\rho_{12}+\rho_{22}=\phi\rho_f$. The coefficient ρ_{12} represents the mass coupling parameter between the fluid and solid phases and is always negative $\rho_{12}=-\phi\rho_f(\alpha-1)$, α being the tortuosity of the medium. To express the viscous exchanges between the fluid and the structure which play an important role in damping the acoustic wave in porous material, the tortuosity α becomes a function of frequency, called the dynamic tortuosity[25–28] $\alpha(\omega)$. The parts of the fluid affected by this exchange can be estimated by the ratio of a microscopic characteristic length of the medium, for example pore size, to the viscous skin depth thickness $\delta=(2\eta/\omega\rho_f)^{1/2}$ (η:fluid viscosity, ω: angular frequency). This domain corresponds to the region of the fluid in which the velocity distribution is disturbed by the frictional forces at the interface between the fluid and the frame. At high frequencies, the viscous skin thickness is very thin near the radius of the pore r. The viscous effects are concentrated in a small volume near the surface of the frame $\delta/r\ll 1$. In this case, the expression of the dynamic tortuosity $\alpha(\omega)$ is given by[25]

$$\alpha(\omega)=\alpha_\infty\left[1+\frac{2}{\Lambda}\left(\frac{\eta}{j\omega\rho_f}\right)^{1/2}\right], \tag{4}$$

wherein α_∞ is the tortuosity and Λ is the viscous characteristic length.[25]

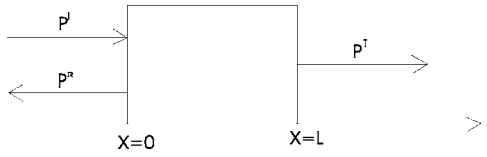


FIG. 1. Problem geometry.

For a slab of cancellous bone occupying the region $0 \leqslant x \leqslant L$ (Fig. 1), the incident $p^i(t)$ and transmitted $p^t(t)$ fields are related in the time domain by the transmission scattering operator $\tilde{T}$:

$$p^t(x,t) = \int_0^t \tilde{T}(\tau) p^i\left[t - \tau - \frac{(x-L)}{c_0}\right] d\tau, \tag{5}$$

where c_0 is the velocity outside the porous material. The transmission operator is independent of the incident signal and depends only on the properties of the cancellous bone. In the frequency domain, the expression of the transmission coefficient $\mathcal{T}(\omega)$, which is the Fourier transform of $\tilde{T}$ is given by[26]

$$\mathcal{T}(\omega) = \frac{j\omega 2\rho_f c_0 F_4(\omega)}{[j\omega\rho_f c_0 F_4(\omega)]^2 - [j\omega F_3(\omega) - 1]^2}, \tag{6}$$

where $F_4(\omega)$ and $F_3(\omega)$ are given in the Appendix.

In the next section, we solve the inverse problem and recover the physical parameters describing the propagation, using the theoretical expression of the transmission coefficient and the experimental transmitted waves propagating through human cancellous bone samples.

III. INVERSE PROBLEM

As seen in the previous section, within the framework of the modified Biot theory, the propagation of ultrasonic waves in a slab of cancellous bone is conditioned by many parameters: porosity ϕ, tortuosity α_∞, viscous characteristic length Λ, fluid viscosity η, Young's modulus of the elastic solid E_s, Young's modulus of porous skeletal frame E_b, Poisson's ratio of the elastic solid ν_s, Poisson's ratio of the porous skeletal frame ν_b, the solid density ρ_s, the bulk modulus of the saturating fluid K_f, and the fluid density ρ_f. It is therefore important to develop new experimental methods and efficient tools[29,30] for their estimation. The basic inverse problem associated with the slab of cancellous bone may be stated as follows: from measurements of the signal transmitted outside the slab, find the values of the medium's parameters.

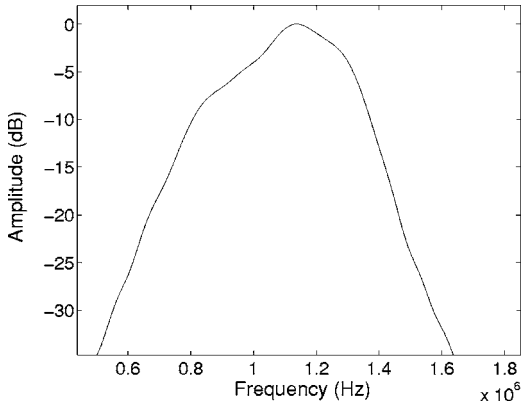


FIG. 3. Spectrum of the incident signal.

Solving the inverse problem for all the Biot parameters using only the transmitted experimental data is difficult, if not impossible. To achieve this task, requires more experimental data for obtaining a unique solution. For this reason, in this contribution we limit the inversion to the five parameters: $E_b, \nu_b, \phi, \alpha_\infty$, and Λ. In our previous paper,[26] we studied the sensitivity of transmitted wave forms to variations of ϕ, α_∞, and Λ. The sensitivity of E_b and ν_b is examined in the next paragraph.

Consider a sample of cancellous bone having the following characteristics: thickness $L=12.5$ mm, $\phi=0.9$, $\alpha_\infty=1.13$, $\eta=10^{-3}$ kg m s^{-1}, $\rho_f=1000$ kg m^{-3}, $\Lambda=8$ μm, $\rho_s=1990$ kg m^{-3}, $K_f=2.4$ GPa, $\nu_s=0.35$, $E_s=10$ GPa, $\nu_b=0.25$, and $E_b=4.16$ GPa. The incident signal used in the simulation is exhibited in Fig. 2 and its spectrum in Fig. 3. A simulated transmitted signal can be calculated in the time domain using Eqs. (5) and (6). Employing the incident signal given in Fig. 2, and the Biot parameters given below, we present in Fig. 4(a) comparison between two simulated transmitted signals corresponding to a Young modulus of the porous skeletal frame, $E_b=4.16$ GPa (solid line) and $E_b=2.08$ GPa (dashed line). The fast and slow waves can be easily identified. The

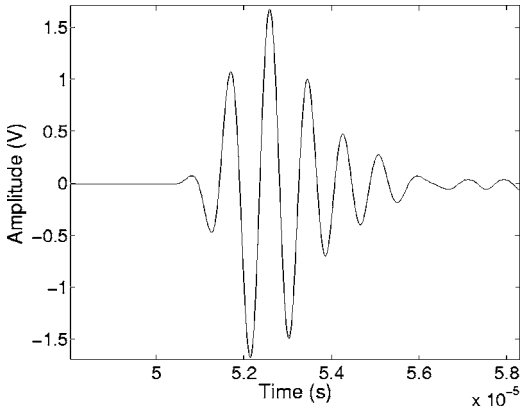


FIG. 2. Incident signal.

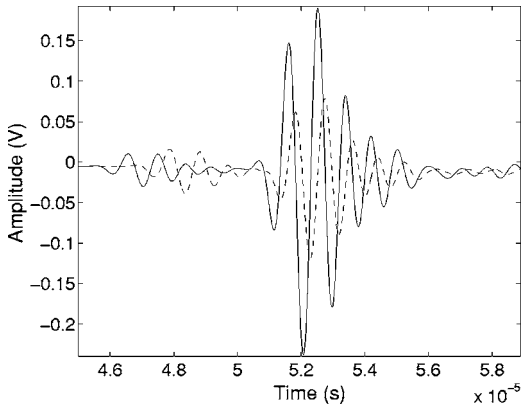


FIG. 4. Comparison between simulated transmitted signals corresponding to $E_b=4.16$ GPa (solid line) and $E_b=2.08$ GPa (dashed line).

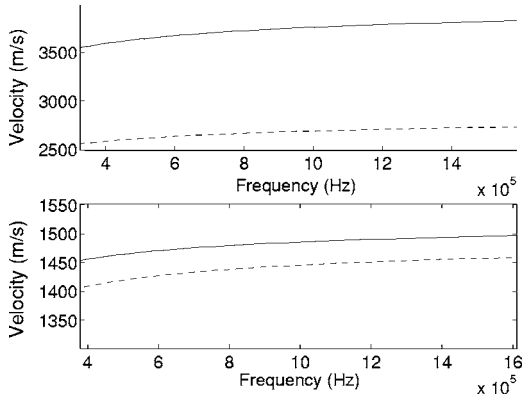


FIG. 5. Comparison between simulated velocities of the fast and slow waves corresponding to E_b=4.16 GPa (solid line) and E_b=2.08 GPa (dashed line).

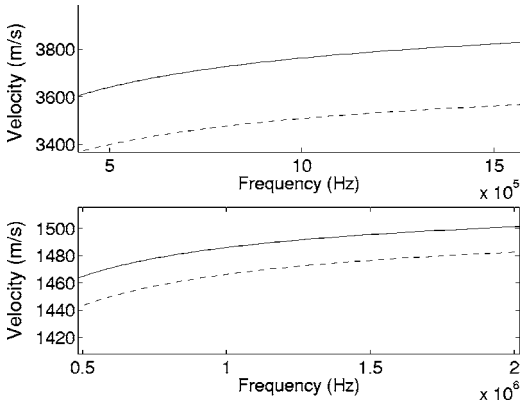


FIG. 7. Comparison between simulated velocities of the fast and slow waves corresponding to ν_b=0.25 (solid line) and ν_b=0.125 (dashed line).

simulation shows that by decreasing the value of E_b, the attenuation of the slow wave increases, while the amplitude of the fast wave remains unchanged.

In Fig. 5(a) comparison is made between the simulated velocities of the fast and slow wave, respectively, for E_b =4.16 GPa (solid line) and E_b=2.08 GPa (dashed line). It is seen that the two velocities are sensitive to the Young modulus of the skeletal frame, especially the fast wave.

When the Poisson ratio of the porous skeletal frame ν_b decreases by 50% of its initial value, the transmitted signal changes also. Figure 6 compares the transmitted signals for two Poisson ratios. The first one (solid line) corresponds to a Poisson ratio ν_b=0.25 and the second one (dashed line) to a Poisson ratio of ν_b=0.125. It can be seen that the arrival times of the slow and fast wave have changed. By decreasing the Poisson ratio value, the velocities of the two waves become lower. Figure 7 illustrates the decrease of the two velocities in the frequency bandwidth of the incident signal used in the simulation. We also note (Fig. 6) that the amplitude of the slow wave is much attenuated when the Poisson ratio decreases, while a very small change appears for the amplitude of the fast wave.

The sensitivity of E_b and ν_b with respect to the transmitted wave depends strongly on the coupling between the solid and fluid phases of the porous material and thus on the other parameters which were kept constant during this study. This analysis indicates a real sensitivity of transmitted wave forms to E_b and ν_b (i.e., to the attenuations and velocities of fast and slow waves), so that it may be possible to solve the inverse problem for E_b and ν_b.

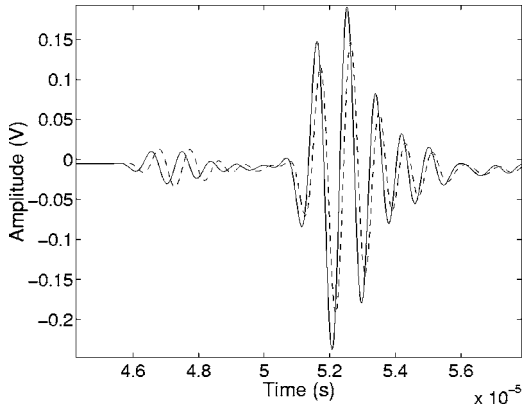


FIG. 6. Comparison between simulated transmitted signals corresponding to ν_b=0.25 (solid line) and ν_b=0.125 (dashed line).

The solution of the direct problem involves the transmission coefficient expressed as a function of the physical parameters. The inversion algorithm for identifying the values of the slab parameters in the transmitted mode is based on the procedure: find the values of the parameters $\phi, \alpha_\infty, \Lambda, E_b$, and ν_b such that the transmitted signal describes the scattering problem in the best possible way (e.g., in the least-square sense). The inverse problem is to find the parameters ϕ, α_∞, Λ, E_b, and ν_b which minimize the discrepancy function

$$U(\phi, \alpha_\infty, \Lambda, E_b, \nu_b) = \int_0^t [p^t_{\exp}(x,t) - p^t(x,t)]^2 dt,$$

where $p^t_{\exp}(x,t)$ is the experimentally determined transmitted signal and $p^t(x,t)$ the transmitted wave predicted from Eq. (5). However, because the equations are strongly nonlinear, the solution of the inverse problem using the conventional least-square method cannot be found analytically. In our case, a numerical solution of the least-square procedure consists in minimizing $U(\phi, \alpha_\infty, \Lambda, E_b, \nu_b)$ defined by

$$U(\phi, \alpha_\infty, \Lambda, E_b, \nu_b) = \sum_{i=1}^{i=n} [p^t_{\exp}(x,t_i) - p^t(x,t_i)]^2, \quad (7)$$

wherein $p^t_{\exp}(x,t)_{i=1,2,\ldots,n}$ is the discrete set of values of the experimental transmitted signal and $p^t(x,t_i)_{i=1,2,\ldots,n}$ the discrete set of values of the simulated transmitted signal. The next section deals with the solution of the inverse problem based on experimental transmitted data. For the iterative solution of the inverse problem, we used the simplex search method (Nedler Mead)[31] which does not require numerical or analytic gradients.

IV. ULTRASONIC MEASUREMENTS

As an application of this model, some numerical simulations are compared with experimental results. Experiments are performed in water using two broadband Panametrics A 303S plane piezoelectric transducers with a central frequency

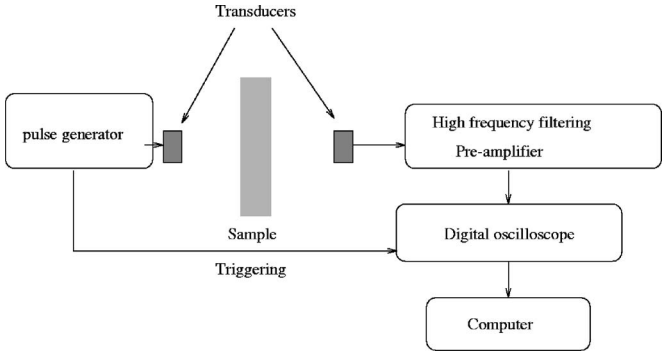


FIG. 8. Experimental setup for ultrasonic measurements.

of 1 MHz in water, and diameter of 1 cm. 400 V pulses are provided by a 5058PR Panametrics pulser/receiver. Electronic interference is removed by averaging 1000 acquisitions. The experimental setup is shown in Fig. 8. The parallel-faced samples were machined from femoral heads and femoral necks of human cancellous bone. The liquid in the pore space (blood and marrow) is removed from the bone sample and substituted by water. The size of the ultrasound beam is very small compared to the size of the specimens. The emitting transducer insonifies the sample at normal incidence with a short (in time domain) pulse. When the pulse hits the front surface of the sample, a part is reflected, a part is transmitted as a fast wave, and a part is transmitted as a slow wave. When any of these components, traveling at different speeds, hit the second surface, a similar effect takes place: a part is transmitted into the fluid, and a part is reflected as a fast or slow wave. The experimental transmitted wave forms are traveling through the cancellous bone in the same direction as the trabecular alignment (x direction). The fluid characteristics[26] are: bulk modulus $K_f=2.28$ GPa, density $\rho_f=1000$ kg m^{-3}, viscosity $\eta=10^{-3}$ kg m s^{-1}.

Consider a sample of human cancellous bone $M1$ (femoral neck) of thickness 11.2 mm and solid density $\rho_s=1990$ kg m^{-3}. The Young's modulus $E_s=13$ GPa and Poisson ratio $\nu_s=0.3$ of the solid bone are taken from the literature.[8] Figure 9 shows the experimental incident signal. The inverse problem is solved by minimizing the function $U(\phi,\alpha_\infty,\Lambda,E_b,\nu_s)$ given by Eq. (7). A large variation range is applied of each estimating parameter value in solving the

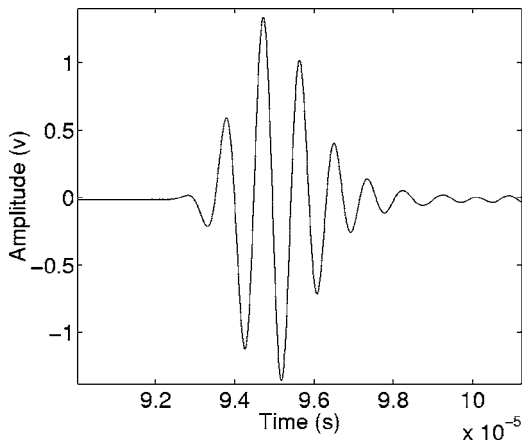


FIG. 9. Experimental incident signal for bone sample $M1$.

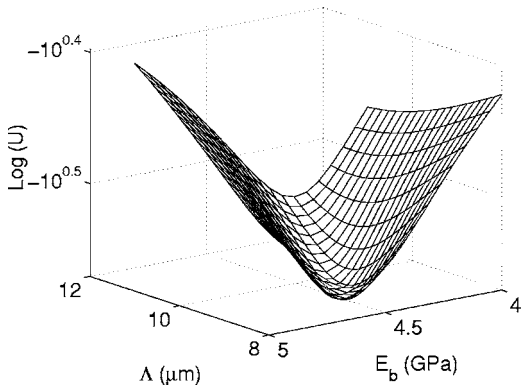


FIG. 10. Variation of the minimization function U with the viscous characteristic length Λ and the Young modulus of the skeletal frame E_b.

inverse problem. The variation range of the parameters is: $\alpha_\infty\in[1,2],\Lambda\in[1,200]$ μm, $\phi\in[0.5,0.99]$, $\nu\in[0.1,0.5]$, and $E_b\in[0.5,5]$GPa. The variations of the cost function with the physical parameters present one clear minimum corresponding to the mathematical solution of the inverse problem. This shows that the inverse problem is well posed mathematically, and that the solution is unique. The minima, corresponding to the solution of the inverse problem, are clearly observed for each parameter. After solving the inverse problem, we find the following optimized values: $\phi=0.64$, $\alpha_\infty=1.018$, $\Lambda=10,44$ μm, $\nu_b=0.28$, and $E_b=4.49$ GPa. Using these values, we present in Figs. 10–14 the variations in the discrepancy function U with respect to two values of the inverted parameters. For showing clearly the solution of the inverse problem, the variation of U in Figs. 10–14 is given only around the minima values of the inverted parameters. In Fig. 15, a comparison is made between the experimental transmitted signal and the simulated transmitted signals using the reconstructed values of $\alpha_\infty,\phi,\Lambda,\nu_b$, and E_b. The difference between the two curves is small, which leads us to conclude that the optimized values of the physical parameters are correct. The fast and slow waves predicted by the Biot theory are easily detected in the transmitted signal. The slow wave seems to be less attenuated than the fast wave. In the other applications,[32] the slow wave is generally more attenuated and dispersive than the fast wave. We usually observe the opposite phenomena for

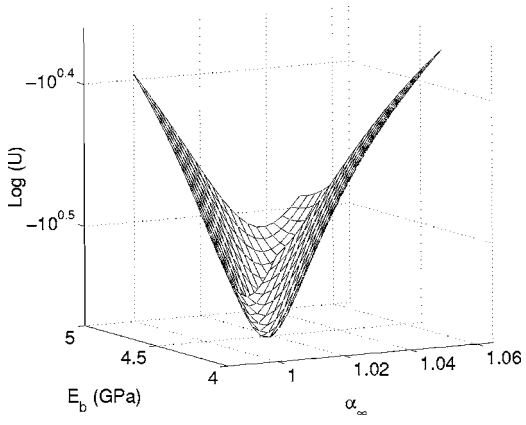


FIG. 11. Variation of the minimization function U with the Young modulus of the skeletal frame E_b and the tortuosity α_∞.

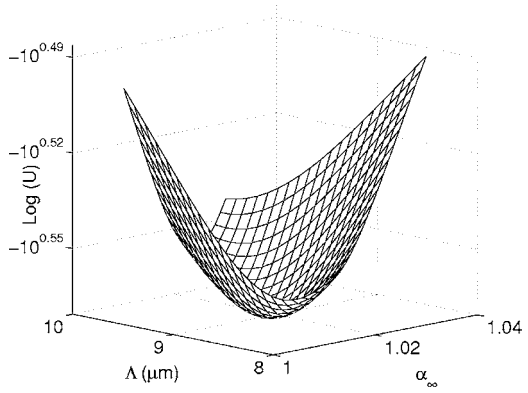


FIG. 12. Variation of the minimization function U with the viscous characteristic length Λ and the tortuosity α_∞.

cancellous bone samples; this can be explained by the different orders of magnitude of the physical parameters (high porosity, low tortuosity, etc.).

Let us now solve the inverse problem for sample $M2$ (femoral neck) of thickness 12 mm. By solving the inverse problem, the obtained optimized values are: $\phi=0.79$, $\alpha_\infty=1.052$, $\Lambda=10,12$ μm, $\nu_b=0.25$, and $E_b=2.47$ GPa. Figure 16 shows a comparison of experimental transmitted signal and a simulated signal obtained by optimization after solving the inverse problem. Here, again, the correlation between theoretical predictions and experimental data is satisfactory.

Using another sample of cancellous bone (femoral head) $M3$ of thickness 10.2 mm. The results after solving the inverse problem are: $\phi=0.72$, $\alpha_\infty=1.1$, $\Lambda=14.97$ μm, $\nu_b=0.22$, and $E_b=3.1$ GPa. In Fig. 17, we compare the experimental transmitted signal to the transmitted simulated signal using reconstruct values of the physical parameters. The correlation between the two curves is excellent.

In a second step, the bone samples are drained and their physical parameters (ϕ, α_∞, Λ, E_b, and ν_b) are measured by techniques[20,33,34] developed initially for air-saturated porous materials such as plastic foams and fibrous mats. When the liquid saturating the cancellous bone is drained from the pores and replaced by air, partial decoupling of the Biot waves occurs[20,28] due to the tremendous difference in density between the frame and air. The fluid particles do not have enough mass to generate motion in the heavy solid frame, and thus the slow wave propagates in the fluid wherein it is

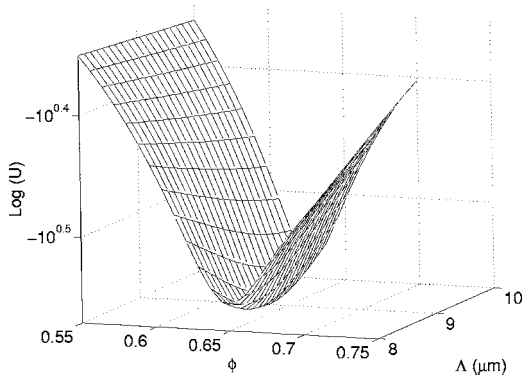


FIG. 13. Variation of the minimization function U with the porosity ϕ and the viscous characteristic length Λ.

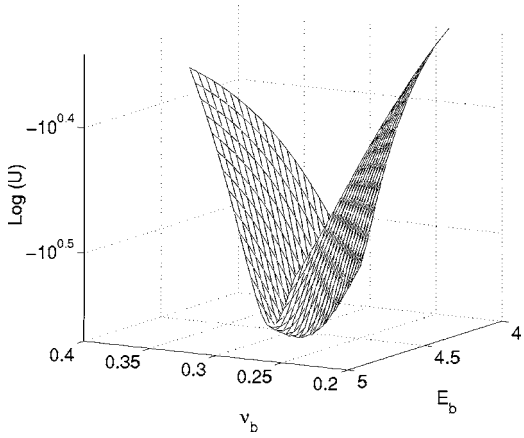


FIG. 14. Variation of the minimization function U with the Poisson ratio ν_b and the Young modulus of the skeletal frame E_b.

detected (in noncontact manner) by a transducer. The three parameters; α_∞, ϕ, and Λ are determined by measuring the slow wave propagating in air-saturated cancellous bone. For example, the porosity ϕ and the tortuosity α_∞ are determined by measuring the wave reflected by the first interface of the bone sample at oblique incidence.[33] The viscous characteristic length Λ is evaluated by measuring the transmitted wave.[34] With contact excitation,[20] the fast wave travels in the solid frame and some air particles move along with the frame. The velocity of the fast wave approaches the velocity in the frame as measured in vacuum and is given by $v_L=\sqrt{(K_b+4/3N)/(1-\phi)\rho_s}$. By measuring the fast wave velocity of a sample whose pores are filled with air one finds $K_b+4/3N$. The shear modulus N can be evaluated independently by measuring the velocity of the shear wave. The expression of the shear wave velocity is given by[28] $v_T=\sqrt{N/(1-\phi)\rho_s}$. By measuring experimentally v_L and v_T, we deduce K_b and N, and then the values of E_b and ν_b using the relations (3). The experimental values of the longitudinal and transverse velocities v_L and v_T, for the bone samples $M1$, $M2$, and $M3$, and their deduced values of E_b and ν_b are reported in the Table I.

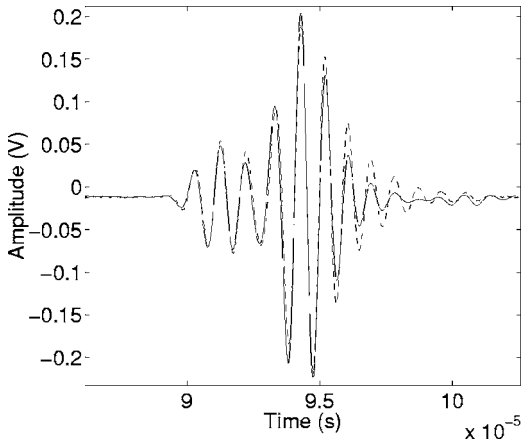


FIG. 15. Comparison between the experimental transmitted signal (solid line) and the simulated transmitted signals (dashed line) using the reconstructed values of α_∞, ϕ, Λ, ν_b, and E_b (sample $M1$).

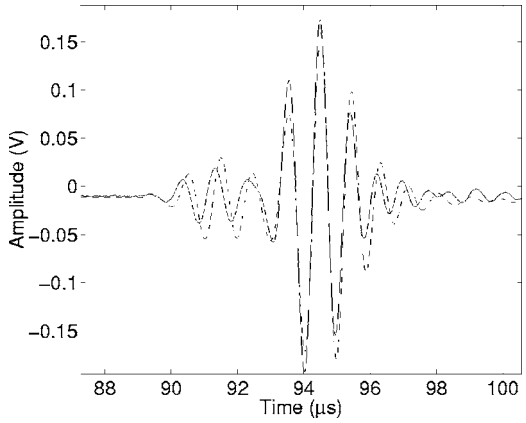


FIG. 16. Comparison between the experimental transmitted signal (solid line) and the simulated transmitted signals (dashed line) using the reconstructed values of α_∞, ϕ, Λ, ν_b, and E_b (sample $M2$).

A comparison between the optimized values of ϕ, α_∞, Λ, ν_b, and E_b obtained by solving the inverse problem, and those evaluated with drained bone samples is given, for the three samples $M1$, $M2$, and $M3$ in the Table II. It can seen that the optimized values obtained by solving the inverse problem are close to those obtained for the drained bone samples. Except for the viscous characteristic length value of the specimen $M2$, for which the discrepancy is of 50%. To note that the viscous characteristic length is the most difficult parameter to obtain with a good precision. This parameter depends strongly of the attenuation of transmitted wave by drained cancellous bone, which is very important in air. The simulated transmitted signals obtained using optimized values (Figs. 15 and 17) reproduce correctly the experimental transmitted signals. This leads us to conclude that this method is well adapted for the characterization of cancellous bone.

In the former studies, McKelvie[18,21] demonstrated that predictions by the Biot theory agreed better with experimental results obtained from calcaneus than predictions made by scattering theories. The authors[18,21] predict correctly the acoustic attenuation but not the trend in ultrasound velocity. Conversely, Williams[19] found good results for the prediction of the fast wave velocity using a limited formulation of the

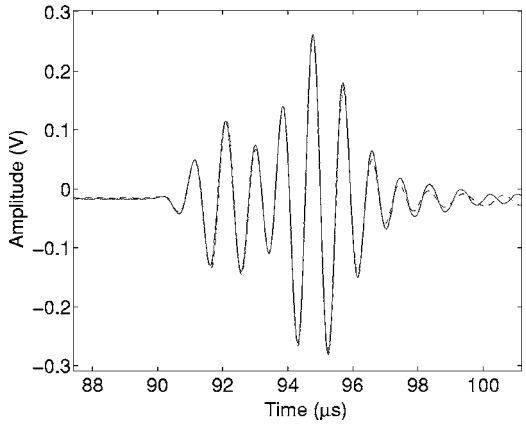


FIG. 17. Comparison between the experimental transmitted signal (solid line) and the simulated transmitted signals (dashed line) using the reconstructed values of α_∞, ϕ, Λ, ν_b, and E_b (sample $M3$).

TABLE I. Experimental values of the longitudinal and transverse velocities v_L and v_T, for the air-saturated bone samples $M1$, $M2$, and $M3$, and their deduced values of E_b and ν_b.

Cancellous bone samples	$M1$	$M2$	$M3$
Longitudinal wave velocity v_L (m/s)	3031	2797	2149
Shear wave velocity v_T (m/s)	1573	1486	1381
Young modulus of the skeletal frame: E_b (GPa)	4.65	2.4	2.59
Poisson ratio of the skeletal frame: ν_b	0.31	0.24	0.21

Biot theory from tibial and femoral bovine cancellous bone samples. Williams[19] expanded his formulation of the Biot theory to consider attenuation using the formulation of dynamic tortuosity[25] (used in this paper). Good agreement for the fast wave velocity was again found. For attenuation, although the predicted trends were similar to those observed experimentally in cancellous bone, the experimental values were considerably higher than those predicted by the Biot theory. Hosokawa and Otani[23] obtained better results for the wave velocities (fast and slow) than for the acoustic attenuation. In most of these studies, the authors do not take into account the losses dues to the reflections at the interfaces by considering the reflection and transmission by a slab of cancellous bone. In addition, to our knowledge, the inverse problem was not considered for the determination of the physical parameters from experimental data.

In our previous paper,[26] we attempted to apply the modified[25] Biot theory for solving the direct problem of the ultrasonic propagation in human cancellous bone. The direct scattering problem involves determining the scattered field as well as the internal field, arising when a known incident field impinges on a cancellous bone with known physical properties. The reflected and transmitted fields are deduced from the internal field and boundary conditions. In the present paper, we attempt to solve the inverse problem for determining the value of the medium's parameters using the incident and transmitted experimental data. In this characterization problem, the losses due to reflections at the interfaces of the samples, and those due to viscous exchange between fluid and structure are taken into account. The comparison between experimental and theoretical attenuations and waves velocities (fast and slow) are given simultaneously in time domain using the transmitted signals. The results obtained in this study are encouraging for the ultrasonic characterization of cancellous bone, we will try in the future to consider the experimental reflected waves.

V. CONCLUSION

In this paper the characterization of cancellous bone is treated by solving the inverse problem numerically using experimental transmitted signals. Five physical parameters (porosity ϕ, tortuosity α_∞, viscous characteristic length Λ, Poisson ratio ν_b, and Young modulus of the skeletal frame E_b) are inverted. The sensitivity analysis of v_b and E_b was studied in this paper, showing the importance of the values of these parameters in fast and slow wave arrival times (speeds) and attenuation, respectively. The optimized values of the physi-

TABLE II. Comparison between the optimized parameters of the bone samples $M1$, $M2$, and $M3$ obtained by solving the inverse problem and those evaluated with drained bone samples.

Parameters and methods	E_b (GPa)	ν_b	ϕ	α_∞	$\Lambda(\mu\text{m})$
Inverse problem ($M1$)	4.49	0.28	0.64	1.018	9.1
Drained bone ($M1$)	4.65	0.31	0.71	1.02	10.44
Inverse problem ($M2$)	2.47	0.25	0.79	1.052	10.12
Drained bone ($M2$)	2.4	0.24	0.75	1.045	15
Inverse problem ($M3$)	3.1	0.22	0.64	1.1	14.97
Drained bone ($M3$)	2.59	0.21	0.59	1.08	19.5

cal parameters are compared with those obtained with techniques developed initially for air-saturated porous material (drained bone) given a good results. The comparison between experiment and theory validate the proposed method.

APPENDIX: EXPRESSION OF THE TRANSMISSION COEFFICIENT

The expression of the transmission coefficient is given by[26]

$$\mathcal{T}(\omega)=\frac{j\omega 2\rho_f c_0 F_4(\omega)}{[j\omega\rho_f c_0 F_4(\omega)]^2-[j\omega F_3(\omega)-1]^2},$$

where

$$F_i(\omega)=\{1+\phi[\mathfrak{I}_i(\omega)-1]\}\times\sqrt{\lambda_i(\omega)}\frac{\Psi_i(\omega)}{\sinh(l\sqrt{\lambda_i(\omega)})}\frac{2}{\Psi(\omega)},\quad i=1,2,$$

$$F_3(\omega)=\rho_f c_0\{F_1(\omega)\cosh[l\sqrt{\lambda_1(\omega)}]+F_2(\omega)\cosh[l\sqrt{\lambda_2(\omega)}]\},$$

$$F_4(\omega)=F_1(\omega)+F_2(\omega).$$

The functions $\lambda_1(\omega)$ and $\lambda_2(\omega)$ are given by

$$\lambda_1(\omega)=\tfrac{1}{2}\Big[-\tau_1\omega^2+\tau_2(j\omega)^{3/2}-\sqrt{(\tau_1^2-4\tau_3)\omega^4+2(\tau_1\tau_2-2\tau_4)(j\omega)^{7/2}+\tau_2^2(j\omega)^3}\Big],$$

$$\lambda_2(\omega)=\tfrac{1}{2}\Big[-\tau_1\omega^2+\tau_2(j\omega)^{3/2}+\sqrt{(\tau_1^2-4\tau_3)\omega^4+2(\tau_1\tau_2-2\tau_4)(j\omega)^{7/2}+\tau_2^2(j\omega)^3}\Big],$$

with

$$\tau_1=R'\rho_{11}+P'\rho_{22}-2Q'\rho_{12},\quad \tau_2=A(P'+R'+2Q'),$$

$$\tau_3=(P'R'-Q'^2)(\rho_{11}\rho_{22}-\rho_{12}^2),$$

$$\text{and}\quad \tau_4=A(P'R'-Q'^2)\,(\rho_{11}+\rho_{22}-2\rho_{12}).$$

Coefficients R', P', and Q' are given by

$$R'=\frac{R}{PR-Q^2},\quad Q'=\frac{Q}{PR-Q^2},\quad \text{and}\quad P'=\frac{P}{PR-Q^2}.$$

The functions $\mathfrak{I}_1(\omega)$ and $\mathfrak{I}_2(\omega)$ are given by

$$\mathfrak{I}_1(\omega)=\frac{(2\tau_5-\tau_1)\omega^2+(\tau_2-2\tau_6)(j\omega)^{3/2}-\sqrt{(\tau_1^2-4\tau_3)\omega^4+2(\tau_1\tau_2-2\tau_4)(j\omega)^{7/2}+\tau_2^2(j\omega)^3}}{2[-\tau_7\omega^2-\tau_6(j\omega)^{3/2}]},$$

$$\mathfrak{I}_2(\omega)=\frac{(2\tau_5-\tau_1)\omega^2+(\tau_2-2\tau_6)(j\omega)^{3/2}+\sqrt{(\tau_1^2-4\tau_3)\omega^4+2(\tau_1\tau_2-2\tau_4)(j\omega)^{7/2}+\tau_2^2(j\omega)^3}}{2[-\tau_7\omega^2-\tau_6(j\omega)^{3/2}]},$$

where

$$\tau_5=(R'\rho_{11}-Q'\rho_{12})\quad \tau_6=A(R'+Q'),$$

$$\tau_7=(R'\rho_{12}-Q'\rho_{22}).$$

The coefficients $\Psi_1(\omega),\Psi_2(\omega)$, and $\Psi(\omega)$ are given by

$$\Psi_1(\omega)=\phi Z_2(\omega)-(1-\phi)Z_4(\omega),$$

$$\Psi_2(\omega)=(1-\phi)Z_3(\omega)-\phi Z_1(\omega),$$

$$\Psi(\omega)=2[Z_1(\omega)Z_4(\omega)-Z_2(\omega)Z_3(\omega)],$$

and the coefficients $Z_1(\omega)$, $Z_2(\omega)$, $Z_3(\omega)$, and $Z_4(\omega)$ by

$$Z_1(\omega)=[P+Q\mathfrak{I}_1(\omega)]\lambda_1(\omega),$$

$$Z_2(\omega)=[P+Q\mathfrak{I}_2(\omega)]\lambda_2(\omega),$$

$$Z_3(\omega)=[Q+R\mathfrak{I}_1(\omega)]\lambda_1(\omega),$$

$$Z_4(\omega)=[Q+R\mathfrak{I}_2(\omega)]\lambda_2(\omega).$$

[1]A. M. Parfitt, "Trabecular bone architecture in the pathogenesis and prevention of fracture," Am. J. Med. **82** (1B), 68–72 (1987).
[2]S. M. Ott, R. F. Kilcoyne, and C. Chestnut III, "Ability of four different techniques of measuring bone mass to diagnose vertebral fractures in postmenopausal women," J. Bone Miner. Res. **2**, 201–210 (1987).
[3]J. M. Alves, W. Xu, D. Lin, R. S. Siffert, J. T. Ryaby, and J. J. Kaufman, "Ultrasonic assessment of human and bovine trabecular bone: A comparison study," IEEE Trans. Biomed. Eng. **43**(3), 249–258 (1996).
[4]C. C. Gluer, "Quantitative ultrasound techniques for the assessment os osteoporosis: Expert agreement on current status," J. Bone Miner. Res. **12** (8), 1280–1288 (1997).
[5]F. J. Fry and J. E. Barger, "Acoustical properties of the human skull," J. Acoust. Soc. Am. **63**, 1576–1590 (1978).
[6]J. Y. Rho, "An ultrasonic method for measuring the elastic properties of human tibial cortical and cancellous bone," Ultrasonics **34**, 777–783 (1996).
[7]R. B. Ashman, J. D. Corin, and C. H. Turner, "Elastic properties of cancellous bone: Measurement by an ultrasonic technique," J. Biomech. **10**, 979–989 (1987).
[8]R. B. Ashman and J. Y. Rho, "Elastic modulus of trabecular bone material," J. Biomech. **21**, 177–181 (1988).
[9]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).
[10]E. R. Hughes, T. G. Leighton, G. W. Petley, and P. R. White, "Ultrasonic propagation in cancellous bone: A new stratified model," Ultrasound Med. Biol. **25**, 811–821 (1999).
[11]F. Padilla and P. Laugier, "Phase and group velocities of fast and slow compressional waves win trabecular bone," J. Acoust. Soc. Am. **108** (4), 1949–1952 (2000).
[12]F. Luppe, J. M. Conoir, and H. Franklin, "Scattering by a fluid cylinder in a porous medium: Application to trabecular bone," J. Acoust. Soc. Am. **111**, 2573–2582 (2002).
[13]S. Chaffai, V. Roberjot, F. Peyrin, G. Berger, and P. Laugier, "Frequency dependence of ultrasonic backscattering in cancellous bone: Autocorrelation model and experimental results," J. Acoust. Soc. Am. **108**, 2403–2411 (2000).
[14]K. A. Wear, "Frequency dependence of ultrasonic backscatter from human trabecular bone: Theory and experiment," J. Acoust. Soc. Am. **106**, 3659–3664 (1999).
[15]M. Schoenberg, "Wave propagation in alternating solid and fluid layers," Wave Motion **6**, 303–321 (1984).
[16]M. A. Biot, "Generalized theory of acoustic propagation in porous dissipative media," J. Acoust. Soc. Am. **34**(4), 1254–1264 (1962).
[17]M. A. Biot, "The theory of propagation of elastic waves in fluid-saturated porous solid. I. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[18]M. L. McKelvie, "Ultrasonic propagation in cancellous bone," Ph.D. thesis, Hull, UK, University of Hull, 1988.
[19]J. L. Williams, "Ultrasonic wave propagation in cancellous bone and cortical bone: Prediction of some experimental results by Biot's theory," J. Acoust. Soc. Am. **91**, 1106–1112 (1992).
[20]W. Lauriks, J. Thoen, I. Van Asbroeck, G. Lowet, and G. Vanderperre, "Propagation of ultrasonic pulses through trabecular bone," J. Phys. (Paris), Colloq. **4**, 1255–1258 (1994).
[21]M. L. Mckelvie and S. B. Palmer, "The interaction of ultrasound with cancellous bone," Phys. Med. Biol. **36**(10), 1331–1340 (1991).
[22]T. J. Haire and C. M. Langton, "Biot theory: A review of its application on ultrasound propagation through cancellous bone," Bone (N.Y.) **24** (4), 291–295 (1999).
[23]A. Hosokawa and T. Otani, "Ultrasonic wave propagation in bovine cancellous bone," J. Acoust. Soc. Am. **101**, 558–562 (1997).
[24]A. Hosokawa and T. Otani, "Acoustic anisotropy in bovine cancellous bone," J. Acoust. Soc. Am. **103**, 2718–2722 (1998).
[25]D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).
[26]Z. E. A. Fellah, J. Y. Chapelon, S. Berger, W. Lauriks, and C. Depollier, "Ultrasonic wave propagation in human cancellous bone: Application of Biot theory," J. Acoust. Soc. Am. **116** (1), 61–73 (2004).
[27]D. L. Johnson, T. J. Plona, and H. Kojima, "Probing porous media with first and second sound. II. Acoustic properties of water-saturated porous media," J. Appl. Phys. **76** (1), 115–125 (1994).
[28]J. F. Allard, *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Chapman and Hall, London, 1993).
[29]N. P. Chortiros, "An inversion for Biot parameters in water-saturated sand," J. Acoust. Soc. Am. **112** (5), 1853–1868 (2002).
[30]J. L. Buchanan, R. P. Gilbert, and K. Khashanah, "Recovery of the poroelastic parameters of cancellous bone using low frequency acoustic interrogation," in *Acoustics, Mechanics, and the Related Topics of Mathematical Analysis*, edited by A. Wirgin (World Scientific, Singapore, 2002), pp. 41–47.
[31]J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder-Mead simplex method in low dimensions," SIAM J. Optim. **9**, 112–147 (1998).
[32]T. J. Plona, "Observation of a second bulk compressional wave in a porous medium at ultrasonic frequencies," Appl. Phys. Lett. **36** (4), 259–261 (1980).
[33]Z. E. A. Fellah, S. Berger, W. Lauriks, C. Depollier, C. Aristegui, and J. Y. Chapelon, "Measuring the porosity and tortuosity of porous materials via reflected waves at oblique incidence," J. Acoust. Soc. Am. **113** (5), 2424–2433 (2003).
[34]Z. E. A. Fellah, M. Fellah, W. Lauriks, and C. Depollier, "Direct and inverse scattering of transient acoustic waves by a slab of rigid porous material," J. Acoust. Soc. Am. **113**, 61–73 (2003).

Theory of sound propagation from a moving source in a three-layer Pekeris waveguide

Michael J. Buckingham[a] and Eric M. Giddens
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0238

(Received 2 March 2006; revised 29 June 2006; accepted 30 June 2006)

A theory is developed for the acoustic field in a three-layer waveguide, representing the atmosphere, shallow ocean and sediment. The unaccelerated source is moving horizontally in the atmosphere. Two solutions are presented. The first, for a line source normal to the direction of travel, is a single wavenumber integral yielding the two-dimensional (2-D) field in each layer; and the second, for a point source, is a double wavenumber integral for the 3-D field in each layer. In both cases, the moving-source dispersion relationship for the three-layer environment is derived. From the 2-D dispersion relation, asymmetries fore and aft of the source, due to source motion, are shown to exist in the field in all three layers. In the water column, complex Doppler effects modify asymmetrically the effective depth of the channel and hence also the mode shapes. Evidence of the fore-aft modal asymmetry appears in the high-resolution spectrum of the field in the channel, which exhibits several sharp peaks on either side of the unshifted frequency, each associated with an up- or downshifted mode. A numerical evaluation of the 3-D solution provides a graphic illustration of the asymmetrical character of the field in all three layers.
[DOI: 10.1121/1.2258095]

PACS number(s): 43.20.Bi, 43.30.Bp, 43.30.Es [AIT] Pages: 1825–1841

I. INTRODUCTION

The effects of source motion on the propagation of sound in an infinite, homogeneous medium have been examined by a number of authors, for instance, Lowson,[1] Leppington and Levine,[2] Morse and Ingard,[3] Pierce,[4] and, of course, Johann Doppler,[5] for whom the frequency shift due to source motion is named. Sound transmission from moving sources in inhomogeneous media has also received considerable attention, particularly in connection with the deep ocean,[6–8] where the sound speed profile creates an acoustic waveguide known as the deep sound channel. Indeed, differential Doppler between individual peaks in very-long-range (several thousand kilometer) ocean-acoustic transmissions through the deep sound channel has been exploited by Dzieciuch and Munk[9] as a means of identifying ray arrivals.

Doppler effects in shallow-ocean waveguides have been addressed by several authors, including Jacyna *et al.*[10], who adopted a ray argument to interpret the acoustic field in a channel with a depth-dependent sound speed profile. Neubert[11] took a different approach, using normal modes to investigate the effect of Doppler on long-range acoustic tracking in oceanic channels; and a normal-mode theory of acoustic Doppler effects in shallow oceanic waveguides has been developed by Hawker.[12] A more general problem was considered by Schmidt and Kuperman,[13] who derived a spectral representation (wavenumber integral) for the field from a horizontally moving source in a stratified oceanic waveguide with multiple layers. As examples of their modeling technique, they discuss a moving source and moving receiver in a shallow water channel overlying a sedimentary half-space, and also acoustic emissions from the Arctic ice cover.

It is implicit in most discussions of Doppler phenomena in the ocean that the moving source is either on or beneath the sea surface. Such sources generally move slowly relative to the speed of sound in seawater, with typical Mach numbers of 0.01 or less, and hence the associated Doppler effects are small if not negligible. By comparison, an airborne acoustic source such as a light aircraft may produce more significant Doppler effects, since the Mach number (in air) is typically around 0.15, which is at least an order of magnitude higher than that of a water-borne source.

From the early observations of Urick[14] and Medwin,[15] and more recently the measurements of Richardson *et al.*,[16] it is known that sound from an aircraft couples into the ocean and is detectable on a submerged receiver. Moreover, Doppler effects in aircraft sound in both the atmosphere and the ocean have been observed by Ferguson,[17,18] who made measurements of the 68 Hz propeller-blade harmonic from a turboprop, fixed-wing aeroplane as it flew over a microphone mounted just above ground level and a hydrophone at a depth of 20 m beneath the sea surface. On the microphone, the difference-frequency between the upshifted tone on approach and downshifted tone on departure was found to be a factor of about 4.5 greater than the corresponding difference frequency on the hydrophone. This is consistent with a simple ray argument, combined with Snell's law for sound penetrating the air-sea boundary, which shows that the approach-departure difference frequency scales inversely with the sound speed of the medium in which the observation is made. Thus, the 4.5:1 ratio of the difference frequen-

[a]Also affiliated with the Institute of Sound and Vibration Research, The University, Southampton SO17 1BJ, England. Electronic mail: mjb@mpl.ucsd.edu

cies observed by Ferguson[17,18] on the microphone and hydrophone is, in fact, a measure of the sound speed in seawater relative to that in air.

The relationship between the local speed of sound and the Doppler difference frequency in a stratified medium has been exploited recently by Buckingham *et al.*[19,20] as a means of measuring the speed of sound in the sediment immediately beneath the sea floor. In their experiments, the Doppler-shifted engine and propeller harmonics from a single-engine, light aircraft were detected on acoustic sensors located in the atmosphere, the ocean and the sediment. An inversion technique, designated Doppler spectroscopy, was then used to return the speed of sound in each of the three layers, along with a number of other parameters, including the altitude and speed of the aircraft.

In the early experiments,[19,20] the Doppler spectroscopy inversions were based on geometrical ray theory. Our purpose in this paper is to improve on the rudimentary ray-based analysis by developing a full wave-theoretic (forward) model of the sound field from a moving airborne source in a stratified, three-layer environment (atmosphere-ocean-sediment). It will be shown that, in all three layers, Doppler effects give rise to significant asymmetries in the acoustic field fore and aft of the source. These asymmetric features are characterized in some detail in the discussion.

The basis of the model is as follows. A harmonic airborne source (an aircraft) is in horizontal, unaccelerated motion in a homogeneous, isotropic, semi-infinite atmosphere. Beneath the atmosphere, the ocean is treated as a classic Pekeris[21] channel, with flat surface, uniform depth, and constant sound speed. Below the water column is a homogeneous, isotropic semi-infinite fluid sediment.

In fact, two closely related wave-theoretic models are discussed in the paper. To investigate the essential physics of a moving source in a three-layer fluid medium, a two-dimensional, analytical model is constructed, giving the range, depth, and time dependence of the field from a moving, airborne *line source* oriented normal to the direction of aircraft motion. This 2-D model yields a solution for the field in each layer in the form of a *single* wavenumber integral. Simple algebraic approximations are developed for these integrals, which provide valuable insight into the asymmetric structure of the field in each layer fore and aft of the source. Of course, asymmetries similar to those in the 2-D field are also present in the 3-D field from a moving *point source* in the atmosphere, which is treated in the second model. In the 3-D case, the field solutions are in the form of *double* wavenumber integrals. These field integrals are evaluated numerically to provide graphic illustrations of the fore-aft asymmetry that exists in all three layers.

It is perhaps worth commenting on the *double-integral* formulation of the 3-D field from the moving point source. For the special case of a *stationary* source, the fore-aft asymmetries vanish, the 3-D field is azimuthally uniform, varying with only two spatial coordinates, range and depth, in which case the solution in each layer may be expressed as a *single* wavenumber integral. This is the form of solution developed by Pekeris[21] for his two-layer ocean-sediment problem. When source *motion* is present, however, the axial symmetry is broken and the field is no longer azimuthally uniform. The 3-D field then depends on three spatial coordinates, giving rise to a solution in each layer having the form of a *double* wavenumber integral.

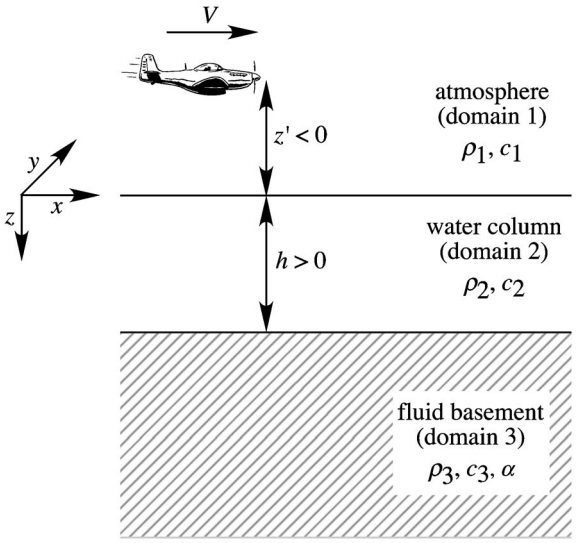


FIG. 1. Three-layer (atmosphere-ocean-sediment) waveguide, showing the Cartesian coordinate system and moving, unaccelerated sound source at constant altitude z' in the atmosphere (not to scale).

II. SOLUTION FOR THE 2-D DOPPLER-SHIFTED FIELD IN A THREE-LAYER WAVEGUIDE

Figure 1 shows the three-layer geometry to be considered in the following analysis. Regions 1, 2, and 3 represent the atmosphere, seawater, and sediment, each with density ρ_i and sound speed c_i, $i=1,2,3$. In the horizontal, x is the range coordinate in the direction of source motion and y is the direction normal to the source track. Depth (or altitude) is z, increasing downward, with the origin $z=0$ at the sea surface. The airborne source is at fixed altitude $z=z'<0$ and the seabed is at a uniform depth $z=h>0$. A stationary, point receiver at $(x,0,z)$ may be in any of the three layers. The source, moving horizontally with constant (positive) speed V in the positive x direction, is at horizontal range x' when time $t=0$.

For the 2-D problem, the source is treated as an infinite horizontal line normal to the source track, in which case symmetry dictates that the field is everywhere independent of y. Assuming a monotonic time dependence, possibly representative of an aircraft engine or propeller harmonic, of (unshifted) angular frequency Ω, the 2-D wave equation is written in Cartesian coordinates for each of the three layers as follows:

$$\frac{\partial^2\phi_1}{\partial x^2}+\frac{\partial^2\phi_1}{\partial z^2}-\frac{1}{c_1^2}\frac{\partial^2\phi_1}{\partial t^2}=-Q_L\delta(z-z') \times\delta(x-x'-Vt)e^{i\Omega t},\quad z<0, \tag{1}$$

$$\frac{\partial^2\phi_2}{\partial x^2}+\frac{\partial^2\phi_2}{\partial z^2}-\frac{1}{c_2^2}\frac{\partial^2\phi_2}{\partial t^2}=0,\quad 0\leqslant z\leqslant h, \tag{2}$$

$$\frac{\partial^2\phi_3}{\partial x^2}+\frac{\partial^2\phi_3}{\partial z^2}-\frac{1}{c_3^2}\frac{\partial^2\phi_3}{\partial t^2}=0,\quad z>h, \tag{3}$$

where $\phi_i=\phi_i(x,z,t)$ is the velocity potential in the ith layer, Q_L is the source strength per unit length [with dimensions of (length)^2/time] and $\delta(\cdots)$ is the Dirac delta function. The full time-dependent solution comes from adding a conjugate source term to the right of Eq. (1) or, equivalently, by taking the real part of the final expression for the field. The three wave equations are to be solved subject to the boundary conditions, namely that the pressure and normal component of particle velocity should be continuous across the sea surface and the seabed:

$$\rho_1\phi_1(0)=\rho_2\phi_2(0), \tag{4}$$

$$\phi_1'(0)=\phi_2'(0), \tag{5}$$

$$\rho_2\phi_2(h)=\rho_3\phi_3(h), \tag{6}$$

$$\phi_2'(h)=\phi_3'(h), \tag{7}$$

where the abbreviated notation $\phi_i(u)=\phi_i(x,u,t)$ and $\phi_i'(u)=\partial\phi_i(x,z,t)/\partial z|_{z=u}$ has been introduced.

Following an analysis similar to that in Buckingham and Giddens,[22] the solutions of the wave equations are obtained using standard integral transform techniques. The exact result for the field in each of the three layers takes the form of a wavenumber integral:

$$\phi_1(x,z,t)=\frac{Q_L e^{i\Omega t}}{8\pi^2 i}\int_{-\infty}^{\infty}e^{ip(x-x'-Vt)}F_1(\eta_1,\eta_2,\eta_3)dp,\quad z<0, \tag{8}$$

$$\phi_2(x,z,t)=\frac{Q_L b_{12} e^{i\Omega t}}{4\pi^2 i}\int_{-\infty}^{\infty}e^{ip(x-x'-Vt)}F_2(\eta_1,\eta_2,\eta_3)dp,\quad 0\leqslant z\leqslant h, \tag{9}$$

$$\phi_3(x,z,t)=\frac{Q_L b_{13} e^{i\Omega t}}{4\pi^2 i}\int_{-\infty}^{\infty}e^{ip(x-x'-Vt)}F_3(\eta_1,\eta_2,\eta_3)dp,\quad z>h, \tag{10}$$

where the integration variable, p, is the horizontal wavenumber and

$$\left.\begin{aligned}\eta_i&=\sqrt{(k_i-p\beta_i)^2-p^2}\\ k_i&=\frac{\Omega}{c_i}\\ \beta_i&=\frac{V}{c_i}\end{aligned}\right\},\quad i=1,2,3. \tag{11}$$

Thus, (η_i,k_i,β_i) are the Doppler-shifted vertical wavenumber, the acoustic wavenumber, and the Mach number for layer $i=1,2$, or 3. Note that the presence of source motion introduces an asymmetry into the vertical wavenumbers, η_i, in that they are mixed functions of p, rather than even functions, as they would be if V were zero. The functions F_i in the integrands of Eqs. (8) and (10) are

$$F_1(\eta_1,\eta_2,\eta_3)=\frac{e^{-i\eta_1|z-z'|}}{\eta_1}+\frac{e^{-i\eta_1|z+z'|}}{\eta_1}\left\{\frac{\eta_2(\eta_1-b_{13}\eta_3)\cos(\eta_2 h)-i(b_{12}\eta_2^2-b_{23}\eta_1\eta_3)\sin(\eta_2 h)}{\eta_2(\eta_1+b_{13}\eta_3)\cos(\eta_2 h)+i(b_{12}\eta_2^2+b_{23}\eta_1\eta_3)\sin(\eta_2 h)}\right\}, \tag{12}$$

$$F_2(\eta_1,\eta_2,\eta_3)=e^{i\eta_1 z'}\left\{\frac{\eta_2\cos\,\eta_2(h-z)+ib_{23}\eta_3\sin\,\eta_2(h-z)}{\eta_2(\eta_1+b_{13}\eta_3)\cos(\eta_2 h)+i(b_{23}\eta_1\eta_3+b_{12}\eta_2^2)\sin(\eta_2 h)}\right\}, \tag{13}$$

and

$$F_3(\eta_1,\eta_2,\eta_3)=\frac{\eta_2 e^{i[\eta_1 z'-\eta_3(z-h)]}}{\eta_2(\eta_1+b_{13}\eta_3)\cos(\eta_2 h)+i(b_{23}\eta_1\eta_3+b_{12}\eta_2^2)\sin(\eta_2 h)}. \tag{14}$$

These expressions have been derived using the conventions

$$\mathrm{Im}(\eta_i)<0,\quad i=1\text{ and }3, \tag{15}$$

which are radiation conditions ensuring that the solutions for the field in the atmosphere and the sediment decay to zero in the limit of high $|z|$. Note that all three of the F_i are even functions of η_2 and mixed in both η_1 and η_3. The field expressions contain several density ratios, expressed by the factors

$$b_{ij}=\frac{\rho_i}{\rho_j},\quad i,j=1,2,\text{ or }3. \tag{16}$$

Since the density of air is three orders of magnitude less than that of seawater or sediment, the premultipliers b_{12} and b_{13}, respectively, in Eqs. (9) and (10) represent significant at

tenuating factors in the expressions for the *velocity potential* in the water column and the sediment. These very low density ratios are responsible for the inefficient transmission of acoustic *energy* across the air-sea and sea-sediment interfaces. Of course, a hydrophone responds to *pressure*, which, as stated in Eqs. (4) and (6), remains constant across the air-sea and sea-sediment boundaries. Thus, the pressure is not affected in the same way as the velocity potential by density contrast factors.

III. THE 2-D FIELD IN THE ATMOSPHERE

The expressions in Eqs. (8) and (12) account for all the types of waves that are present in the atmosphere due to the moving airborne source. The first term on the right of Eq. (12) represents the Doppler-shifted direct path arrival, while the second term includes all the remaining types of wave, each of which is also Doppler shifted. Thus, this second term represents the sea-surface-reflected arrival, the normal modes, which may be interpreted in terms of multiple arrivals from partial reflections off the boundaries of the shallow-water channel, and the lateral, or head, wave propagating at the critical angle of the air-sea interface. In effect, the term in parentheses on the right of Eq. (12) is the reflection coefficient of the sea surface for those wave components of the incident field with horizontal wavenumber p.

In many circumstances, including recent Doppler spectroscopy experiments with light aircraft,[19,20] the normal modes and the lateral wave make a negligible contribution to the total field in the atmosphere. These multipath arrivals may be removed from the solution in Eqs. (8) and (12) by allowing the channel depth, h, to become indefinitely large under the condition

$$\mathrm{Im}(\eta_2) > 0. \tag{17}$$

The expression for F_1, then reduces to

$$F_1(\eta_1,\eta_2,\eta_3) = \frac{e^{-i\eta_1|z-z'|}}{\eta_1} + \frac{e^{-i\eta_1|z+z'|}}{\eta_1}\left\{\frac{\eta_1 + b_{12}\eta_2}{\eta_1 - b_{12}\eta_2}\right\}, \tag{18}$$

where, on the right, the reflection coefficient has been derived using the relationship $b_{13}=b_{12}b_{23}$. A further simplification may be achieved by recognizing that the terms involving b_{12}, associated with the lateral wave in the atmosphere, are negligible since the air-to-seawater density ratio is $b_{12}\approx 10^{-3}$. Under this condition, Eq. (18) becomes

$$F_1(\eta_1,\eta_2,\eta_3) = \frac{e^{-i\eta_1|z-z'|}}{\eta_1} + \frac{e^{-i\eta_1|z+z'|}}{\eta_1}, \tag{19}$$

indicating that the sea surface approximates a rigid boundary to sound incident from above. On the right of Eq. (19), the first and second terms represent, respectively, the Doppler-shifted source and its image in the acoustically rigid sea surface.

When Eq. (19) is substituted into Eq. (8), the expression for the Doppler-shifted field in the atmosphere becomes

$$\phi_1(x,z,t) = \frac{Q_L e^{i\Omega t}}{8\pi^2 i}\int_{-\infty}^{\infty}\frac{e^{ip(x-x'-Vt)}}{\eta_1} \times (e^{-i\eta_1|z-z'|} + e^{-i\eta_1|z+z'|})dp. \tag{20}$$

This integral may be evaluated by completing the square under the radical η_1 and then making a straightforward substitution, which reduces it to a standard form.[23] The result for the time-dependent velocity potential is the real part of the expression

$$\phi_1(x,z,t) = \frac{Q_L}{8\pi i\sqrt{1-\beta_1^2}}\exp ik_1\left\{\frac{c_1 t - \beta_1(x-x')}{(1-\beta_1^2)}\right\} \times\left\{H_0^{(2)}\left[\frac{k_1}{1-\beta_1^2}\sqrt{(x-x'-\beta_1 c_1 t)^2 + (1-\beta_1^2)|z-z'|^2}\right] + H_0^{(2)}\left[\frac{k_1}{1-\beta_1^2}\sqrt{(x-x'-\beta_1 c_1 t)^2 + (1-\beta_1^2)|z+z'|^2}\right]\right\}, \tag{21}$$

where $H_0^{(2)}[\cdots]$ is the Hankel function of the second kind of order zero.

The two Hankel functions in Eq. (21) represent the Doppler-shifted fields from the source and its image in the (rigid) sea surface. Besides the Doppler shifts, the main feature of the field expressed by Eq. (21) is the asymmetric Lloyd's mirror structure arising from the interference between the direct and reflected arrivals. Clearly, when the source is stationary (i.e., $\beta_1=0$), Eq. (21) reduces correctly to the two cylindrically spreading terms that are characteristic of a line source.

To understand the algebraic origin of the fore-aft asymmetry in the field, Eq. (21) may be simplified, with no loss of generality, by letting the receiver range be fixed at $x=0$ and the source range be $x'=0$ when $t=0$. The source is then directly over the receiver, that is, at the closest point of approach (CPA), at the origin of time. With this coordinate system, the (real) velocity potential, from Eq. (21), may be expressed as

$$\phi_1(0,z,t) = \frac{Q_L}{8\pi\sqrt{1-\beta_1^2}}\left[\sin\left(\frac{k_1 c_1 t}{1-\beta_1^2}\right)\left\{J_0\left[\frac{k_1 R_-}{1-\beta_1^2}\right] + J_0\left[\frac{k_1 R_+}{1-\beta_1^2}\right]\right\} - \cos\left(\frac{k_1 c_1 t}{1-\beta_1^2}\right) \times \left\{Y_0\left[\frac{k_1 R_-}{1-\beta_1^2}\right] + Y_0\left[\frac{k_1 R_+}{1-\beta_1^2}\right]\right\}\right]. \quad (22)$$

where $J_0(\cdots)$, $Y_0(\cdots)$ are, respectively, Bessel functions of the first and second kind of order zero, and

$$R_\pm = \sqrt{(1-\beta_1^2)|z \pm z'|^2 + \beta_1^2 c_1^2 t^2}. \quad (23)$$

The sign of the radical in this expression is always positive. In Eq. (22), since $R_\pm$ is even in t, the terms containing the sine and cosine functions are, respectively, odd and even functions of t. This accounts for the asymmetry in the field, since t is negative as the source approaches the sensor (inbound) and positive on departure (outbound). As a check on the solution in Eq. (22), it is easily shown that when the source is far from the receiver, either inbound ($t \to -\infty$) or outbound ($t \to +\infty$), the Doppler-shifted frequencies correctly reduce to $\Omega/(1 \pm \beta_1)$, where the minus (plus) sign applies on approach (departure). To arrive at this result, the Bessel functions in Eq. (22) are replaced by their asymptotic expansions and the expression for the field then collapses to a single trigonometric term, which yields the up- and down-shifted frequencies directly.

IV. THE 2-D FIELD IN THE WATER COLUMN

The expressions for the field in the water column, Eqs. (9) and (13), take full account of the penetrable bottom boundary and the fluid-fluid nature of the air-sea interface. The function F_2 in the integrand contains six branch points in the complex p plane. From the first of Eqs. (11), it is evident that the branch points are Doppler shifted, falling in the second quadrant at $p = -k_i/(1-\beta_i)$ and the fourth quadrant at $p = +k_i/(1+\beta_i)$, $i = 1,2,3$. It is implicit here that the imaginary parts of the acoustic wavenumbers, k_i, are negative.

Poles also appear in F_2, the residues of which represent Doppler-shifted normal modes. These modes extend throughout the water column, the atmosphere, and the sediment and are represented in the latter two cases by the residues of F_1 and F_3, respectively. (In the atmosphere, of course, the modes make a negligible contribution to the field and were neglected in the above discussion.) The poles of F_2, as well as F_1 and F_3 for that matter, coincide with the zeros of the denominator in Eq. (13) and hence are solutions of the transcendental equation

$$\tan\ \eta_2 h = i\frac{\eta_2(\eta_1 + b_{13}\eta_3)}{b_{23}\eta_1\eta_3 + b_{12}\eta_2^2}. \quad (24)$$

Equation (24) is the complete moving-source, 2-D dispersion relation for the three-layer channel from which the eigenvalues of the Doppler-shifted normal modes are obtained. Since Eq. (24) is independent of the source location, it holds, regardless of whether the source is in the atmosphere, the water column, or the sediment. Equation (24) is a generalization of the Pekeris[21] dispersion relation and reduces identically to the Pekeris form when the source speed is set to zero and the atmosphere is represented as a vacuum.

The integral in Eq. (9) is evaluated by taking a D-shaped contour in the complex p plane but indented around the appropriate branch cuts. An example of such a contour, for the case of a stationary source in a Pekeris channel, is shown in Fig. 2 of Buckingham and Giddens.[22] The cut lines are chosen such that each follows the locus $\mathrm{Im}(\eta_i) = 0$, $i = 1,2,3$. This type of branch cut is attributed to Ewing, Jardetzky, and Press[24] and is commonly referred to as an EJP cut. Everywhere on the top Riemann surface produced by the EJP cuts, where the contour integrations are performed, the radiation conditions in Eq. (15) are satisfied,[25] thus ensuring that a well-behaved, convergent solution for the field in the three-layer waveguide is obtained.

By integrating around the indented D contour in the complex p plane, it is evident that the field in the water column consists of a discrete component, in the form of a finite sum of convergent, or "proper," normal modes, plus a continuous spectrum arising from contour integrations around the η_1 and η_3 branch lines:

$$\phi_2(x,z,t) = \frac{Q_L b_{12} e^{i\Omega t}}{4\pi^2 i}\left\{\sum_{m=1}^{M} \text{normal modes} + \int_{\eta_1\ cut} \cdots dp + \int_{\eta_3\ cut} \cdots dp\right\}, \quad (25)$$

where M is the total number of proper modes. There is no contribution to the field from the η_2 cut because F_2 is even in η_2, making the integrand an odd function of η_2, as a result of which the integral is identically zero. The η_1 integral yields the (usually negligible) evanescent field immediately beneath the air-sea interface; and the integral around the η_3 cut represents the continuous, subcritical field in the water column, which partially penetrates into the sediment, and also the lateral (head) wave associated with the critical angle of the bottom boundary.

Again it is convenient to set $x = 0$, and $x' = 0$ when $t = 0$, in which case the source is inbound to the receiver station when $t < 0$ and outbound when $t > 0$. From Cauchy's theorem and Jordan's lemma,[26] the contour integration leading to Eq. (25) is taken around the upper half plane inbound and the lower half plane outbound. The branch line integrals in Eq. (25) may be made more tractable by a substitution of variable. From Eq. (11), the horizontal wavenumber, p, may be expressed as

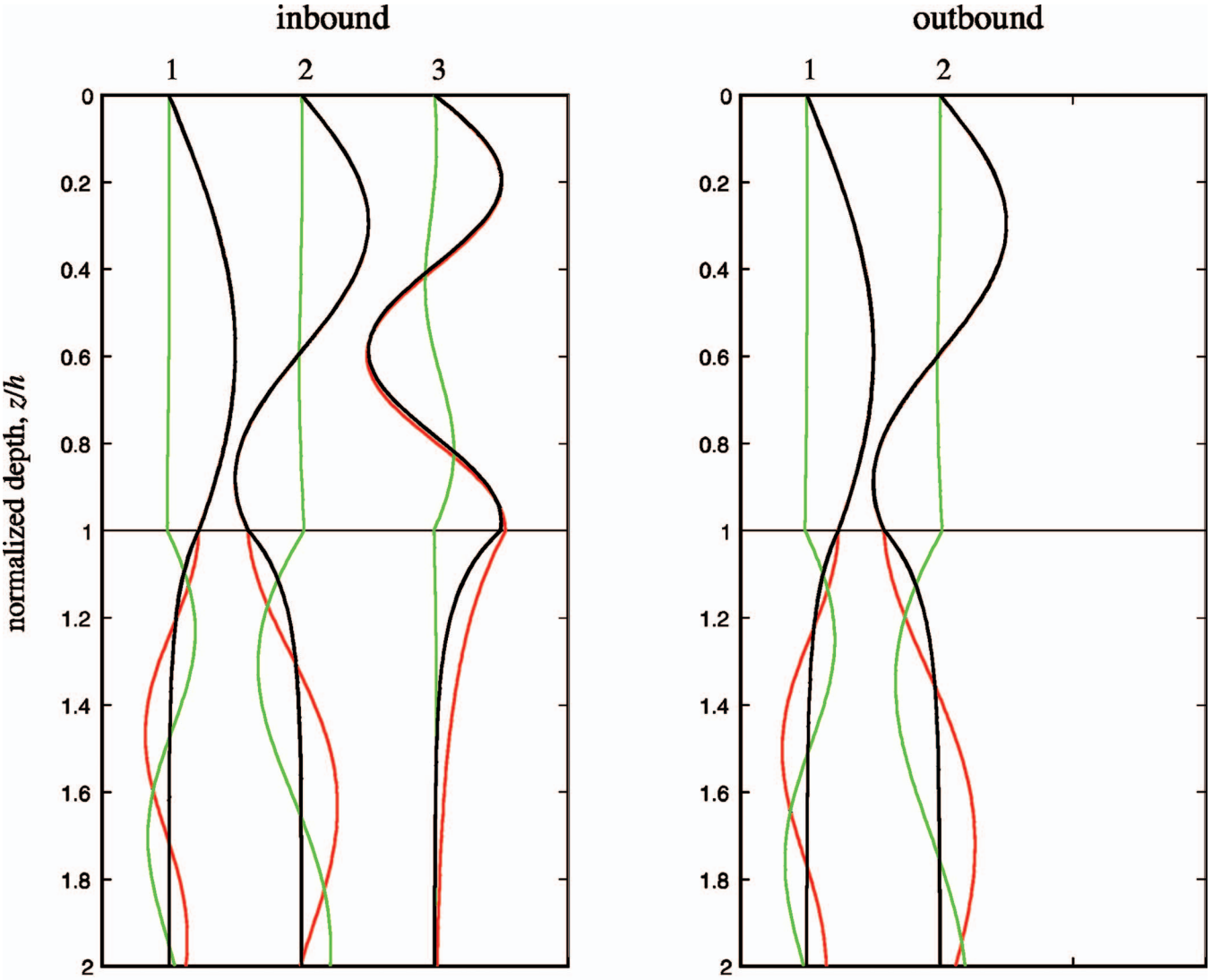


FIG. 2. Real (red) and imaginary (green) parts of the mode shapes fore and aft of the source, evaluated from the exact expressions in Eqs. (53) and (64) using the parameters of the Zhang and Tindle channel (Table I), a source frequency of 88 Hz, and source speed of 50 m/s. The corresponding approximate (black) mode shapes from Eqs. (54) and (65). In the channel, the red and black mode shapes are almost identical.

$$p = \frac{-k_i\beta_i \pm \sqrt{k_i^2 - (1-\beta_i^2)\eta_i^2}}{(1-\beta_i^2)} \tag{26a}$$

from which

$$dp = \mp \frac{\eta_i}{\sqrt{k_i^2 - (1-\beta_i^2)\eta_i^2}} d\eta_i. \tag{26b}$$

$$\phi_2(x,z,t) = \frac{Q_L b_{12} e^{i\Omega t}}{4\pi^2 i}\left\{ 2\pi i \sum_{m=1}^{M} \text{residues} + \int_{-\infty}^{\infty} e^{-ipVt} \frac{F_2(\eta_1,\eta_2,\eta_3)}{\sqrt{k_1^2 - (1-\beta_1^2)\eta_1^2}} \eta_1 \, d\eta_1 + \int_{-\infty}^{\infty} e^{-ipVt} \frac{F_2(\eta_1,\eta_2,\eta_3)}{\sqrt{k_3^2 - (1-\beta_3^2)\eta_3^2}} \eta_3 \, d\eta_3 \right\}, \quad \text{for } t<0, \tag{27}$$

Where there is a choice of sign in Eqs. (26), the upper sign applies for $t<0$, when the source is inbound and the integration contour is around the top half-plane; otherwise, for $t>0$, when the source is outbound and the integration is around the lower half-plane, the lower sign applies. For an inbound run, the expression for the field in Eq. (25) now takes the form

where the infinite limits on the integrals are the extremum values of the (real) radicals η_1 and η_3 around the respective cuts. The variable p in the integrands in Eq. (27) is given in terms of the appropriate η by Eq. (26a). An analogous expression to Eq. (27), with appropriate changes of sign, applies when $t>0$ and the source is outbound from the receiver.

A. Normal modes

1. The "moving source" dispersion relation

As in the atmosphere, some simplification to the solution for the field in the water column is achieved by exploiting the large density contrast across the air-sea boundary. Since b_{12}, $b_{13} \ll 1$, the dispersion relation for the poles in Eq. (24) may be approximated as

$$\tan \eta_2 h = i \frac{\eta_2}{b_{23} \eta_3}, \tag{28}$$

which has a similar form to the Pekeris[21] dispersion relation, except that the vertical wavenumbers in Eq. (28) include Doppler effects due to the motion of the source. Equation (28) is a transcendental equation that possesses an infinite number of solutions but of these, only a finite number, M, also satisfy the radiation condition $\mathrm{Im}(\eta_3) < 0$. Each of these M solutions represents a proper normal mode. Since the horizontal wavenumbers in Eq. (28) exhibit a fore-aft asymmetry, M, may differ according to whether the source is inbound to or outbound from the receiver.

2. Iterative solution of the dispersion relation

As it is not possible to derive an explicit closed-form solution of the dispersion relation in Eq. (28), the exact eigenvalues of the modes can only be determined numerically, using an appropriate iterative procedure. Zhang and Tindle[27] solved a similar problem, in their case for a stationary source in the water column, on the basis of a generalized expression for the "effective depth" of the channel.[28,29] An alternative approach, introduced by Buckingham and Giddens[22] for a stationary source, is extended below to the moving-source problem: the roots of the dispersion relation are obtained directly using a standard Newton-Raphson iterative procedure.

To begin, the dispersion relation in Eq. (28) is expressed in the form

$$\eta_{2m} h = m\pi + \tan^{-1}\left\{ i \frac{\eta_{2m}}{b_{23} \eta_{3m}} \right\}, \quad m = 1,2, \ldots, \tag{29}$$

where m is the mode number and the subscript m denotes the mth root of Eq. (29). By manipulating Eqs. (11), the vertical wavenumber η_{3m} is expressed in terms of η_{2m} as follows:

$$\eta_{3m} = -i \sqrt{\frac{(1+\beta_2^2)}{(1-\beta_2^2)^2}(k_2^2 - k_3^2) - \frac{(k_2^2 - k_3^2\beta_2^2)}{k_2^2(1-\beta_2^2)} \eta_{2m}^2 \mp 2\beta_2 \frac{(k_2^2 - k_3^2)}{k_2(1-\beta_2^2)^2} \sqrt{k_2^2 - (1-\beta_2^2)\eta_{2m}^2}}, \tag{30}$$

where the real part of the radical is positive. For convenience the following notation is now introduced:

$$X = \eta_{2m} h, \tag{31}$$

$$\hat{a} = \frac{(1+\beta_2^2)}{(1-\beta_2^2)^2}(k_2^2 - k_3^2)h^2, \tag{32a}$$

$$\hat{b} = \frac{(k_2^2 - k_3^2\beta_2^2)}{k_2^2(1-\beta_2^2)}, \tag{32b}$$

$$\hat{c} = 2\beta_2 h^2 \frac{(k_2^2 - k_3^2)}{(1-\beta_2^2)^2}, \tag{32c}$$

$$\hat{d} = \frac{(1-\beta_2^2)}{k_2^2 h^2} \tag{32d}$$

and

$$g(X) = b_{23} \frac{\sqrt{\hat{a} - \hat{b}X^2 \mp \hat{c}\sqrt{1 - \hat{d}X^2}}}{X}. \tag{33}$$

Returning to Eq. (29), the equation to be solved for X is

$$f(X) = X - \left(m - \tfrac{1}{2}\right)\pi - \tan^{-1}[g(X)] = 0, \quad m = 1,2, \ldots, \tag{34a}$$

the derivative of which is

$$f' = \frac{df}{dX} = 1 + \frac{1}{(1+g^2)X}\left[g + \frac{b_{23}^2}{g}\left\{ \hat{b} \mp \frac{\hat{c}\hat{d}}{2\sqrt{1 - \hat{d}X^2}} \right\} \right]. \tag{34b}$$

Where there is a choice of sign in the above expressions, the upper (lower) sign is selected when the source is inbound (outbound). If the nth approximation for the root is X_n, then

$$X_{n+1} = X_n - \frac{f(X_n)}{f'(X_n)}, \tag{34c}$$

which converges after just a few iterations to the required solution. A good starting value is $X_0 = (m - 1/2)\pi$. Once the vertical wavenumbers of the modes have been determined, the eigenvalues can be obtained from the expression

$$p_m = \frac{-k_2\beta_2 \pm \sqrt{k_2^2 - (1-\beta_2^2)\eta_{2m}^2}}{(1-\beta_2^2)}, \quad m = 1,2, \ldots, \tag{35}$$

where the minus (plus) sign in front of the radical applies when the source is inbound (outbound).

The Newton-Raphson routine in Eqs. (34) returns a finite number of solutions, $m \leq M$, representing proper modes, which satisfy both the dispersion relation and the radiation condition $\mathrm{Im}(\eta_{3m}) < 0$. Such modes are convergent in the sense of being square integrable over the channel and the semi-infinite sediment. Equations (34) also return an infinite number of solutions, $m > M$, that violate the radiation condi-

TABLE I. Shallow-water parameters for the Zhang and Tindle (Z & T) channel and the channel north of Scripps pier (SIO).

Channel	Z & T	SIO
Depth, m	54	20
c_2, m/s	1500	1500
c_3, m/s	1600	1650
b_{23}	0.8	0.56
Atten., dB/m/kHz	0.3125	0.3125

tion in the sediment. These nonphysical solutions represent divergent, improper modes, which are not part of the solution for the field. It is easy to identify the admissible solutions of the dispersion relation, simply by inspection of the imaginary part of η_{3m}.

As a check on the Newton-Raphson procedure in Eqs. (34), the iteration was performed for a stationary source ($V=0$) in the Zhang and Tindle[27] channel (Table I) using their seabed attenuation of 0.3125 dB/m/kHz. At a frequency of 100 Hz, it is found from Eqs. (34) that three proper modes are supported, the (complex) eigenvalues of which are as shown in the second column of Table II. These eigenvalues are identical to those in Table I of Zhang and Tindle.[27] Even with unrealistically high source speeds and extreme bottom attenuations, the Newton-Raphson procedure always returns solutions that satisfy the dispersion relation in Eq. (28).

With the source moving at 50 m/s over the Zhang and Tindle channel, but all else the same, the Newton-Raphson routine shows that three modes are supported both on approach and departure (see the third and fourth columns of Table II). It is interesting that this differs from the lossless case: in the absence of bottom attenuation, three modes are supported on approach, whereas, on departure, the third mode is cutoff, leaving only modes $m=1$ and $m=2$ to propagate in the channel. Clearly, in this example, when the third mode on the departure side is supported, it owes its existence to the presence of attenuation in the seabed.[22] Evidently, realistic levels of bottom loss can be sufficient to influence the proper mode count in the channel.

3. Total number of modes, lossless seabed

An exact analytical expression can be derived for the total number of proper modes, M_0, when losses in the bottom are negligible, although the convergence condition that the imaginary part of η_3 must be negative is retained. After some straightforward algebra, η_3 in Eq. (11) may be expressed in terms of η_2 as follows:

$$\eta_3 = -i\sqrt{\sin^2(\alpha_c)(k_2 - p\beta_2)^2 - \eta_2^2}, \tag{36}$$

where $\alpha_c = \cos^{-1}(c_2/c_3)$ is the critical grazing angle of the bottom and the real part of the radical is positive. The dispersion relation, Eq. (28), may then be written as

$$\tan\ \eta_2 h = -\frac{\eta_2}{b_{23}\sqrt{\sin^2(\alpha_c)(k_2 - p\beta_2)^2 - \eta_2^2}}. \tag{37}$$

The M_0 real solutions for η_2 all occur when the right hand-side of Eq. (37) is real. The mth solution, η_{2m}, can be expressed as

$$\eta_{2m}h = m\pi - \tan^{-1}\left\{\frac{\eta_{2m}}{b_{23}\sqrt{\sin^2(\alpha_c)(k_2 - p_m\beta_2)^2 - \eta_{2m}^2}}\right\}, \quad m = 1,2,\ldots,M_0, \tag{38}$$

where p_m is the eigenvalue of the mth mode.

The criterion for modal cutoff is that the upper limit on the grazing angle of an equivalent modal ray is α_c. Under this condition, the radical in Eq. (38) is zero, from which it follows, on setting $m=M_0$, that

$$\eta_{2M_0} = \left(M_0 - \tfrac{1}{2}\right)\frac{\pi}{h} = \sin(\alpha_c)(k_2 - p_{M_0}\beta_2). \tag{39}$$

To determine M_0 from this result it is necessary to identify the eigenvalue p_{M_0} of the highest proper mode. From Eq. (11),

$$\eta_{2M_0}^2 = k_2^2 - 2p_{M_0}k_2\beta_2 - p_{2M_0}^2(1-\beta_2^2), \tag{40}$$

which, when combined with the last term of Eq. (39), yields

$$p_{M_0} = \left.\begin{array}{ll} -\dfrac{k_2\cos(\alpha_c)}{1-\beta_2\cos(\alpha_c)} & \text{inbound} \\ \dfrac{k_2\cos(\alpha_c)}{1+\beta_2\cos(\alpha_c)} & \text{outbound} \end{array}\right\}. \tag{41}$$

On substituting these expressions for the highest modal eigenvalue into Eq. (39), the total number of propagating modes with the source inbound to the receiver is found to be

$$M_{0(\text{in})} = \frac{k_2 h\ \sin(\alpha_c)}{\pi[1-\beta_2\cos(\alpha_c)]} + \frac{1}{2} \tag{42a}$$

and, outbound,

$$M_{0(\text{out})} = \frac{k_2 h\ \sin(\alpha_c)}{\pi[1+\beta_2\cos(\alpha_c)]} + \frac{1}{2}, \tag{42b}$$

where the right-hand sides of these expressions are to be rounded down to the nearest integer. When the Mach number

TABLE II. Eigenvalues of the proper modes in the Zhang and Tindle channel (Table I) with an unshifted source frequency of 100 Hz. Column 2 is for a stationary source and columns 3 and 4 for a source moving at speed $V=50$ m/s. The real eigenvalues, in round brackets, are for a lossless seabed but with all else the same. Note that on the outbound run, mode 3 is absent with a lossless bottom but is supported when the attenuation is present.

Mode (m)	p_m (stationary)	p_m (inbound)	p_m (outbound)
1	0.415 853 4−0.000 054 7i (0.415 858 4)	−0.430 270 0+0.000 053 6i (−0.430 274 9)	0.402 368 4−0.000 055 8i (0.402 373 5)
2	0.406 681 2−0.000 239 4i (0.406 705 2)	−0.421 008 2+0.000 232 5i (−0.421 031 3)	0.393 285 6−0.000 246 4i(0.393 310 7)
3	0.391 841 08−0.001 322 0i (0.392 709 5)	−0.405 885 7+0.001 071 5i (−0.406 271 4)	0.378 662 5−0.001 647 8i (no mode)

is zero, it is clear that both expressions reduce identically to the correct result for a stationary source in the Pekeris waveguide.[30] From inspection, it is evident that Eqs. (42) could be obtained simply by replacing the unshifted angular frequency, Ω, in the stationary-source expression for M_0 with the shifted angular frequencies $\Omega/[1\pm/\beta_2\cos(\alpha_c)]$. The latter, of course, are just the Doppler up- and down-shifts on a ray inclined at the critical grazing angle to the horizontal.

As an example of the asymmetry expressed through Eqs. (42), consider a shallow-water channel with a lossless bottom and the remaining parameters, as shown in Table I, as discussed by Zhang and Tindle.[27] Taking a source speed of $V=50$ m/s and a propeller harmonic of unshifted frequency 100 Hz, both typical of a light aircraft, the number of modes inbound is $M_{o(\mathrm{in})}=3$ and outbound is $M_{0(\mathrm{out})}=2$ (see Table II and the discussion toward the end of Sec. IV A 4). For the shallower channel (Table I) used in the Doppler spectroscopy experiments of Buckingham *et al.*,[19,20] with $V=60$ m/s and an unshifted frequency of 135 Hz, the number of modes inbound and outbound, respectively, from Eqs. (42) is $M_{0(\mathrm{in})}=2$ and $M_{0(\mathrm{out})}=1$.

In practice, all seabeds exhibit some degree of loss, which tends to increase the number of propagating modes above the values expressed by Eqs. (42). That is to say, when k_3 is complex, the dispersion relation in Eq. (28) generally admits more solutions than would be allowed with k_3 purely real. A detailed analysis of this phenomenon, for the case of a stationary source in a Pekeris channel, has been performed by Buckingham and Giddens.[22] Table II includes a realistic example in which the introduction of bottom attenuation increases the mode count outbound but leaves it unaffected inbound.

4. Residues

To obtain the residue of the mth mode, the denominator of F_2 in Eq. (13) must be expanded to first order in $(p-p_m)$, where p_m is the mth eigenvalue given by Eq. (35). Neglecting the very small terms in b_{12} and b_{13}, the Taylor expansion of the denominator, D, yields

$$D=(p-p_m)\left.\frac{dD}{dp}\right|_{p=p_m}+\cdots=-(p-p_m)\times\left\{\frac{[k_2\beta_2+p_m(1-\beta_2^2)]}{\eta_{2m}}[\cos(\eta_{2m}h)-\eta_{2m}h\sin(\eta_{2m}h)+ib_{23}\eta_{3m}h\cos(\eta_{2m}h)]+\frac{ib_{23}\sin(\eta_{2m}h)}{\eta_{3m}}[k_3\beta_3+p_m(1-\beta_3^2)]\right\}+\cdots. \quad (43)$$

Now the wavenumber η_{3m} may be expressed in terms of η_{2m} through the dispersion relationship in Eq. (28). The denominator then becomes

$$D=-(p-p_m)\frac{L_m}{\eta_{2m}\sin(\eta_{2m}h)}+\cdots, \quad (44)$$

where

$$L_m=\left\{[k_2\beta_2+p_m(1-\beta_2^2)][\eta_{2m}h-\sin(\eta_{2m}h)\cos(\eta_{2m}h)]-\frac{b_{23}^2\sin^3(\eta_{2m}h)}{\cos(\eta_{2m}h)}[k_3\beta_3+p_m(1-\beta_3^2)]\right\}. \quad (45)$$

If R_m is the residue of the mth mode, it follows from Eqs. (9), (13), and (44) that

$$R_m=e^{-ip_mV_t}\frac{e^{i\eta_{1m}z'}}{\eta_{1m}}\frac{\eta_{2m}\sin(\eta_{2m}z)\sin(\eta_{2m}h)[\eta_{2m}\sin(\eta_{2m}h)-ib_{23}\eta_{3m}\cos(\eta_{2m}h)]}{L_m},\quad 0\leqslant z\leqslant h, \quad (46)$$

where, following earlier practice, x and x' have been set to zero and

$$\eta_{1m}=\sqrt{(k_1-p_m\beta_1)^2-p_m^2}. \quad (47)$$

Again using the dispersion relationship in Eq. (28), this time to eliminate η_{3m} in the numerator of Eq. (46), the final expression for the residues becomes

$$R_m=e^{-ip_mV_t}\frac{e^{i\eta_{1m}z'}}{\eta_{1m}}\frac{\eta_{2m}^2\sin(\eta_{2m}z)}{L_m},\quad 0\leqslant z\leqslant h. \quad (48)$$

For the special case of stationary source, L_m in Eq. (45) reduces identically to the corresponding, Doppler-free expression derived by Pekeris.[21]

5. The "effective depth" approximation

Although an exact numerical value for η_{2m} may be computed from the simple Newton-Raphson algorithm in Eqs. (34), it is also possible to derive an analytical approximation for η_{2m}, which offers some insight into the physics governing the mode shapes and the modal attenuation.

The derivation is based on the assumption that, for those modes not too close to cutoff, the vertical wavenumbers in the water column are small, in which case the terms in η_{2m}^2 may be neglected in Eq. (30). Some algebraic rearrangement then leads to the result

$$\eta_{3m}\approx -i\frac{\sqrt{k_2^2-k_3^2}}{(1\pm\beta_2)}, \quad (49)$$

which, when substituted into Eq. (29), after approximating the arctangent by its argument, leads to the required solution

$$\eta_{2m} \approx \frac{m\pi}{h\left[1 + \frac{(1 \pm \beta_2)}{b_{23} h \sqrt{k_2^2 - k_3^2}}\right]}. \tag{50}$$

In general, this solution for η_{2m} is complex due to losses in the bottom, represented by the condition $\mathrm{Im}(k_3) < 0$.

When bottom losses are negligible, however, the denominator on the right of Eq. (50) is real and so too are the vertical wavenumbers, which may expressed in the form

$$\eta_{2m} \approx \frac{m\pi}{H}, \tag{51}$$

where H is the Doppler-shifted version of the effective depth of the channel.[28,29] When the source is inbound,

$$H \equiv H_{\mathrm{in}} = h\left\{1 + \frac{1 - \beta_2}{b_{23} k_2 h \sin(\alpha_c)}\right\}, \tag{52a}$$

and outbound,

$$H \equiv H_{\mathrm{out}} = h\left\{1 + \frac{1 + \beta_2}{b_{23} k_2 h \sin(\alpha_c)}\right\}, \tag{52b}$$

where $\alpha_c = \cos^{-1}(k_3/k_2)$ is the critical grazing angle of the bottom.

Since the Mach number, β_2, is positive, it is clear from Eqs. (52) that $H_{\mathrm{out}} > H_{\mathrm{in}}$. This asymmetry in the fore and aft effective depths arises because the frequency is Doppler upshifted inbound and downshifted outbound. An inspection of Eqs. (52) reveals that these frequency shifts are given by $\Omega/(1 \pm \beta_2)$, where the absence of a cosine (grazing angle) factor multiplying the Mach number is a consequence of neglecting η_{2m}^2 in Eq. (30). In effect, this "low-order-mode" approximation amounts to treating the modal equivalent rays as though they were horizontal, with zero grazing angle.

6. Mode shape functions

The depth dependence of the modes, given by the residues in Eq. (48), may be conveniently expressed in terms of mode shape functions:

$$s_m(z) = \sin(\eta_{2m} z), \quad 0 \leqslant z \leqslant h, \tag{53}$$

where η_{2m} is the mth root of the dispersion relation [Eqs. (29) and (30)]. From the solution for η_{2m} in Eq. (51), the shape functions may be approximated in terms of the effective depth as follows:

$$s_m(z) \approx \sin\left(\frac{m\pi z}{H}\right), \quad 0 \leqslant z \leqslant h. \tag{54}$$

According to this result, the mode shapes within the channel are the same as if the seabed were a pressure-release boundary at a depth $H > h$ beneath the sea surface. Clearly, the fore-aft asymmetry in H, as expressed by Eqs. (52), gives rise to different shape functions ahead of and behind the moving source. Although the effect is usually small, as illustrated in Fig. 2 for the case of the Zhang and Tindle[27] channel, it is evident from Eq. (54) that the turning points in the modes are shallower with the source inbound than outbound. Of course, the main fore-aft asymmetry in Fig. 2 is the presence of three modes inbound and only two outbound.

7. Modal attenuation

According to Eq. (48) for the residues, each mode undergoes attenuation as it propagates along the channel. Since Vt is, in effect, the source-to-receiver range, it is evident that this modal decay, arising from losses in the seabed, is associated with the first exponential function in Eq. (48) and that the attenuation coefficient of the mth mode is

$$\alpha_m = \mathrm{Im}(p_m), \tag{55}$$

where the complex eigenvalue, p_m, is given by Eq. (35).

An approximate expression for the modal attenuation coefficient, α_m, may be obtained from Eq. (50) by making the acoustic wavenumber, k_3, complex, thus allowing for losses in the bottom. It is convenient to write

$$k_3 = k_3'(1 - i\gamma), \tag{56}$$

where $k_3' = \Omega/c_3$ is real and γ is the loss tangent of plane waves propagating through the bottom. [*N.B.*: The imaginary term in Eq. (56), $k_3'\gamma$, is the plane-wave attenuation coefficient of the bottom.] For most marine sediments, γ is a small number close to 0.01. On substituting Eq. (56) into Eq. (50) and expanding the result to first order in γ, the following approximation is obtained:

$$\eta_{2m} \approx \frac{m\pi}{H}\left[1 + \frac{i\gamma(1 \pm \beta_2)\cot^2(\alpha_c)}{b_{23} k_2 H \sin(\alpha_c)}\right], \tag{57}$$

where the critical grazing angle is now defined as

$$\alpha_c = \cos^{-1}\left(\frac{k_3'}{k_2}\right) \tag{58}$$

and H is the effective depth, either H_{in} or H_{out} in Eqs. (52). The choice of sign in Eq. (57) is such that the minus (plus) applies inbound (outbound).

Turning now to the expression for p_m in Eq. (35), on substituting for η_{2m} from Eq. (57) and expanding to first order in γ and to second order in the mode number m, the eigenvalues are approximated as

$$p_m \approx \frac{-k_2\beta_2 \pm \sqrt{k_2^2 - (1 - \beta_2^2)(m^2\pi^2/H^2)}}{(1 - \beta_2^2)} \mp \frac{i\gamma(1 \pm \beta_2)m^2\pi^2 \cot^2(\alpha_c)}{b_{23} k_2^2 H^3 \sin(\alpha_c)}. \tag{59}$$

According to Eq. (55), the imaginary part of this expression is the attenuation coefficient of the mth mode:

$$\alpha_m \approx \frac{\gamma(1 \pm \beta_2)m^2\pi^2 \cot^2(\alpha_c)}{b_{23} k_2^2 H^3 \sin(\alpha_c)}, \tag{60}$$

where on approach (departure) the minus (plus) sign applies and the effective depth H is given by $H_{\mathrm{in}}(H_{\mathrm{out}})$ in Eq. (52). To the level of approximation in Eq. (60), the modal attenuation scales as the square of the mode number, giving rise to the phenomenon of mode stripping, and inversely as the cube of the effective depth. When β_2 is set to zero, and after allowing for differences in notation, it can be seen that Eq. (60) reduces identically to Buckingham's[29] expression for the modal attenuation with a stationary source. In passing, it should be mentioned that the real

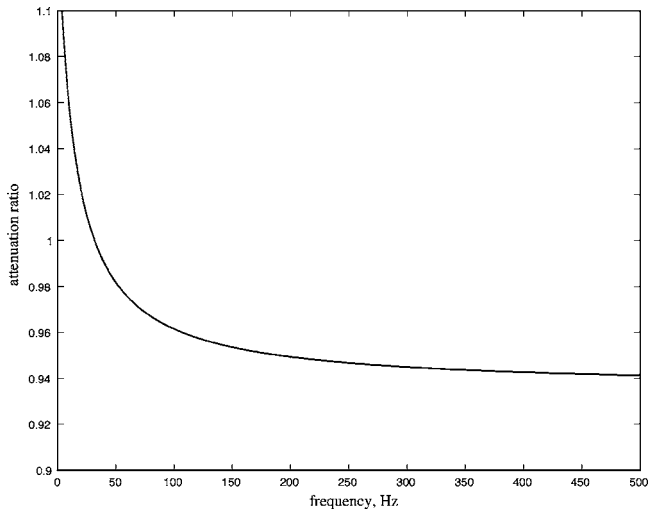


FIG. 3. Modal attenuation ratio as a function of source (unshifted) frequency, evaluated from Eq. (61) for the Zhang and Tindle channel (Table I) with source speed $V=50$ m/s.

parts of the eigenvalues in Eq. (59) switch sign as the source passes through the zenith (CPA), reflecting the fact that the acoustic arrivals at the receiver propagate in the positive x direction on approach (inbound), but in the negative x direction on departure (outbound).

Interestingly, the effects of source motion on the modal attenuation do not have a simple interpretation in terms of near-horizontal modal equivalent rays. The Mach number, β_2, appears in Eq. (60) through the terms $(1\pm\beta_2)$ in the numerator and H^3 in the denominator. These terms introduce a fore-aft asymmetry into the modal attenuation, the magnitude of which may be appreciated by forming the inbound to outbound attenuation ratio:

$$A = \frac{\alpha_{m(\mathrm{in})}}{\alpha_{m(\mathrm{out})}} \approx \frac{(1-\beta_2)}{(1+\beta_2)}\left(\frac{H_{\mathrm{out}}}{H_{\mathrm{in}}}\right)^3, \tag{61}$$

This ratio is independent of mode number but depends on frequency through the cube of the effective depths. Figure 3 shows A versus frequency for the parameters of the Zhang and Tindle[27] channel, as listed in Table I. In the limit of high frequency, both effective depths in Eq. (61) approach asymptotically the actual depth, h, and A takes the value $(1-\beta_2)/(1+\beta_2)$, indicating that the modal attenuation is *lower* on approach than on departure. At lower frequencies, however, the effective depths in Eq. (61) are no longer equal, leading to an increase in A with decreasing frequency. For the case illustrated in Fig. 3, the attenuation ratio, A, passes through unity at a frequency of about 30 Hz, below which the modal attenuation is *higher* on approach than on departure. In general, equality between the inbound and outbound attenuation occurs at angular frequency,

$$\Omega_{A=1} \approx \frac{2c_2}{b_{23}h\sin(\alpha_c)}\left(1-\tfrac{4}{9}\beta_2^2\right), \tag{62}$$

where the approximation is valid to second order in the Mach number β_2. Naturally, for any given mode, the condition in Eq. (62) would not be observed if the mode cutoff frequency were above $\Omega_{A=1}/2\pi$.

B. Branch line integrals

The branch line integrals in Eq. (27) represent field components that tend to decay relatively quickly with distance from the source. Thus, these integrals are significant only when the moving source is close to the zenith. Although the branch line integrals cannot be expressed explicitly, they may be approximated in several different ways using asymptotic techniques. For the case of a stationary source in the water column, Pekeris[21] presented without proof a particular approximation for the η_3 integral representing the lateral wave; and an approximation for the continuous field, in the form of an infinite sum of "virtual modes," has been developed from the η_3 integral by Tindle and colleagues.[31,32]

In the presence of source motion, the algebra involved in the asymptotic analyses of the branch line integrals tends to be excessively cumbersome. Since such analyses are not particularly enlightening, at least as far as moving-source effects are concerned, they are not pursued here.

V. THE FIELD IN THE SEDIMENT

As in the water column, the field in the sediment consists of a finite sum of normal modes plus two branch line integrals. Again, the branch line integrals are held in abeyance, while the normal-mode field is examined only briefly, since the modes in the sediment are simply one-to-one extensions of the modes in the water column, satisfying the same dispersion relationship [Eq. (28)], and thus possessing identical eigenvalues [Eq. (35)] and attenuation [Eq. (60)]. The residues and the mode shapes, of course, differ from those in the water column.

Following an argument similar to that for the residues in the water column, the residues in the sediment are found, from Eqs. (10), (14), and (44), to be

$$R_m = e^{-p_m V t} e^{i[\eta_{1m}z' - \eta_{3m}(z-h)]}\frac{\eta_{2m}^2}{\eta_{1m}}\frac{\sin(\eta_{2m}h)}{L_m}, \quad h < z < \infty. \tag{63}$$

Therefore, the mode shape functions in the sediment are

$$s_m(z) = \sin(\eta_{2m}h)e^{-i\eta_{3m}(z-h)}, \quad h < z < \infty. \tag{64}$$

From the approximations for the vertical wavenumbers in Eqs. (49) and (51), these shape functions may be expressed as

$$s_m(z) \approx \sin\left(\frac{m\pi h}{H}\right)e^{-k_2(z-h)\sin(\alpha_c)/(1\pm\beta_2)}. \tag{65}$$

When the source is inbound (outbound), the sign preceding the Mach number is negative (positive) and the effective depth, H, is given by $H_{\mathrm{in}}(H_{\mathrm{out}})$ in Eq. (52). It is evident from Eq. (65) that, due to the source motion, the modes decay with depth in the sediment more rapidly on approach than on departure. This, and the other fore-aft asymmetries in the mode shapes, are illustrated in Fig. 2.

VI. SOLUTION FOR THE 3-D DOPPLER-SHIFTED FIELD IN A THREE-LAYER WAVEGUIDE

In practice, an aircraft is probably better represented as a point source rather than a line source. The Doppler-shifted field then becomes three-dimensional, depending on (x,y,z), where, as shown in Fig. 1, the additional coordinate y is the direction normal to the aircraft's track. Fore and aft of the point source, the Doppler-shifted field is similar to that of the line source, except for differences in the geometrical spreading. Assuming that the aircraft's track is in the vertical plane $y=0$, then symmetry dictates that the field be an even function of y. Incidentally, a related problem, that of a stationary point source in a moving medium, may be found in the book by Brekhovskikh and Godin.[33]

The wave equations to be solved for the Doppler-shifted 3-D field due to an unaccelerated, horizontally moving airborne point source are

$$\frac{\partial^2\phi_1}{\partial x^2}+\frac{\partial^2\phi_1}{\partial y^2}+\frac{\partial^2\phi_1}{\partial z^2}-\frac{1}{c_1^2}\frac{\partial^2\phi_1}{\partial t^2} = -Q_p\delta(x-x'-Vt)\delta(y)\delta(z-z')e^{i\Omega t}, \quad z<0, \tag{66}$$

$$\frac{\partial^2\phi_2}{\partial x^2}+\frac{\partial^2\phi_2}{\partial y^2}+\frac{\partial^2\phi_2}{\partial z^2}-\frac{1}{c_2^2}\frac{\partial^2\phi_2}{\partial t^2}=0, \quad 0<z<h, \tag{67}$$

$$\frac{\partial^2\phi_3}{\partial x^2}+\frac{\partial^2\phi_3}{\partial y^2}+\frac{\partial^2\phi_3}{\partial z^2}-\frac{1}{c_3^2}\frac{\partial^2\phi_3}{\partial t^2}=0, \quad z>h, \tag{68}$$

where $\phi_i=\phi_i(x,y,z,t)$ is the velocity potential in layer $i=1,2,3$, Q_P is the source strength [with dimensions of (length)^3/time]. Following an analysis almost identical to that for the 2-D field, but with an additional Fourier transform over y, these equations may be solved for the field in each of the three layers. The results are as follows:

$$\phi_1(x,y,z,t)=\frac{Q_P e^{i\Omega t}}{16\pi^3 i}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{ip(x-x'-Vt)}e^{isy}\times F_1(\eta_1,\eta_2,\eta_3)dp\,ds, \quad z<0, \tag{69}$$

$$\phi_2(x,y,z,t)=\frac{Q_P b_{12} e^{i\Omega t}}{8\pi^3 i}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{ip(x-x'-Vt)}e^{isy}\times F_2(\eta_1,\eta_2,\eta_3)dp\,ds, \quad 0\leqslant z\leqslant h, \tag{70}$$

$$\phi_3(x,y,z,t)=\frac{Q_P b_{13} e^{i\Omega t}}{8\pi^3 i}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{ip(x-x'-Vt)}e^{isy}\times F_3(\eta_1,\eta_2,\eta_3)dp\,ds, \quad z>h, \tag{71}$$

where the Fourier transform variable s is the horizontal wavenumber in the y direction. The functions F_i in these double integrals are exactly the same as in the 2-D case, that is, they are given identically by Eqs. (12)–(14). The vertical wavenumbers, η_i, however, are now

$$\eta_i=\sqrt{(k_i-p\beta_i)^2-p^2-s^2}, \quad i=1,2,3, \tag{72}$$

where it is important to note the presence of both horizontal wavenumbers, p and s, instead of just p in the 2-D case. As before, the constraints

$$\mathrm{Im}(\eta_i)<0, \quad i=1,3, \tag{73}$$

must hold in order to ensure that the fields in the atmosphere and sediment satisfy the radiation conditions by converging to zero as $|z|$ goes to infinity.

From inspection of the functions F_i, it is evident that all three possess the same denominator, which has zeros when the 3-D dispersion relationship is satisfied:

$$\tan\,\eta_2 h=i\frac{\eta_2(\eta_1+b_{13}\eta_3)}{b_{23}\eta_1\eta_3+b_{12}\eta_2^2}. \tag{74}$$

Clearly, this has exactly the same functional form as the dispersion relationship in Eq. (24) for the 2-D field but, as already mentioned, the η_i, given in Eq. (72), are now functions of two Fourier transform variables, p and s. On neglecting the terms involving the very small density ratios b_{12} and b_{13}, the exact 3-D dispersion relationship in Eq. (74) reduces to

$$\tan\,\eta_2 h=i\frac{\eta_2}{b_{23}\eta_3}, \tag{75}$$

just as in the 2-D case.

Although the dispersion relations in two and three dimensions have the same functional form, the solution of Eq. (75) is not so straightforward as for the 2-D field. The difficulty may be appreciated by returning to Eq. (30), which, for the 2-D case, expresses η_3 uniquely in terms of η_2 along with a number of physical constants (acoustic wavenumbers and Mach numbers). This simple relationship holds for the 2-D field, even in the presence of source motion. On moving to three dimensions, the expression for η_3 corresponding to Eq. (30) involves η_2 and the same physical constants, but now, because of the reduced symmetry due to the source motion, it also depends on one or other of the Fourier variables (p,s). It follows that the solution of Eq. (75) for η_2 is a function of one of these integration variables, say s. Hence, in the 3-D case, in the presence of source motion, the poles in the complex p plane, along with the modal eigenvalues, the mode shape functions, the effective depth, and the modal attenuation, are all functions of s. With a stationary source, of course, symmetry is recovered, the s dependence disappears from η_2 and elsewhere, and the solution of Eq. (75) proceeds exactly as in the 2-D case.

Clearly, even with a moving source, a Newton-Raphson solution of Eq. (75) could be developed, allowing η_2 to be computed for each value of s. The integrals over p in Eqs. (69)–(71) could then be expressed in terms of normal-mode sums plus additional terms, all of which would be dependent on s; and the integrals over s could then be evaluated numerically. However, such an exercise is of doubtful utility. Instead, it is easier to compute the double integrals in Eqs. (69)–(71) directly. The 3-D fields that are returned exhibit asymmetric features that are qualitatively similar to those discussed earlier, in connection with the 2-D situation.

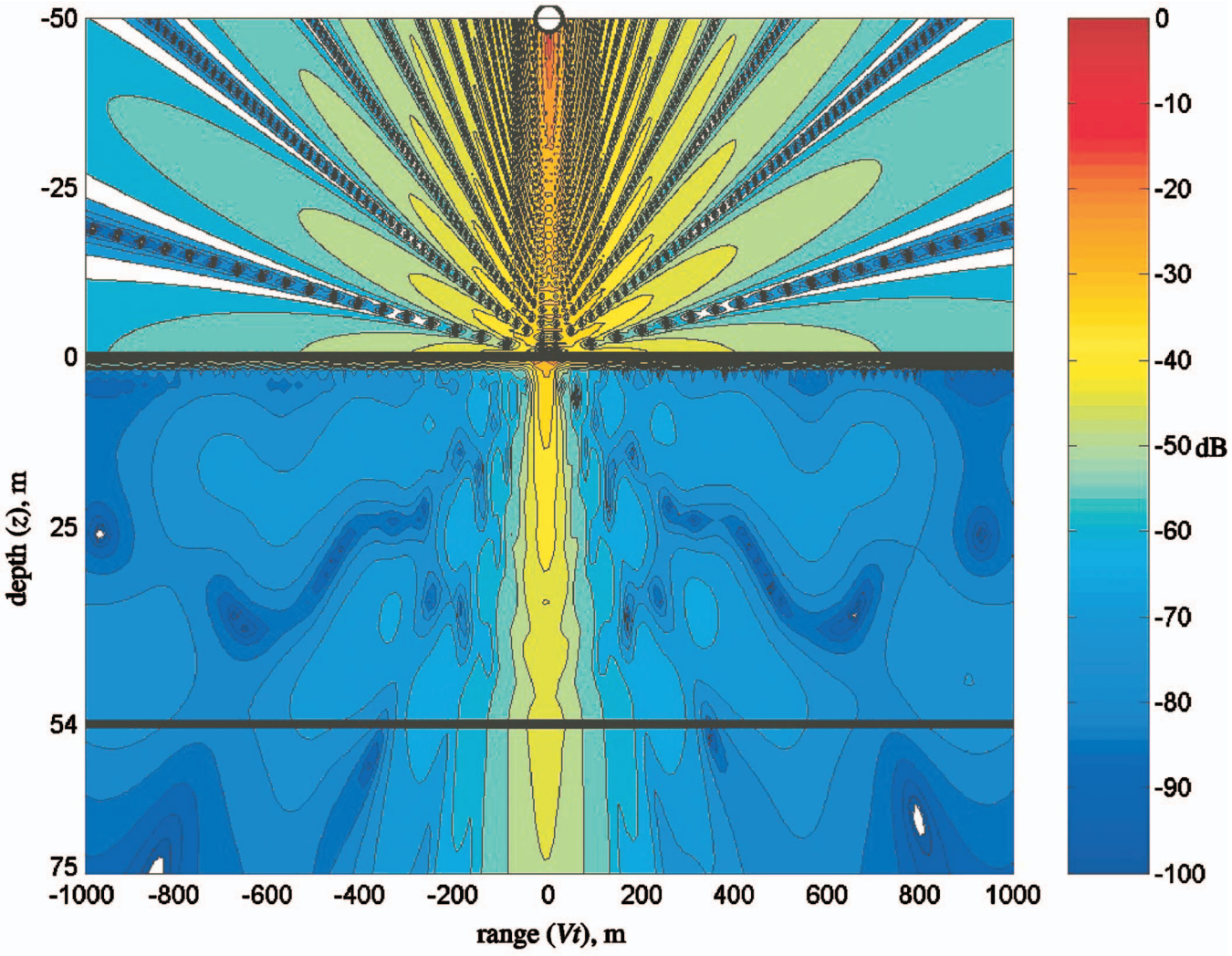


FIG. 4. Pressure field, computed from Eqs. (69) and (71), in the vertical plane containing the source track in the Zhang and Tindle channel (Table I) with source (unshifted) frequency $f=88$ Hz, speed $V=50$ m/s and altitude $|z'|=50$ m. The color bar is dB relative to the free-space source level at 1 m (e.g., if the free-space source level were 120 dB *re* 1 μPa^2 at 1 m and the pressure at a point in the channel is −50 dB, then the actual signal level at the point is 70 dB *re* 1 μPa^2).

Figure 4 shows the pressure field in the vertical plane containing the source track (i.e., the $y=0$ plane) in the Zhang and Tindle[27] channel (Table I), as computed from the time derivatives of the velocity potentials in Eqs. (69)–(71). The horizontal axis represents horizontal range, Vt, between the source and receiver and the source speed is $V=50$ m/s. Negative (positive) ranges correspond to approach (departure). In effect, Fig. 4 is a transmission loss (TL) plot showing the field on the vertical line $(x=0,\ y=0,z)$ extending throughout the three layers of the waveguide. Figure 5 shows a horizontal cut through the same pressure field at a depth of 53 m in the channel, which in this case is 1 m above the seabed. Both figures illustrate the fore-aft asymmetry in the field that is introduced by the source motion.

Although the source is monotonic, the frequency of the pressure fields in Figs. 4 and 5 is not spatially uniform but varies with the relative positions of the source and receiver, an effect due to the Doppler shifts that are introduced through the motion of the source. For instance, with the receiver ahead of (behind) the source the frequency is greater (less) than the unshifted source frequency. This multifrequency characteristic obviously distinguishes moving-source fields like those in Figs. 4 and 5 from conventional, single-frequency TL plots associated with a stationary source.

The time series of the field in Fig. 4 at depth $z=53$ m in the channel is shown in Fig. 6(a). This is a nonstationary, multifrequency waveform extending 20 s either side of CPA and with a spectral structure that is determined not only by the source motion but also by the propagation conditions in the three-layer waveguide. Figure 6(b) shows the power spectrum of the entire nonstationary time series in Fig. 6(a), as obtained from a single FFT of 40 s duration, corresponding to a frequency cell width of 0.025 Hz.

It can be seen that the "high resolution" spectrum in Fig. 6(b) exhibits Doppler spreading, which extends approximately 3 Hz either side of the unshifted source frequency, $f=88$ Hz. The spread spectrum contains a number of sharply defined, Doppler-shifted peaks. Generally, a pair of peaks is associated with each normal mode in the channel, one of the pair being upshifted (approach) and the other downshifted (departure). As discussed in more detail later, the lower the mode number, the nearer the equivalent modal rays are to the

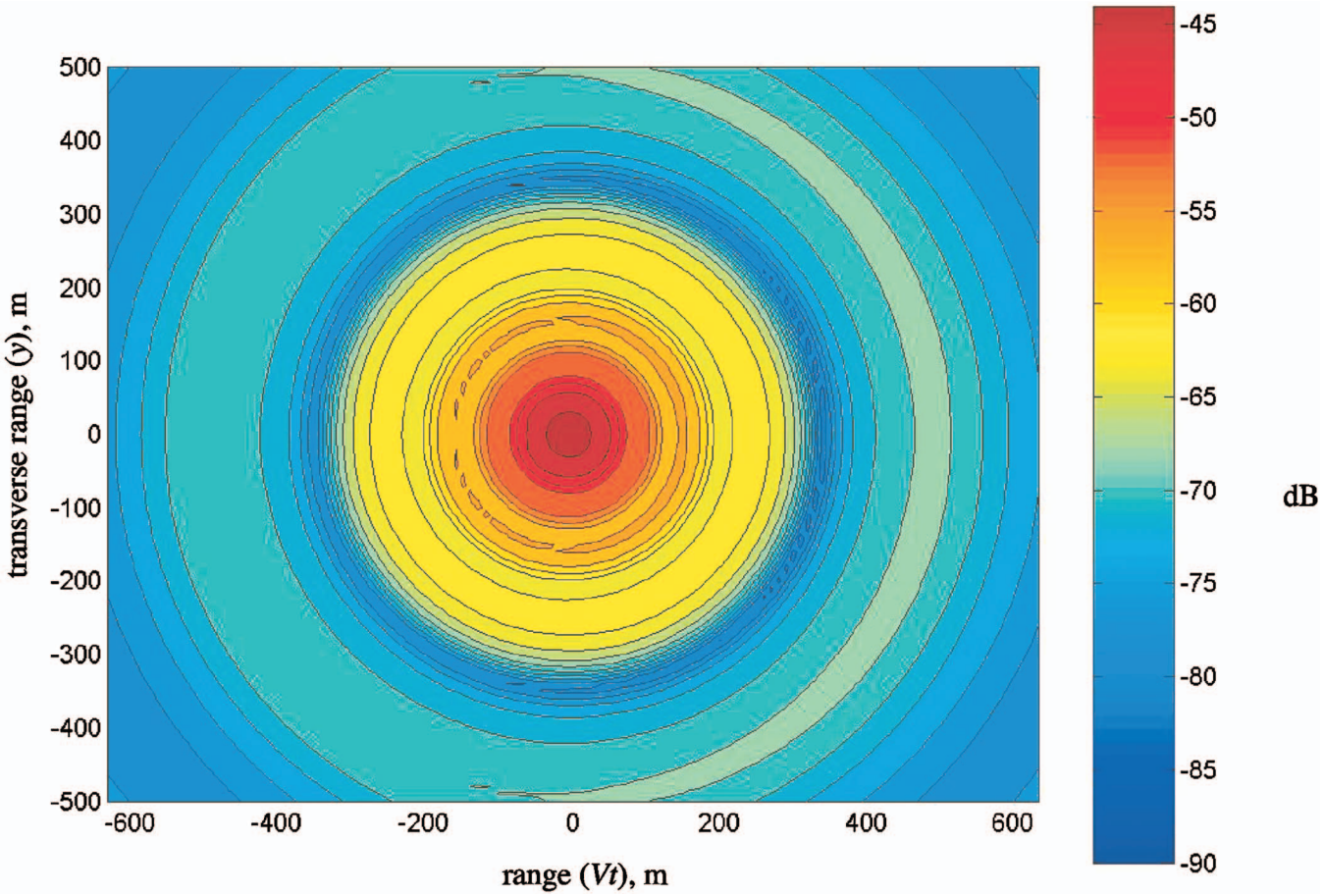


FIG. 5. Pressure field, computed from Eqs. (69) and (71), in the horizontal plane at depth $z=53$ m in the Zhang and Tindle channel (Table I) with source (unshifted) frequency $f=88$ Hz, speed $V=50$ m/s, and altitude $|z'|=50$ m. The reference of the dB scale on the color bar is as defined in Fig. 4.

horizontal and the greater are the shifts in the frequencies of these spectral peaks: mode 1 exhibits the greatest (up- and down-) shift, mode 2 a little less, and so on. Of course, if the source were stationary, there would be no Doppler shifting and the spectrum in Fig. 6(b) would collapse onto a single line at the unshifted frequency of 88 Hz.

In Fig. 6(b), the two small peaks on the flanks of the spectrum, at 91.007 and 85.189 Hz, respectively, are identified with mode 1; mode 2 corresponds to the two peaks at 90.921 and 85.273 Hz; and mode 3 to the peak at 90.773 Hz. On the downshifted side, mode 3 is actually cutoff, even though a sharp, mode-like peak is evident at 85.419 Hz that at first sight would seem to be the downshifted third mode. However, this peak and the remaining spectral peaks in Fig. 6(b) are associated with Doppler-shifted *virtual modes*, which, as discussed by Tindle *et al.*[30,31] for the case of a stationary source in the water column, are discrete approximations to the continuous field in the channel. [The wavenumber integral in Eq. (70) is complete and exact and can be expressed as a finite sum of normal modes plus a branch line integral, the latter representing both the continuous field and the lateral wave. An asymptotic analysis of the branch line integral allows it to be approximated as a sum of discrete, mode-like terms, known as virtual modes. As a mode becomes cut off, say, by reducing the frequency, it moves out of the (true) mode sum, where it was the highest mode, to become the lowest-order virtual mode in the branch line integral.]

VII. RAY INTERPRETATION OF THE MODAL PEAKS

A moving, monotonic acoustic source in a homogeneous medium emits rays that exhibit a Doppler frequency shift that depends on the launch angle. This is illustrated schematically in Fig. 7, which shows a horizontally traveling source of unshifted frequency $f_0=\Omega/2\pi$. Rays propagating ahead of the source are upshifted in frequency, with the maximum frequency occurring along the source track in the direction of source motion. Similarly, rays behind the source are downshifted in frequency with the minimum frequency occurring along the source track in the direction opposite to the source motion. Normal to the source track, the ray propagates at the unshifted frequency, f_0. In general, at an intermediate launch angle, α, the frequency of a ray is given by the familiar Doppler expression

$$f=\frac{f_0}{1-\frac{V}{c_1}\cos\alpha}. \tag{76}$$

It may be inferred from Eq. (76) that the direction of a ray identifies its frequency, and *vice versa*.

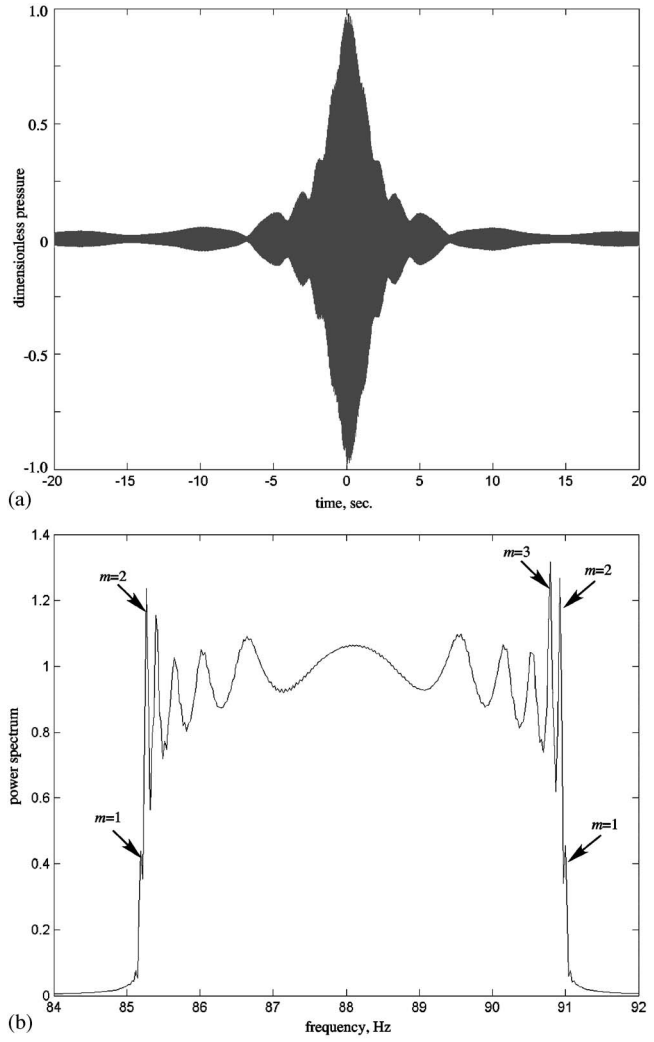


FIG. 6. (a) Time series, computed from Eq. (70), at depth $z=53$ m in the Zhang and Tindle channel (Table I) directly beneath the source track with source (unshifted) frequency $f=88$ Hz, speed $V=50$ m/s, and altitude $|z'|=50$ m. (b) Corresponding high-resolution power spectrum showing the up- and downshifted modal peaks, identified by mode number, m. The frequencies of the three upshifted peaks are 91.007, 90.921, and 90.773 Hz and the two downshifted peaks are at 85.189 and 85.273 Hz.

For rays incident from above, the air-sea interface has a critical angle of approximately 13°. Shallower rays are totally reflected but steeper rays penetrate the boundary, where they are refracted according to Snell's law, as illustrated in Fig. 8. The frequency along a given ray path is, of course, constant. If a refracted ray in the water column has the same grazing angle as an equivalent modal ray, that mode will be

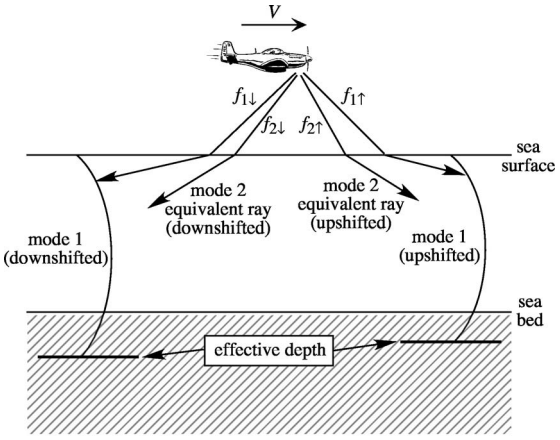


FIG. 8. Ray schematic of sound from a monotonic (frequency f_0), moving, airborne source coupling into normal modes in the channel. The arrows in the water column represent equivalent modal rays, each of which propagates at a specific, discrete grazing angle that is governed by the mode number, the Doppler-shifted frequency, the depth of the channel, and the sound speed in the bottom. To avoid clutter, the second mode shape has been omitted from the diagram.

excited and will propagate through the channel. Since the rays in the water column from a moving airborne source have grazing angles that extend continuously from 0° to 180°, it follows that all possible modes supported by the channel are excited by the source.

Each mode has a unique frequency, characterized by the launch angle of the ray that excited it. Those modes propagating ahead of the source (from left to right in Fig. 8) are upshifted in frequency, with the lowest-order mode experiencing the greatest Doppler upshift because its equivalent ray has the shallowest grazing angle. Since their equivalent rays are steeper, higher-order modes have successively lower frequencies, but all are higher than the unshifted frequency, f_0. A similar description applies to modes propagating behind the source (from right to left in Fig. 8). They will be downshifted in frequency, with the lowest-order mode experiencing the greatest Doppler downshift, again because its equivalent ray has the shallowest grazing angle. Higher-order modes have successively higher frequencies but all are lower than the unshifted frequency, f_0. To summarize, the mode frequencies satisfy the inequalities

$$f_{1\downarrow} < f_{2\downarrow} < \cdots < f_0 < \cdots < f_{2\uparrow} < f_{1\uparrow}, \tag{77}$$

where the numerical subscripts denote the mode number and the arrows (up or down) represent the direction of the Doppler shift.

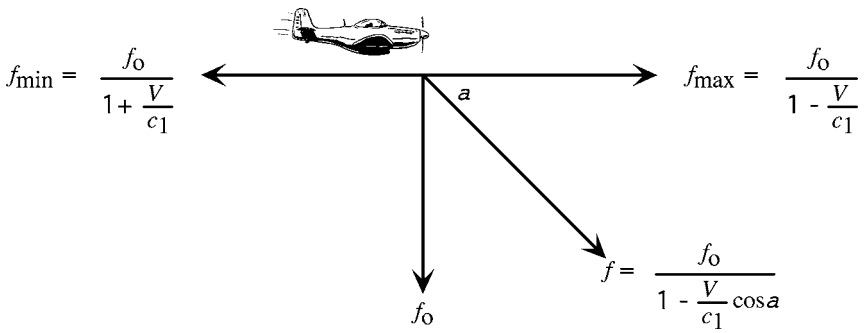


FIG. 7. The Doppler-shifted frequency of an acoustic ray emitted from a moving source decreases as the launch angle, α, increases.

This picture of modal equivalent rays in the channel fore and aft of the moving source is consistent with the presence of the very sharp modal peaks on either flank of the spread spectrum in Fig. 6(b). The mode-1 upshifted and downshifted spectral peaks are associated with the shallowest equivalent modal rays propagating, respectively, ahead of and behind the source. It is clear from Eq. (77) that these rays experience the greatest Doppler shift. Successive model peaks in the spectrum exhibit progressively smaller Doppler shifts, consistent with the steeper grazing angles of the equivalent modal rays.

VIII. CONCLUDING REMARKS

An analytical theory of sound from a horizontally moving, unaccelerated airborne source (an aircraft) in a three-layer (atmosphere-ocean-sediment) Pekeris waveguide is developed in this paper. Each layer is taken to be homogeneous and isotropic, implying a uniform sound speed throughout, and all three layers are assumed to be fluid, incapable of supporting shear. Full solutions for two source geometries are presented.

First, a line source normal to the direction of travel is considered, which gives rise to a 2-D field, varying in horizontal range and depth. This exact 2-D solution takes the form of a *single* wavenumber integral for each layer. From this solution, the dispersion relation [Eq. (24)] for the three-layer, moving-source problem is derived. This dispersion relation is complete and exact, accounting fully for all source-motion effects as well as the penetrable nature of the sea bed and the fluid-fluid boundary condition at the sea surface. It is clear from the dispersion relation that source motion introduces intricate Doppler shifts into the field, which, as a result, exhibits significant fore-aft asymmetries. Various asymmetrical properties of the field are investigated on the basis of the 2-D dispersion relation, including the effective depth of the channel, the shapes of the normal modes, the mode count, and the modal attenuation. The discussion yields considerable insight into the essential physics underlying the Doppler-shifted field in the channel and the sediment. In the atmosphere, of course, the modal component of the field, although present, is negligible compared with the direct and surface reflected arrivals, which interfere with each other to form a Doppler-shifted version of a Lloyd's-mirror field exhibiting significant fore-aft asymmetry.

The second solution presented in the paper is for a moving point source, which produces a 3-D field and is perhaps more representative of an aircraft than a line source. In the 3-D case, the solution for the field in each layer takes the form of a *double* wavenumber integral, from which the exact 3-D dispersion relation [Eq. (74)], incorporating full Doppler effects, is derived. Numerical examples of the field in the three layers, as computed from the 3-D double integrals, illustrate graphically the asymmetries that appear fore and aft of the source.

In the water column and basement particularly, the complexities of the Doppler-shifts ahead of and behind the moving source depend on a number of factors, including the bottom boundary conditions, which themselves depend on the geoacoustic parameters of the sediment. The fore-aft asymmetry introduced into the field by the moving, airborne source is the basis of the Doppler-spectroscopy inversion technique that is currently being developed for recovering the geoacoustic properties of the seabed.[34]

ACKNOWLEDGMENTS

We are indebted to an anonymous reviewer for bringing to our attention the discussion in Ref. 33 of a stationary point source in a moving medium. This work was supported by Dr. Ellen Livingston, Ocean Acoustics Code 321OA, the Office of Naval Research, under Grant No. N00014-04-1-0063.

[1] M. V. Lowson, "The sound field for singularities in motion," Proc. R. Soc. London, Ser. A **286**, 559–572 (1965).
[2] F. G. Leppington and H. Levine, "The sound field of a pulsating sphere in unsteady rectilinear motion," Proc. R. Soc. London, Ser. A **412**, 199–221 (1987).
[3] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[4] A. D. Pierce, *Acoustics: An Introduction to its Physical Principals and Applications* (McGraw-Hill, New York, 1981).
[5] J. C. Doppler, "Remarks on my theory of the colored light from double stars, with regard to the objections raised by Dr. Ballot of Utrecht," Ann. Phys. Chem. **68**, 1–35 (1846).
[6] R. P. Flanagan, N. L. Weinberg, and J. G. Clark, "Coherent analysis of ray propagation with moving source and fixed receiver," J. Acoust. Soc. Am. **56**, 1673–1680 (1974).
[7] J. G. Clark, R. P. Flanagan, and N. L. Weinberg, "Multipath acoustic propagation with a moving source in a bounded deep ocean channel," J. Acoust. Soc. Am. **60**, 1274–1284 (1976).
[8] G. M. Jacyna and M. J. Jacobson, "Analysis of source-motion effects on sound transmission in the deep ocean," J. Acoust. Soc. Am. **61**, 1153–1162 (1977).
[9] M. Dzieciuch and W. Munk, "Differential Doppler as a diagnostic," J. Acoust. Soc. Am. **96**, 2414–2424 (1994).
[10] G. M. Jacyna, M. J. Jacobson, and J. G. Clark, "General treatment of source motion on the total acoustic field with application to an isospeed channel," J. Acoust. Soc. Am. **60**, 815–824 (1976).
[11] J. A. Neubert, "The effect of Doppler on long-range sound propagation," J. Acoust. Soc. Am. **62**, 1404–1411 (1977).
[12] K. E. Hawker, "A normal mode theory of acoustic Doppler effects in the oceanic waveguide," J. Acoust. Soc. Am. **65**, 675–681 (1979).
[13] H. Schmidt and W. A. Kuperman, "Spectral and modal representations of the Doppler-shifted field in ocean waveguides," J. Acoust. Soc. Am. **96**, 386–395 (1994).
[14] R. J. Urick, "Noise signature of an aircraft in level flight over a hydrophone in the sea," J. Acoust. Soc. Am. **52**, 993–999 (1972).
[15] H. Medwin, R. A. Helbig, and J. D. Hagy, Jr., "Spectral characteristics of sound transmission through the rough sea surface," J. Acoust. Soc. Am. **54**, 99–109 (1973).
[16] W. J. Richardson, C. R. Greene, Jr., C. I. Malme, and D. H. Thomson, *Marine Mammals and Noise* (Academic, New York, 1995).
[17] B. G. Ferguson, "A ground-based narrow-band passive acoustic technique for estimating the altitude and speed of a propeller-driven aircraft," J. Acoust. Soc. Am. **92**, 1403–1407 (1992).
[18] B. G. Ferguson, "Doppler effect for sound emitted by a moving airborne source and received by acoustic sensors located above and below the sea surface," J. Acoust. Soc. Am. **94**, 3244–3247 (1993).
[19] M. J. Buckingham, E. M. Giddens, J. B. Pompa, F. Simonet, and T. R. Hahn, "Sound from a light aircraft for underwater acoustics experiments?," Acta. Acust. Acust. **88**, 752–755 (2002).
[20] M. J. Buckingham, E. M. Giddens, F. Simonet, and T. R. Hahn, "Propeller noise from a light aircraft for low-frequency measurements of the speed of sound in a marine sediment," J. Comput. Acoust. **10**, 445–464 (2002).
[21] C. L. Pekeris, "Theory of propagation of explosive sound in shallow water," in *Geological Society of America Memoir, Propagation of Sound in the Ocean* (Geological Society of America, New York, 1948), Vol. **27**, pp. 1–117.
[22] M. J. Buckingham and E. M. Giddens, "On the acoustic field in a Pekeris

waveguide with attenuation in the bottom half-space," J. Acoust. Soc. Am. **119**, 123–142 (2006).
[23]G. N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd ed. (Cambridge University Press, London, 1958).
[24]W. M. Ewing, W. S. Jardetzky, and F. Press, *Elastic Waves in Layered Media* (McGraw-Hill, New York, 1957).
[25]L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves* (Prentice–Hall, Englewood Cliffs, NJ, 1973).
[26]G. Arfken, *Mathematical Methods for Physicists*, 3rd ed. (Academic, San Diego, 1985).
[27]Z. Y. Zhang and C. T. Tindle, "Complex effective depth of the ocean bottom," J. Acoust. Soc. Am. **93**, 205–213 (1993).
[28]D. E. Weston, "A Moiré fringe analog of sound propagation in shallow water," J. Acoust. Soc. Am. **32**, 647–654 (1960).
[29]M. J. Buckingham, "Array gain of a broadside vertical line array in shallow water," J. Acoust. Soc. Am. **65**, 148–161 (1979).
[30]G. V. Frisk, *Ocean and Seabed Acoustics: A Theory of Wave Propagation* (Prentice–Hall, Englewood Cliffs, NJ, 1994).
[31]C. T. Tindle, A. T. Stamp, and K. M. Guthrie, "Virtual modes and the surface boundary condition in underwater acoustics," J. Sound Vib. **49**, 231–240 (1976).
[32]C. T. Tindle, A. T. Stamp, and K. M. Guthrie, "Virtual modes and mode amplitudes near cutoff," J. Acoust. Soc. Am. **65**, 1423–1428 (1979).
[33]O. A. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media. II: Point Sources and Bounded Beams* (Springer-Verlag, Berlin, 1999).
[34]M. J. Buckingham, "Acoustical remote sensing of the sea bed using propeller noise from a light aircraft," in *Sounds in the Sea: From Ocean Acoustics to Acoustical Oceanography*, edited by H. Medwin (Cambridge University Press, Cambridge, 2005), pp. 581–597.

Regions that influence acoustic propagation in the sea at moderate frequencies, and the consequent departures from the ray-acoustic description

John L. Spiesberger
Department of Earth and Environmental Science, 240 S. 33rd Street, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6316

(Received 4 April 2004; revised 23 July 2006; accepted 24 July 2006)

In the limit where a transient signal is comprised of very large frequencies, spatial regions within an inhomogeneous medium that influence the propagation from a source to a receiver lie along one or more ray paths. At lower frequencies for which the geometrical acoustic approximation is of borderline applicability, the regions that influence such transient signals are extended because of diffraction. Previous research has addressed the numerical determination of those spatial regions that influence propagation at low frequency. The present paper addresses the question of how high the center frequency need be so that the regions of influence are nearly described as ray paths for a model ocean in which the speed of sound increases nearly linearly with depth from a perfectly reflecting surface. Computations indicate that near 2500 Hz and at a range of 50 km, the region of influence resembles a ray. Noticeable departures from the ray picture are found at a range of 500 km. Various physical and mathematical causes for the departures from the ray propagation model for lower frequencies and for greater ranges are identified and discussed. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2336991]

PACS number(s): 43.20.El [ADP] Pages: 1842–1850

I. INTRODUCTION

Quantifying where a wavelike signal is influenced between a source and receiver by a medium and its fluctuations is useful in acoustics, communication, scattering theory, optics, and fluid and cosmic gravity waves problems. At infinite frequency, the received signal is influenced only on one or more infinitesimally thin rays. At finite frequencies, the notion of influence is often quantified by considering experiments in which a screen is placed between a source and receiver. For homogeneous media, one imagines increasing the radius of a circular opening in the screen until the received signal is similar to that obtained without a screen. For transmission at a single frequency, the first Fresnel zone provides a radius that approximately admits a field similar to that found without a screen.[1] There remains a bright spot at the center that reduces to that found without a screen when the radius is much larger.[1] When a signal has a nonzero bandwidth, the free-field solution is obtained exactly, within a pulse resolution of the time of the direct arrival, when the radius is given by the zone of influence,[2–6] $R_I(x) \cong (2cTx(d-x)/d)^{1/2}$. Here, the speed of the wave is c, its temporal resolution is T, the distance between source and receiver is d, and x measures distance from the source to receiver. Circular or otherwise shaped openings interfere with the transmitted signal, and in particular their edges lead to edge-diffracted rays.[1,7] Such rays are entirely caused by diffraction and appear to exist for all edges even when the screen is perfectly absorbing.[1,7] When the transmission consists of a single frequency, edge-diffracted rays cause the anomalously bright spot mentioned above. For transmissions with nonzero bandwidth, the edge-diffracted rays arrive exactly one pulse resolution later than the signal traveling directly between the source and receiver when the circular opening has radius given by the zone of influence.

For any transient signal at finite frequencies, an exact method has been developed to compute the region in any medium that significantly influences the received signal for any specified window of signal travel time.[8] This window is sometimes chosen to surround a peak. Results at finite frequencies differ from those at infinite frequency because of diffraction. The purpose of this paper is to investigate how high the center frequency of a signal need be in an ideal oceanic acoustic waveguide to yield a region of influence that resembles one or more ray paths. For homogeneous media, the exact method is different than the idea of a zone of influence. Perhaps the most interesting difference is that the method allows one to look within the zone of influence to see at high spatial resolution how each region of space influences the signal. This high-resolution picture is possible to compute for inhomogeneous media as well.[8]

Solutions are given for 50- and 500-km ranges of propagation. A waveguide speed is chosen to increase nearly linearly with depth such that an analytical solution to the wave equation is available. Results based on the analytical solution are compared with those derived from the sound-speed insensitive parabolic approximation.[9] If these results are similar, then results previously given for this approximation are probably accurate.[8]

James Bowlin pioneered the use of the Huygens-Fresnel principle (pp. 370–375 in Ref. 1) to estimate paths of transient signals between a source and a receiver in the ocean.[10] The results in Ref. 8 expand on his ideas in three ways. They are the following: (1) The integral theorem of Helmholtz and

 0001-4966/2006/120(4)/1842/9/$22.50

Kirchhoff is used to calculate effects of diffraction. This theory is more accurate than the Huygens-Fresnel principle.[1] (2) A method is identified for obtaining quantitative values of medium influence on the received signal at high spatial resolution within the domain. (3) An interference filter is found useful for visualizing regions of space that significantly influence the received signal. The calculations in this paper use these three methods. We also compare results from the Huygens-Fresnel principle with those obtained from the integral theorem of Helmholtz and Kirchhoff.

II. SUMMARY OF METHOD FOR ESTIMATING REGION OF INFLUENCE

We summarize the methods[8] used to compute a region of influence for a transient signal in any medium. Suppose an infinitely large opaque screen is perpendicular to the x axis at x_{sc}. The screen is in-between a source and receiver. The z axis is parallel to the screen. The integral theorem of Helmholtz and Kirchhoff[1] yields the contribution to the time series at time t at the receiver from the wave field passing through a transparent opening of the screen between coordinates z_r and z_s,

$$e(t,x_{sc},z_r,z_s) = B\int_{z_r}^{z_s}\int_{-\infty}^{\infty} H_1(x_{sc},z,\omega)\exp(-i\omega t)d\omega\, dz, \tag{1}$$

where B is a normalization constant. The radian frequency is ω and

$$H_1(x_{sc},z,\omega) = W_1(x_{sc},z,\omega)\frac{\partial W_0(x_{sc},z,\omega)}{\partial x} - W_0(x_{sc},z,\omega)\frac{\partial W_1(x_{sc},z,\omega)}{\partial x}. \tag{2}$$

The solutions of the Helmholtz equation on the screen due to emissions located at the source and receiver are $W_0(x_{sc},z,\omega)$ and $W_1(x_{sc},z,\omega)$ respectively.[1] We will also work with

$$G(t,x_{sc},z_r,z_s) = \mathcal{F}[e(t,x_{sc},z_r,z_s)], \tag{3}$$

where $\mathcal{F}$ removes the carrier frequency via complex demodulation of analytic signals [Eq. (47) of Ref. 9]. Results using the Huygens-Fresnel principle[1] with an inclination factor of unity can be obtained by substituting

$$H_2(x_{sc},z,\omega) = W_0(x_{sc},z,\omega)W_1(x_{sc},z,\omega), \tag{4}$$

for $H_1(x_{sc},z,\omega)$ in Eq. (1).

The "region of influence" denotes locations through which a transient signal travels that significantly affect the received signal within a specified window of travel time. This region is different than the concept for the zone of influence, which yields a time series identical to that found without a screen within the temporal resolution of the signal. For narrow-band signals in inhomogeneous media, it could be necessary to extend the region toward infinity to guarantee reception of an identical signal. Instead, the region of influence's boundaries demarcate locations of significant contribution.

A term is needed to specify what is being influenced in the received time series. We call it the "measure of influence." Two such measures are defined. The measure of influence of the first type is the largest peak in the time series obtained from apertures $z=0$ through $z=z_j$,

$$M_1(t_0,x_{sc},0,z_j) \equiv \max_{t\in(t_0\pm\delta t/2)}\{V(t,x_{sc},0,z_j)\}. \tag{5}$$

where $V(t,x_{sc},0,z_j)$ is some function of the received time series within the window of travel time of duration δt. The measure of influence of the second type is the energy of the function of the time series in the window,

$$M_2(t_0,x_{sc},0,z_j) \equiv \int_{t_0-\delta t/2}^{t_0+\delta t/2} V^2(t,x_{sc},0,z_j)dt. \tag{6}$$

Examples of $V(t,x_{sc},0,z_j)$ are the time series with and without the carrier frequency [Eqs. (1) and (3)], and a time series that is adjusted for interference with a filter, $\mathcal{I}$,

$$\int_{z_r}^{z_s}\int_{-\infty}^{\infty}\mathcal{I}\{H_1(x_{sc},z,\omega)\}\exp(-i\omega t)d\omega\, dz, \tag{7}$$

discussed later.

The "differential measure of influence" is obtained using a first difference of the measure of influence as

$$\delta M_k(t_0,x_{sc},z_j) = \begin{cases} M_k(t_0,x_{sc},0,z_1) & \text{if } j=1, \\ M_k(t_0,x_{sc},0,z_j) - M_k(t_0,x_{sc},0,z_{j-1}) & \text{if } j>1, \end{cases} \tag{8}$$

where $k=1,2$ denotes measures of the first and second types, respectively. The differential measure of influence contains the information needed to quantify the influence of the signal passing through any aperture of a screen on the received signal. However, when the received phase changes quickly from signals passing through nearby apertures, it can be difficult to visualize regions that yield a significant contribution. The interference filter is designed to make it easier to understand what is happening in this situation. The interference filter outputs only net positive contributions to the measure of influence.[8] The region of influence so filtered is called the "net region of influence."[8] The weighted interference filter is used here because it is more accurate than the unweighted filter.[8]

We choose boundaries for the region of influence by including significant contributions to the measure of influence. For any screen, the contributions from all apertures yield the final value for the measure of influence, $M_i(t_0,x_{sc},0,D)$, where the bottom of the screen is $z=D$. The absolute values of the differential measure of influence are sorted into decreasing order. The cumulative sum of the first N sorted apertures is

$$\mathcal{L}(N) \equiv \sum_{n=1}^{N}\frac{|\delta M_k(t_0,x_{sc},z_{g(n)})|}{|M_k(t_0,x_{sc},0,D)|}, \tag{9}$$

where $g(n)$ denotes sorted order. With the fractional amplitude method, we say that the measure of influence is reconstructed with a fidelity, f, by choosing the smallest value of

$N=1,2,3,\ldots$ such that $\mathcal{L}(N)$ exceeds f without any value $\mathcal{L}(N+p)$ being less than f for integer p greater than zero. For example, if we wish to reconstruct the measure of influence with a fidelity of 0.9, f is 0.9. Perfect fidelity corresponds to $f=1$.

III. SHORTER DISTANCE PROPAGATION

A source and receiver are placed at a depth of 5 m and separated by 50 km. The speed of sound is taken to increase with depth, z, almost linearly according to

$$c(z) = c_0/\sqrt{1-2az}, \quad z \leqslant 1/(2a) \tag{10}$$

because this yields an analytical solution[11] to Helmholtz's equation in cylindrical coordinates,

$$\frac{\partial^2 W(r,z,\omega)}{\partial r^2} + \frac{1}{r}\frac{\partial W(r,z,\omega)}{\partial r} + \frac{\partial^2 W(r,z,\omega)}{\partial z^2} + k^2(z)W(r,z,\omega) = -\frac{2}{r}\delta(z-z_s)\delta(r). \tag{11}$$

The depth of the source is z_s and it is at a radius, r, of zero. The acoustic wave number is $k=\omega/c(z)$.

A. Exact solution from normal modes

Pressure perturbations vanish at the surface and at infinite z so $W(r,0,\omega)=0$ and $\lim_{z\to\infty} W(r,z,\omega)=0$. The solution[11] to Helmholtz's equation is

$$W(r,z,\omega) = \pi i \sum_{l=1}^{\infty} \frac{\mathrm{Ai}(z_s/H-y_l)\mathrm{Ai}(z/H-y_l)H_0^{(1)}(\xi_l r)}{\int_0^{1/2a} \mathrm{Ai}^2(z/H-y_l)dz}, \tag{12}$$

where i is $\sqrt{-1}$, l is mode number, Ai is the Airy function given by Eq. (10.4.2) in Ref. 12, $-y_l$ is the lth zero of the Airy function (all zeros are negative so y_l is positive), z_s is the depth of the acoustic source,

$$H \equiv (2ak_0^2)^{-1/3}, \tag{13}$$

$$\xi_l^2 = k_0^2 - y_l/H, \tag{14}$$

and $k_0 \equiv \omega/c_0$. The time series at a receiver at cylindrical coordinate (r_s,z_r) is obtained from the inverse Fourier transform of Eq. (12).

The group speed of mode l is

$$c_g = c_0 \times (1 - y_l(2ac_0)^{2/3}\omega^{-2/3})^{1/2}(1 - 2y_l(2ac_0)^{2/3}\omega^{-2/3}/3)^{-1}, \tag{15}$$

(Eq. 6.6.24 of Ref. 11). These speeds are used to estimate the mode numbers needed to accurately compute the impulse response for any desired window of travel time about a selected peak. The lth vertical mode, which is the Airy function, $\mathrm{Ai}(z/H-y_l)$, changes from oscillating to exponentially decaying when its argument is zero, i.e., $\check{z}=y_l H$. Since the zeros of Ai are negative real numbers,[12] with $y_{l+1}>y_l$, the turning depths are positive and the turning depth of mode $l+1$ is greater than mode l.

B. Approximate solution from parabolic approximation

The sound-speed insensitive parabolic approximation[9] yields accurate travel times, is efficient due to its split-step algorithm, and obeys reciprocity. It is important that reciprocity is obeyed because a proof[1] for reciprocity is provided by the integral theorem of Helmholtz and Kirchhoff. In this paper the computational grids in depth and range are made sufficiently small to achieve convergence of the solution. The parabolic approximation yields estimates of $W_i(x_{sc},z,\omega)$, $i=0,1$.

This algorithm for the parabolic approximation requires the depth of the bottom and the geoacoustic properties of the subbottom. The subbottom, starting at 5500-m depth, is constructed to absorb incident energy in the following way. The sediment thickness is 300 m. The speed of sound at the top of the sediment divided by that at the bottom of the water column is 0.5. This causes incident energy to refract downwards into the sediment. The density of the sediment and water are taken to be equal to minimize reflections at the interface. The attenuation in the sediment is $\alpha(f)=0.2f$ (dB/m), where f is acoustic frequency in kHz. The derivative of speed with depth is 0.001 s^{-1} in the sediment. The density of the lower basement layer is two and one-half times that of the water. The speed of sound in the basement is twice that at the bottom of the sediment layer. The attenuation in the basement is $\alpha(f)=0.5f^{0.1}$ (dB/m). Diffracted regions will only be considered that do not interact with the bottom because the bottom is absent in the solution based on normal modes.

C. Solution from ray approximation

Regions of influence are compared with rays. The ray program, zray, and its eigenray finder have been described and successfully used to study acoustic propagation for many experiments.[13,14] Its results agree with analytical solutions. Sound speeds used by the ray program are the same as those used by the parabolic approximation on its computational grid. Between grid points, the speed is obtained using a quadratic spline. The spline goes through each grid point without gradient discontinuities.[13]

D. Results

In Eq. (10), we set $c_0=1500$ m s^{-1} and $a=1.2\times10^{-5}$ m^{-1}. The speed increases by about 18 m s^{-1} per 1000 m of depth. Sidelobes are minimized in the time domain by applying a Hann taper in the frequency domain between $f_c\pm20$ Hz where the center frequency is f_c. The taper is zero at $f_c\pm20$ Hz, yielding an effective bandwidth of 20 Hz and a time resolution of $\frac{1}{20}=0.05$ s.

Regions of influence are computed for the pulse having a travel time near 33.2 s. For ray theory, the travel time of this pulse is 33.205 s. The lower turning depth of the ray is about 1000 m. It reflects once from the surface [Figs. 1(a) and 1(f)]. At a pulse resolution of 0.05 s, it is almost temporally resolved from its nearest ray arrival at 33.277 s that reflects twice from the surface [Figs. 1(a) and 1(f)]. These

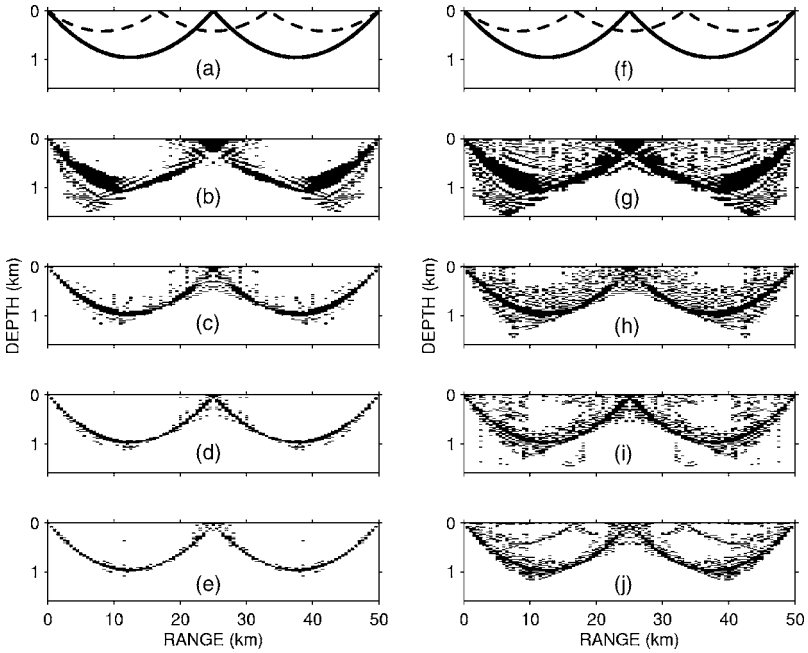


FIG. 1. (a, b) Two rays between source and receiver at 5-m depth and 50-km separation compared with the constructive region of influence as a function of center frequency. The simulated signal has a bandwidth of 20 Hz, and a center frequency of 100, 500, 1250, and 2500 Hz in panels (b)–(e) and (g)–(j), respectively. Results are for the differential measure of influence of the first type (highest peak within the window of travel time) and the Huygens-Fresnel principle with inclination factor equal to unity [Eq. (8) for $k=1$ using Eq. (4)]. The left (right) column shows regions that contribute 0.9 (0.99) of the amplitude of the peak within the window of travel time. The speed of sound varies with depth according to Eq. (10) with $c_0=1500$ m s^{-1} and $a=1.2\times10^{-5}$ m^{-1}.

arrivals appear to be temporally resolved in the impulse responses of all solutions (not shown). As will be seen, however, a little energy from the later arrival leaks into the time window surrounding the earlier arrival.

The region of influence is estimated using the Huygens-Fresnel principle [Eq. (4)] at center frequencies of 100, 500, 1250, and 2500 Hz. We consider the energy arriving within ±0.025 s of the peak corresponding to the arrival at 33.2 s. Solutions for $W_0(x_{sc},z,\omega)$ and $W_1(x_{sc},z,\omega)$ are provided from normal modes. The fractional amplitude method is used to reconstruct the region of influence for fidelities of $f=0.9$ and $f=0.99$. Both ray paths are visible at 2500 Hz with a fidelity of 0.99 [Figs. 1(j) and 2(j)]. At a fidelity of 0.9, the ray that bounces twice from the surface is barely visible at this frequency [Figs. 1(e) and 2(e)]. At a fidelity of 0.99 and a center frequency of 1250 Hz (panel i), the region of influence has features that look unlike a ray. For example, sound hugs the surface over a distance of about 5 km where it reflects from the surface near 25 km of range. Other regions of space do not correspond to a ray path. At 2500 Hz, departures from ray paths are more evident for regions constructed with a fidelity of 0.99 rather than 0.9 [Figs. 1(e), 1(j), 2(e), and 2(j)]. Regions of influence look more raylike as the center frequency increases.

Constructive and destructive regions of influence are shaded as black and gray respectively in Figs. 1 and 2. They are most evident at 100 Hz where they occur as interleaving filaments. It is important to note that Figs. 1 and 2 do **not** show paths of sound between the source and receiver. Rather they shows constructive and destructive regions of influence.

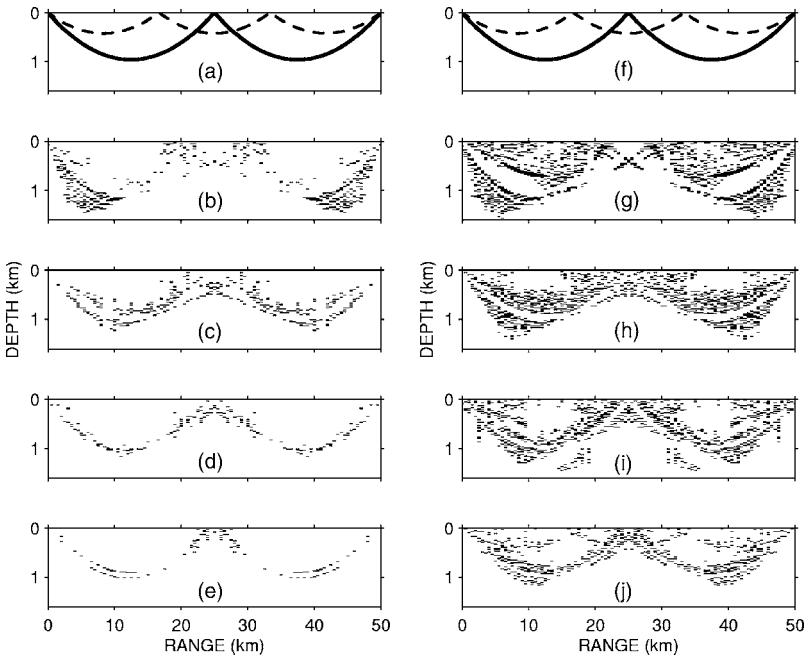


FIG. 2. Same as Fig. 1 except only destructive regions of influence are shown (gray) in (b)–(e) and (g) and (h).

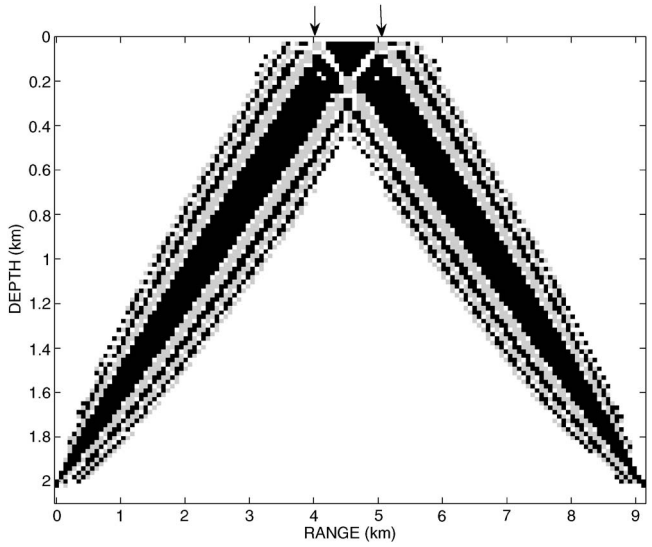


FIG. 3. Region of influence for propagation of path with one surface reflection in an otherwise homogeneous medium with wave speed 1.5 km/s. The source and receiver are at depths of 2 km. To minimize sidelobes in the time domain, a Hann taper is applied in the frequency domain to the emitted signal between 100±40 Hz. The taper is zero at 60 and 140 Hz, yielding an effective bandwidth of 40 Hz and a time resolution of about $\frac{1}{40}=0.025$ s. The arrows indicate where two destructive paths of influence approach the surface. Adapted from Fig. 6 of Ref. 8.

For example, a region of influence shaded black at some range means that the signal from the source passing through an aperture on a screen corresponding to the black region has a significant influence on the measure of influence at the receiver. Where the signal goes before or after the aperture is a different question.[7]

The region of influence near the surface reflection shows a black triangular region [Figs. 1(b) and 1(g)]. It looks like that seen for a single reflection from a flat interface in homogeneous media (Fig. 3). For this later case, this triangle is carved out of the main region that corresponds to a specularly reflecting ray. There are two intense destructive regions of influence whose corresponding paths of influence arrive with phases one-half cycle greater than the signal from the main region.[8] The coherent addition of waves from the main region and those corresponding to the destructive region of influence destructively interfere across the main region. This causes the triangular region at the surface that is carved out of the main region.

The intersection of the destructive region of influence with the surface corresponds to strong destructive paths of influence that seem to reflect from the surface at locations indicated by arrows in Fig. 3. However, they are caused by edge-diffracted rays described by the geometrical theory of diffraction.[7,8] In Fig. 3, we see other constructive and destructive regions of influence. Proceeding to the left from the leftmost arrow, we see alternating destructive and constructive regions corresponding to paths reflecting from the surface indicated in gray and black, respectively. The phases of the waves corresponding to these paths are 0.5, 1, 1.5, 2, and 2.5 cycles greater than the phase along the main region.[8] Their received amplitudes decrease as the cycle difference increases with respect to the main region because they arrive later than the waves corresponding to the main region. Eventually, they arrive so late that their influence diminishes to zero within the selected window of travel time. A similar interpretation applies to the pattern for the region of influence for waves seen in the inhomogeneous waveguide [Figs. 1(b), 1(g), 2(b), and 2(g)]. These constructive and destructive regions of influence are entirely due to diffraction. Their influence diminishes as the center frequency increases (Figs. 1 and 2).

The differential measure of influence given by Eq. (8) has all the information needed to quantify significant contributions to the measure of influence. But we may not be interested in cases where adjoining constructive and destructive regions of influence lead to little net effect on the measure of influence. An interference filter can be applied to the differential measure of influence to guide intuition for

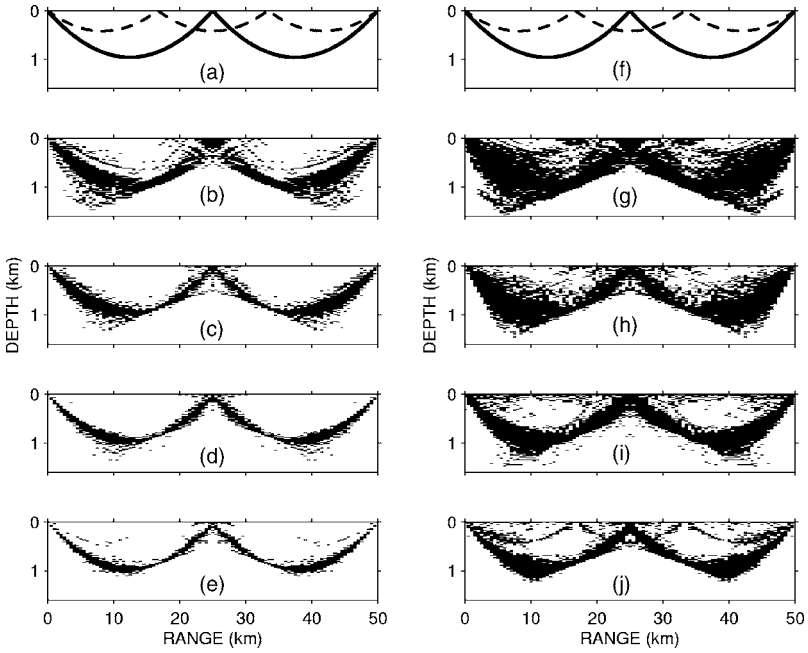


FIG. 4. Same as Figs. 1 and 2 except the interference filter is used.

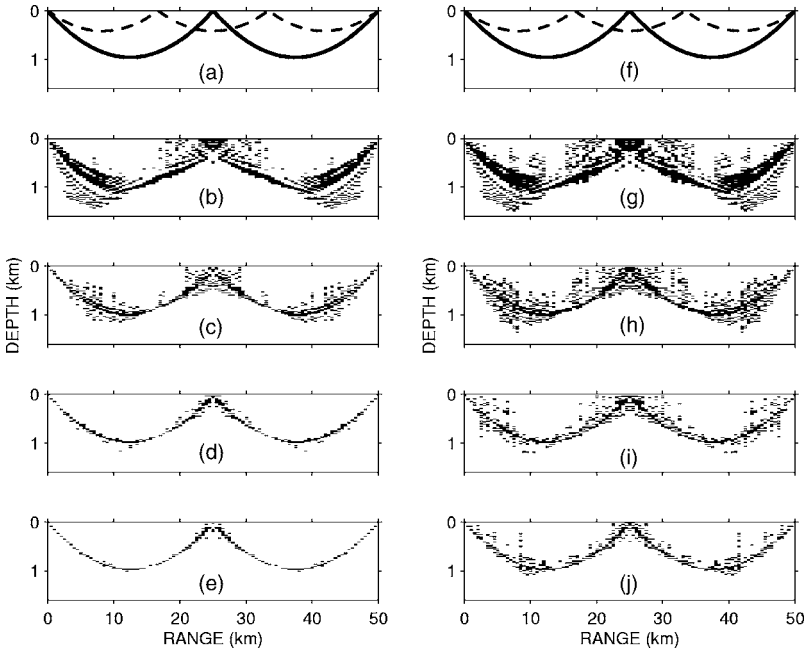


FIG. 5. Same as Fig. 1 except the measure of influence of the second type is used (energy of time series within window of travel time).

regions of influence that are significant[8] (Fig. 4), We see that the filtered region of influence becomes more raylike at higher frequency. The constructive paths of influence have more influence than the destructive paths of influence. This fact leads to the resemblance of the filtered and unfiltered regions of influence. The constructive paths of influence are partially canceled out by the destructive paths of influence, but not totally so. Waves corresponding to the main region of influence (specular reflection) have the largest amplitudes at the receiver. The strongest destructive path of influence is only one-half wavelength longer than the path for the main region. Thus, waves corresponding to the first destructive path of influence undergo about the same geometrical spreading loss as waves corresponding to the main region. However, the first destructive path of influence arrives later than waves from the main region, leading to a significant amount of its energy arriving later than the selected window of travel time. Thus, the net contribution from the main region and first destructive path of influence is won by waves corresponding to the main region. The same explanation applies to subsequent pairs of constructive and destructive paths of influence. The next pair of constructive and destructive paths of influence is dominated by the constructive path because it arrives less late than its destructive partner. The later the arrival, the less the influence in the selected window of travel time. If the constructive and destructive paths of influence arrived almost entirely within the selected window of travel time, their effects on the measure of influence would be insignificant because they would cancel each other. The interference filter is telling us that the constructive paths dominate the destructive paths of influence.

The region of influence looks about the same when we use the measure of influence of the second type (energy of time series within window of travel time) (Figs. 5 and 6). It is interesting that the weak ray path with two surface reflections does not appear with stronger contribution for the chosen values of fidelity. Its effect is still there [e.g., Figs. 5(i) and 6(i)] but is weak. Application of the interference filter yields a region of influence much like that seen for the first measure of influence (Fig. 7). As before, a center frequency of 2500 Hz looks much more raylike than at 100 Hz.

We compare the region of influence at 100 Hz using the exact solution of the wave equation (normal modes) with the solution using the sound-speed insensitive parabolic approximation[9] [Figs. 8(a) and 8(b)]. Panel (a) is the same as Fig. 4(b). Results are similar, which means that this parabolic approximation is accurate in this application. When we use the more accurate theory of diffraction based on the integral theorem of Helmholtz and Kirchhoff,[1] results are very similar [Fig. 8(c)]. The reason for the similarity is that acoustic waves are propagating in a nearly horizontal direction. Thus the effect of the inclination factor[1] in the Huygens-Fresnel theory is small. We have set the inclination factor to unity, which is evidently accurate for this case.

IV. LONGER DISTANCE PROPAGATION

The region of influence at 2500 Hz is computed for a case that is identical to Sec. III D except the range between the source and receiver is 500 km instead of 50 km. A temporally resolved peak is chosen from the normal mode solution having a travel time of 331.7625 s. The neighboring peaks at 331.5625 and 331.9250 s are much further away than the pulse resolution of 0.05 s centered on the peak at 331.7625 s. A raytrace shows the peak to correspond to a single ray leaving the source downward at an angle of 9.7166 deg with a travel time of 331.755 s [Fig. 9(a)]. This differs slightly from the travel time of 331.7625 s computed via normal modes because of diffraction. The net region of influence is estimated using the Huygens-Fresnel principle and the fractional amplitude method as before. The fidelity of the fractional amplitude method is $f=0.9$. We use the measure of influence of the first type (highest peak in selected window of travel time). It looks more like a ray near the

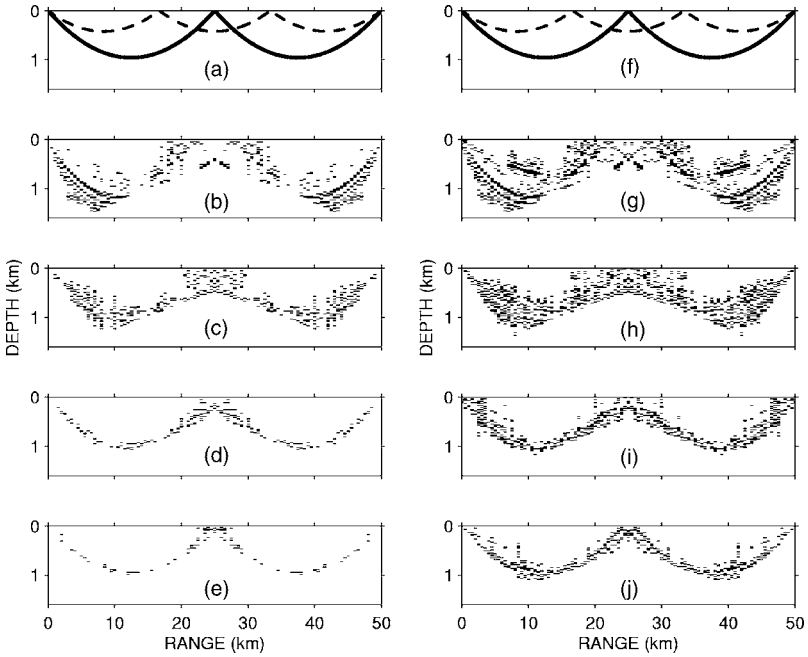


FIG. 6. Same as Fig. 2 except the measure of influence of the second type is used (energy of time series within window of travel time).

source and receiver than in the central regions, where it exhibits many unraylike features such as diffuse regions near the surface [Fig. 9(b)]. Evidently, a center frequency of 2500 Hz is too low for the region of influence to look like a ray. If the fidelity of the reconstruction was increased to $f=0.99$, the region of influence would appear even more unlike a ray, just as seen for the cases at a distance of 50 km. Because the computations are lengthy at 2500 Hz, calculations are not attempted at higher frequency. This simulation uses the first 1000 vertical modes at each acoustic frequency so as to guarantee an accurate solution for Helmholtz's equation.

V. DISCUSSION AND CONCLUSION

We investigated the regions of space that influence the waveform of a transient signal traveling between a source and receiver at 50 and 500 km in an idealized waveguide. At infinite frequency, these regions of influence coincide with one or more ray paths. At finite frequency, the regions of influence depart from the ray picture because of diffraction. At a distance of 50 km, we found it necessary to go to a center frequency near 2500 Hz to obtain a region of influence that resembles a ray. The region of influence could be accurately calculated for any mix of approximations that include the sound-speed insensitive parabolic approximation and the Huygens-Fresnel theory of diffraction using an inclination factor of unity. These approximate solutions are very similar to solutions based on an exact solution of the wave equation via normal modes and the integral theorem of Helmholtz and Kirchhoff. This theorem automatically computes the correct inclination factor on each part of a screen.[1] At a center frequency of 2500 Hz, the region of influence

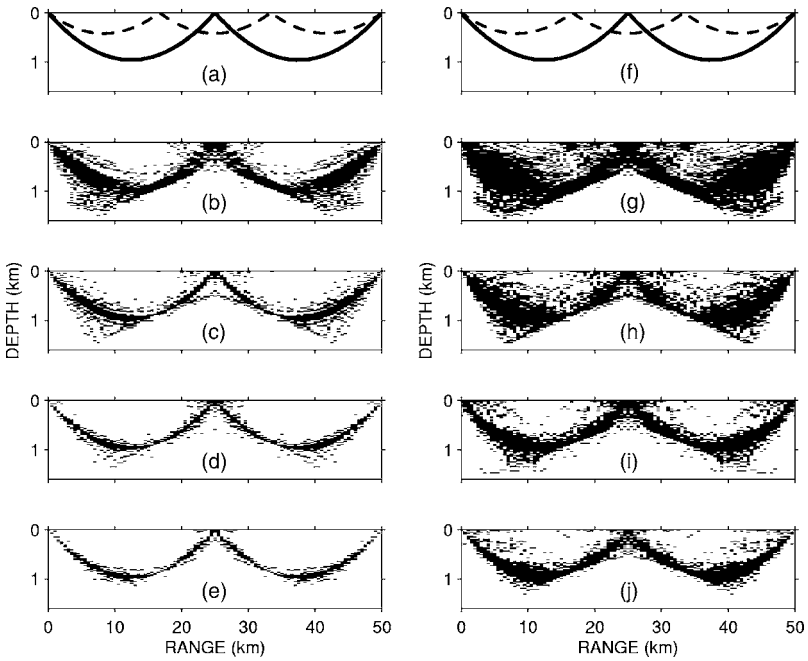


FIG. 7. Same as Figs. 5 and 6 except the interference filter is used.

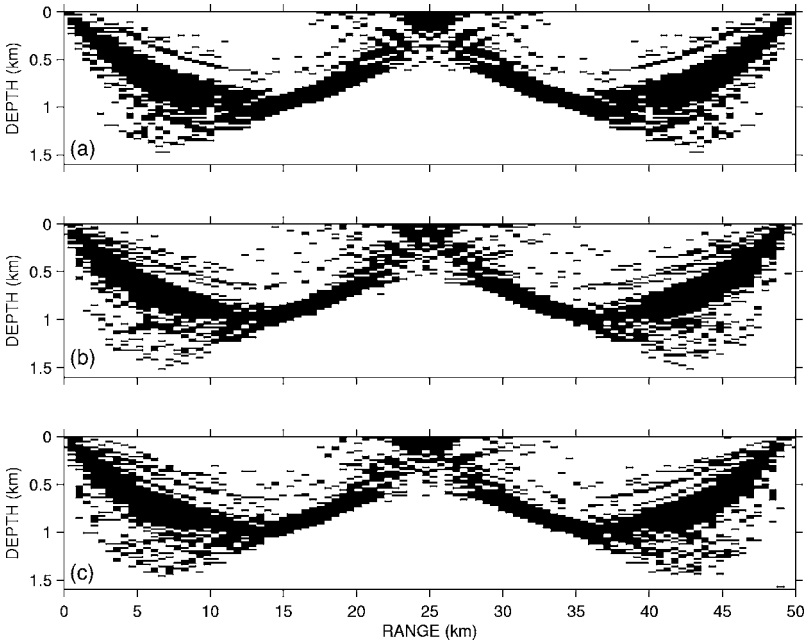


FIG. 8. (a) Region of influence computed with exact solution of wave equation (normal modes, 100 Hz center frequency, 20 Hz bandwidth) and Huygens-Fresnel principle for the measure of influence of the first type (highest peak in selected window of travel time). The interference filter has been applied so this is identical to Fig. 4(b). The region of influence is reconstructed with a fidelity of $f=0.9$. (b) Same except the wave equation is solved with the sound-speed insensitive parabolic approximation.[9] (c) Same as (b) except effects of diffraction are computed from the integral theorem of Helmholtz and Kirchhoff.[1]

can look like a ray at a distance of 50 km, but can exhibit nonraylike features at 500 km. We conclude that both the center frequency and distance of wave propagation are factors contributing to departures from a ray path.

Regions of influence were previously computed using theories of diffraction based on the Huygens-Fresnel principle and the integral theorem of Helmholtz and Kirchhoff.[8,15] These theories are implemented by solving the Green's function at a screen from both the source and receiver for many frequencies. This paper demonstrates that the sound-speed insensitive parabolic approximation[9] yields Green's functions similar to those computed from exact solutions of the wave equation using normal modes. The previous computations for regions of influence[8,15] are further validated here because the previous computations used the sound-speed insensitive parabolic approximation.

We found it possible to explain some features of the region of influence at low frequencies near 100 Hz with the idea of constructive and destructive paths of influence. It appears that some but not all paths of influence are caused by edge-diffracted rays.[7,8] The constructive and destructive paths of influence significantly affect both measures of influence used in this study. Their effects grow smaller as the center frequency increases to 2500 Hz. Since these paths are due to diffraction only, they must vanish at sufficiently high frequencies.

It appears that for some cases at 50-km range, the region of influence looks much like a ray at frequencies above 2500 Hz, even when the fidelity of the measure of influence is reconstructed within 99% ($f=0.99$) of the complete mea-

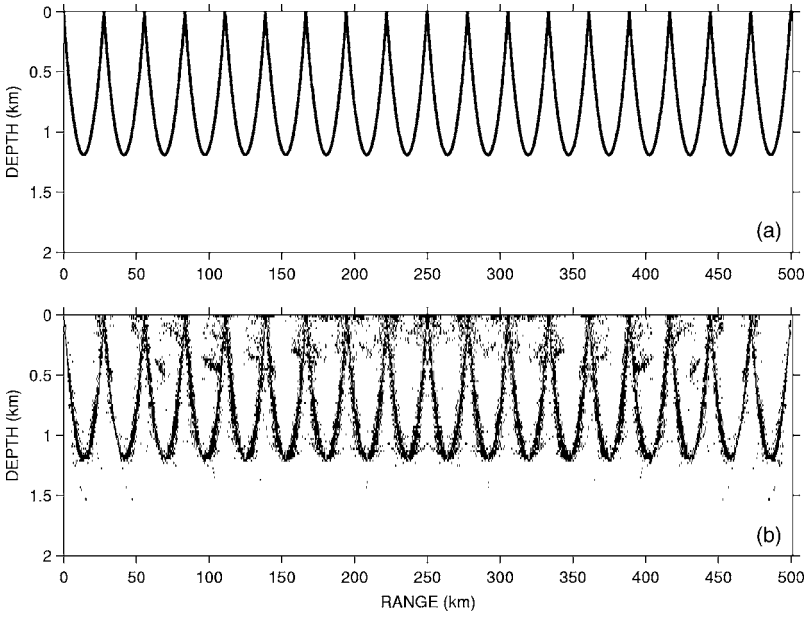


FIG. 9. (a) Ray between source and receiver at 5-m depths and 500-km distance. Sound travels this path in 331.755 s. The sound speed field is the same as used for Fig. 1. (b) The corresponding region of influence for energy centered at 2500 Hz arriving within a pulse resolution of 0.05 s centered on the peak of the impulse response at 331.755 s. The acoustic fields are computed exactly using normal modes. The region of influence is estimated using the interference filter, the Huygens-Fresnel principle, and the fractional amplitude method with a fidelity of $f=0.9$. The net region of influence has many un-ray-like features at this distance, despite the fact that the center frequency is large.

sure of influence. At distances of 500 km, it appears necessary that the center frequency be much larger to yield a region of influence that looks like a ray.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research Contracts No. N00014-03-C-0155, and No. N00014-06-C-0031, and by a grant of computer time from the DOD High Performance Computing Modernization Program at the Naval Oceanographic Office. I thank the reviewers and editor for their comments.

[1]*Principles of Optics: Electromagnetic Theory of Propagation, Interference, and Diffraction*, M. Born and E. Wolf, with contributions by A. Bhatia *et al.* (Cambridge U. P. Cambridge, 1999).
[2]A. W. Trorey, "A simple theory for seismic diffractions," Geophysics **35**, 762–784 (1970).
[3]R. W. Knapp, "Fresnel zones in the light of broadband data," Geophysics **56**, 354–359 (1991).
[4]M. Bruhl, G. J. O. Vermeer, and M. Kiehn, "Fresnel zones for broadband data," Geophysics **61**, 600–604 (1996).
[5]B. R. Zavalishin, "Diffraction problems of 3D seismic imaging," Geophys. Prospect. **48**, 631–645 (2000).
[6]J. Pearce and D. Mittleman, "Defining the Fresnel zone for broadband radiation," Phys. Rev. E **66**, 056602 (2002).
[7]J. B. Keller, "Geometrical theory of diffraction," J. Opt. Soc. Am. **52**, 116–130 (1962).
[8]J. L. Spiesberger, "Regions where transient signals are influenced between a source and receiver," Waves Random Media **16**, 1–21 (2006).
[9]F. Tappert, J. L. Spiesberger, and L. Boden, "New full-wave approximation for ocean acoustic travel time predictions," J. Acoust. Soc. Am. **97**, 2771–2782 (1995).
[10]J. Bowlin, "Generating eigenray tubes from two solutions of the wave equation," J. Acoust. Soc. Am. **89**, 2663–2669 (1991).
[11]L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics* (Springer-Verlag, New York, 1991), p. 270.
[12]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972), p. 1046.
[13]J. B. Bowlin, J. L. Spiesberger, and L. F. Freitag, "Ocean acoustical ray-tracing software RAY," Woods Hole Oceanographic Technical Rept., WHOI-93–10, Woods Hole, MA. (1992).
[14]J. L. Spiesberger, "An updated perspective on basin-scale tomography," J. Acoust. Soc. Am. **109**, 1740–1742 (2001).
[15]J. L. Spiesberger, "Locating where transient signals travel in inhomogeneous media," http://www.arxiv.org/abs/physics/0501162 (2005).

The use of microperforated plates to attenuate cavity resonances

Benjamin Fenech[a)]
Acoustic Technology, Ørsted·DTU, Technical University of Denmark, Building 352, Ørsteds Plads, DK-2800 Kgs. Lyngby, Denmark

Graeme M. Keith[b)]
Ødegaard & Danneskiold-Samsøe, Titangade 15, DK-2200 Copenhagen N, Denmark

Finn Jacobsen[c)]
Acoustic Technology, Ørsted·DTU, Technical University of Denmark, Building 352, Ørsteds Plads, DK-2800 Kgs. Lyngby, Denmark

(Received 31 October 2005; revised 30 June 2006; accepted 3 July 2006)

The use of microperforated plates to introduce damping in a closed cavity is examined. By placing a microperforated plate well inside the cavity instead of near a wall as traditionally done in room acoustics, high attenuation can be obtained for specific acoustic modes, compared with the lower attenuation that can be obtained in a broad frequency range with the conventional position of the plate. An analytical method for predicting the attenuation is presented. The method involves finding complex eigenvalues and eigenfunctions for the modified cavity and makes it possible to predict Green's functions. The results, which are validated experimentally, show that a microperforated plate can provide substantial attenuation of modes in a cavity. One possible application of these findings is the treatment of boiler tones in heat-exchanger cavities. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2258438]

PACS number(s): 43.20.Hq, 43.50.Gf [RR] Pages: 1851–1858

I. INTRODUCTION

For a number of years, microperforated plates have been used as an alternative to fibrous absorptive materials to provide absorption at the boundary surfaces to acoustic cavities. Microperforated plates are particularly well suited to applications requiring tolerance to high temperatures, such as the heat exchanger cavities of boilers, and sterile environments where the introduction of small fibrous particles is unacceptable.

Traditionally, the plates are located a fixed distance from the wall of the cavity and their acoustic characteristics are described in terms of the reflection and absorption of incoming acoustic waves at the plate. This approach is particularly effective at high frequencies where a large number of modes play a significant role in the acoustic behavior of the cavity, and the direction of the incoming waves on the plate can be thought of as being distributed in an almost continuous way, in other words, when the sound field is diffuse.

There are, however, several applications that require attenuation at low frequencies where the acoustic response is dominated by a handful of well-separated modes. One such application is the treatment of so-called boiler tones.[1,2] These are generated by an unstable interaction between the acoustic response to the excitation caused by unsteady flow and the unsteady flow itself and can lead to extremely high levels of tonal noise at even fairly modest flow speeds. Tones can be identified with particular individual acoustic eigenmodes, and the problem is to find a method of increasing the attenuation of a given mode, or a handful of selected modes, rather than to increase the attenuation across a broad range of frequencies.

In these situations, much greater levels of attenuation can be achieved by locating the microperforated plate at some point well inside the cavity. Wherever the plate is located, an accurate analysis of the acoustic characteristics of the plate in the cavity at low frequencies must take account of the modal structure of the cavity. A local analysis at the plate surface is no longer adequate because the "incident" and "reflected" waves are strongly linked through the boundary conditions at the walls of the cavity. The problem becomes essentially a complex eigenvalue problem.

In using microperforated plates in this way to attenuate specific low frequency modes in a cavity, there are two key practical questions that need to be addressed: "What is the optimal location of the plates?" and "What is the flow impedance of the plate that provides the optimal attenuation?" Obviously, such optimization requires a reliable model for predicting the effect of the plates. This paper focuses on the development and validation of such a model. Thus, a theoretical paradigm is presented, a simple analytical model based on this paradigm is derived, and the results of an experimental test of the validity of this model are presented.

II. OUTLINE OF THEORY

A. Boundary conditions and power loss

Consider a volume Ω divided into two subvolumes Ω^+ and Ω^- by a perforated plate lying on a surface Π. At each

[a)]Author to whom correspondence should be addressed. Current affiliation: Aerodynamics & Flight Mechanics Research Group, School of Engineering Sciences, University of Southampton, Southampton, SO17 1BJ, United Kingdom. Electronic mail: bfenech@gmail.com
[b)]Electronic mail: gk@oedan.dk
[c)]Electronic mail: fja@oersted.dtu.dk

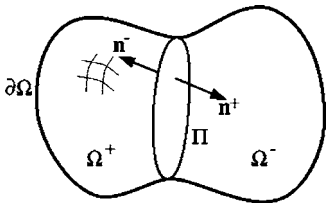


FIG. 1. A volume Ω with boundary $\partial\Omega$ divided by a perforated plate on the surface Π into two subvolumes Ω^+ and Ω^- with outward pointing normals $\mathbf{n}^+$ and $\mathbf{n}^-$.

point $\mathbf{x}_\Pi$ on the plate, the unit vectors $\mathbf{n}^+$ and $\mathbf{n}^-$ are normal to Π and are such that $\mathbf{n}^+$ (respectively, $\mathbf{n}^-$) points out of Ω^+ (respectively, Ω^-); see Fig. 1.

The plate is taken to be sufficiently thin that compressibility effects inside the plate may be neglected. Denoting by $\mathbf{u}^+(\mathbf{x}_\Pi)$ and $\mathbf{u}^-(\mathbf{x}_\Pi)$ the acoustic particle velocities on the two sides of the plate at the point $\mathbf{x}_\Pi \subset \Pi$ continuity of mass gives

$$\mathbf{u}^+(\mathbf{x}_\Pi)\cdot\mathbf{n}^+(\mathbf{x}_\Pi) = \mathbf{u}^-(\mathbf{x}_\Pi)\cdot\mathbf{n}^+(\mathbf{x}_\Pi) = -\mathbf{u}^-(\mathbf{x}_\Pi)\cdot\mathbf{n}^-(\mathbf{x}_\Pi). \tag{1}$$

Viscous losses in the plate support a pressure discontinuity across the surface Π. In the linear limit, this pressure difference is taken to be proportional to the component of the acoustic particle velocity normal to the surface. Denoting by $p^+(\mathbf{x}_\Pi)$ and $p^-(\mathbf{x}_\Pi)$ the pressures on the two sides of the plate at the point $\mathbf{x}_\Pi$, we have

$$p^+(\mathbf{x}_\Pi) - p^-(\mathbf{x}_\Pi) = R(\omega)\mathbf{u}^+(\mathbf{x}_\Pi)\cdot\mathbf{n}^+(\mathbf{x}_\Pi), \tag{2}$$

where $R(\omega)$ is the flow impedance, which is typically a complex valued function of the frequency ω.

The Fourier transform of the linearized Euler momentum equation gives a relation between the acoustic particle velocity and the gradient of the pressure,

$$\nabla p = i\omega\rho_0\mathbf{u} \tag{3}$$

(in which ρ_0 is the density of the medium), and turns the plate conditions given by Eqs. (1) and (2) into conditions involving only the pressure and its gradient,

$$\nabla p^+(\mathbf{x}_\Pi)\cdot\mathbf{n}^+(\mathbf{x}_\Pi) = -\nabla p^-(\mathbf{x}_\Pi)\cdot\mathbf{n}^-(\mathbf{x}_\Pi), \tag{4}$$

$$p^+(\mathbf{x}_\Pi) - p^-(\mathbf{x}_\Pi) = \frac{R(\omega)}{\rho_0 c_0}\frac{1}{ik_0}\nabla p^+(\mathbf{x}_\Pi)\cdot\mathbf{n}^+(\mathbf{x}_\Pi), \tag{5}$$

where c_0 is the speed of sound and $k_0 = \omega/c_0$ is the wave number. For a given pressure field $p(\mathbf{x},\omega)$ inside the volume Ω the time averaged sound power dissipated by the perforated plate is given by

$$P_d(\omega) = \frac{1}{2}\Re\left\{\int_\Pi [p^{+*}(\mathbf{x},\omega)\mathbf{u}^+(\mathbf{x},\omega)\cdot\mathbf{n}^+ + \mathbf{p}^{-*}(\mathbf{x},\omega)\mathbf{u}^-(\mathbf{x},\omega)\cdot\mathbf{n}^-]\mathrm{d}^2\mathbf{x}\right\}, \tag{6}$$

which with Eqs. (1) and (2) can be written

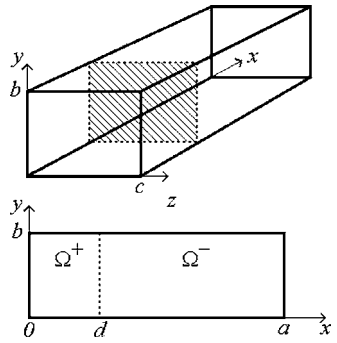


FIG. 2. A rectangular prism with a perforated plate.

$$P_d(\omega) = \frac{1}{2}\Re\{R(\omega)\}\int_\Pi |\mathbf{u}^+\cdot\mathbf{n}^+|^2\mathrm{d}^2\mathbf{x}. \tag{7}$$

Thus the dissipated power is proportional to the real part of the flow impedance and the square of the magnitude of the normal component of the velocity at the plate surface, averaged over the plate. Clearly if the real part of the flow impedance is zero then there is no dissipation. On the other hand, if the real part of the flow impedance is too great then the plate becomes effectively impenetrable, the normal component of the velocity through the plate is zero, and there is no dissipation. Between these two extremes there is an optimal flow impedance, the determination of which requires further knowledge of the pressure field inside the cavity.

All the standard results regarding the acoustics of closed cavities with losses follow.[3] In particular there exists a complete set of orthogonal eigenfunctions $\xi_n(\mathbf{x},\omega)$ with corresponding eigenvalues $\kappa_n(\omega)$ such that

$$(\nabla^2 + \kappa_n^2)\xi_n = 0, \tag{8}$$

subject to the boundary conditions given by Eqs. (4) and (5). Note that due to the frequency dependence of the boundary condition, the eigenvalues and eigenfunctions will in general be functions of the (real) frequency ω. Any given pressure field in Ω can be written as a weighted sum of the eigenfunctions. In particular, the Green's function for the cavity, which satisfies

$$(\nabla^2 + k_0^2)G(\mathbf{x},\mathbf{y};\omega) = -\delta(\mathbf{x}-\mathbf{y}), \tag{9}$$

is given by[3]

$$G(\mathbf{x},\mathbf{y};\omega) = -\sum_{n=1}^{\infty}\frac{\xi_n(\mathbf{x})\xi_n(\mathbf{y})}{V[k_0^2 - \kappa_n^2(\omega)]}, \tag{10}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the source and receiver positions, respectively.

B. A one-dimensional eigenvalue problem

The simplest problem involving a cavity and a perforated plate is that of a rectangular enclosure with the plate located perpendicular to one of the axes; see Fig. 2. This problem can be solved by separation of variables, and since the solutions in the two directions parallel to the plate are trivial the problem essentially reduces to the one-dimensional eigenvalue problem

$$X_l''(x) + \alpha_l^2 X_l(x) = 0, \tag{11}$$

subject to hard walled boundary conditions at the ends of the cavity, viz. $X_l'(0)=X_l'(a)=0$, and the one-dimensional equivalents of Eqs. (4) and (5)

$$X_l'(d^-) = X_l'(d^+), \tag{12}$$

$$X_l(d^-) - X_l(d^+) = \frac{R(\omega)}{\rho_0 c_0}\frac{1}{ik_0}X_l'(d^+). \tag{13}$$

It is convenient to introduce the nondimensional variables

$$\bar{x} = x/a, \quad \delta = d/a, \quad \bar{\alpha} = \alpha a, \quad \varpi = \omega a/c_0, \qquad \bar{R} = R/(\rho_0 c_0). \tag{14}$$

Eigensolutions that satisfy the end boundary conditions can be written

$$X(\bar{x}) = \begin{cases} A\cos[\bar{\alpha}\bar{x}] & \bar{x} < \delta, \\ B\cos[\bar{\alpha}(1-\bar{x})] & \bar{x} > \delta, \end{cases} \tag{15}$$

and the first plate condition, Eq. (12), gives

$$A\sin[\delta\bar{\alpha}] + B\sin[(1-\delta)\bar{\alpha}] = 0. \tag{16}$$

Clearly Eq. (16) is satisfied if both $\sin[\delta\bar{\alpha}]$ and $\sin[(1-\delta)\bar{\alpha}]$ are zero. In this case, it can easily be shown that $\sin[\bar{\alpha}]$ is also zero. These modes correspond to the eigenmodes with no plate present that have zero particle velocity at the location of the plate and thus automatically satisfy both plate conditions. If i and j are integers such that i/j is a fraction in its lowest terms and $i/j=\delta/(1-\delta)$, then $\bar{\alpha}_l=l\pi$ with $l=m(i+j)$ and $m=1,2,3\ldots$, are eigenvalues with eigenfunctions

$$X_l(x) = \sqrt{2}\cos(\bar{\alpha}_l\bar{x}) \tag{17}$$

that automatically satisfy both plate conditions.

The second plate condition, Eq. (13), together with Eq. (16), give the following eigenvalue equation for $\bar{\alpha}$:

$$i\sin[\bar{\alpha}] + \frac{\bar{R}}{\varpi}\bar{\alpha}\sin[\delta\bar{\alpha}]\sin[(1-\delta)\bar{\alpha}] = 0. \tag{18}$$

This equation has been solved for $\bar{\alpha}$ by Newton-Raphson iteration, using the set of eigenvalues corresponding to the case with no plate as starting guesses. Since the ratio $\bar{R}/\varpi$ depends on the frequency a solution strategy was developed that involved tracking each eigenvalue as a function of the ratio from the no-plate case, zero, to its value at the corresponding frequency. The number of intermediate values of the ratio was determined by requiring the change in the eigenvalue corresponding to a change in the ratio not to exceed a specified limit. With tolerances appropriately chosen, this method proved to be extremely robust and remarkably fast. Note that this method does not make any requirement of the ordering of the eigenvalues for a finite flow impedance, as the eigenvalues may (and indeed do) swap places in a list ordered by the size of the real part.

C. The analytical model used for comparison with experiments

The method described in the preceding section furnishes a set of eigenvalues that map continuously to the eigenvalues of the zero flow impedance case. Returning now to dimensional notation, the eigenvalues with the plate can be written $\alpha_l=\bar{\alpha}_l/a$. If l is an integer multiple of $(i+j)$, where i and j are integers with no common factors such that $i/j=\delta/(1-\delta)$, then the zero flow impedance eigenfunction automatically satisfies the plate conditions $\alpha_l=\alpha_{l,0}=l\pi/a$ and the eigenfunction is given by Eq. (17). Otherwise, the eigenfunctions can be written

$$X_l(x) = \begin{cases} -\Lambda_l\sin[(1-\delta)\alpha_l a]\cos[\alpha_l x] & x < d, \\ \Lambda_l\sin[\delta\alpha_l a]\cos[\alpha_l(a-x)] & x > d, \end{cases} \tag{19}$$

where Λ_l is a normalizing factor defined by

$$\int_0^a X_l^2(x)dx = a. \tag{20}$$

Consider now the acoustic cavity shown in Fig. 2 with dimensions a, b, and c. It is trivial to show that eigenfunctions of the Helmholtz equation in such a cavity are given by

$$\xi_{l,m,n}(\mathbf{x}) = X_l(x)Y_m(y)Z_n(z), \tag{21}$$

where $Y_m(y)=\sqrt{2}\cos(\beta_m y)$ with $\beta_m=m\pi/b$ for $m\geq 1$, and $Z_n(z)=\sqrt{2}\cos(\gamma_n z)$ with $\gamma_n=n\pi/c$ for $n\geq 1$. The corresponding eigenvalues are given by

$$\kappa_{lmn}^2 = \alpha_l^2 + \beta_m^2 + \gamma_n^2. \tag{22}$$

For one or more of l, m, or n equal to zero, the corresponding eigenfunctions X_0, Y_0, and Z_0 are unity.

The resulting eigenfunctions may be divided into axial, tangential, and oblique modes in the usual manner. The only modes that are not affected by the presence of the perforated plate are modes which, in the absence of the plate, have zero particle velocity in the x direction at the position of the plate. These include modes with wave motion only in the direction parallel to the plate. An axial mode in the y direction, for example, will not be affected by the presence of the plate.

In the limit of an infinite flow impedance of the plate the eigenmode structure tends to the eigenmode structure of two independent cavities, as one would expect. For each mode in, say, the left subcavity the corresponding mode in the combined cavity is equal to the mode in the left subcavity and identically zero in the right subcavity. Such a mode satisfies the boundary conditions and the governing equations everywhere and has a real-valued eigenfrequency equal to the eigenfrequency of the mode in the left subcavity. With a source in, say, the left subcavity the Green's function is zero if the receiver position is in the right subcavity because all modes that are nonzero in the left subcavity are identically zero in the right subcavity.

Using the eigenfunctions given by Eq. (21) in Eq. (10), the analytical model has been compared with numerical (finite element) calculations performed using the commercial finite element code ACTRAN, and the results were found to be virtually indistinguishable. This showed that the analytical solution provides a solution of the Helmholtz equation sub-

ject to the appropriate boundary and plate conditions [Eqs. (4) and (5)]. In Sec. III B it is investigated through a series of experiments whether these boundary and plate conditions are a reasonable model for the behavior of real perforated plates in a real cavity.

III. EXPERIMENTAL VERIFICATION

A. The flow impedance of a microperforated plate

Microperforated plates are available in various formats, ranging from thin transparent films that can be mounted in front of windows to much more robust metal plates for use at high temperatures or in other harsh environments. Many manufacturers offer microperforated plates as a complete acoustic absorber package (i.e., a perforated plate mounted at a certain distance from some rigid backing). A few others offer individual plates, giving much more flexibility to the noise control engineer.

A plate with the required dimensions was supplied by the Swedish manufacturer Sontech.[4] Microperforated plates under the trade name Acustimet form part of the airborne sound absorption materials group and are available in mild and stainless steel, and aluminum. The perforations are produced by punching rather then drilling; this produces a sharp-edged hole geometry that is extremely difficult to define geometrically. The plate supplied was made of aluminum, and came with the smallest perforation size that the company produces.

One of the most important parameters of the analytical model described in Sec. II is the flow impedance of the microperforated plate. Unfortunately, the manufacturer does not provide such data. Manufacturers rarely provide this quantity, but prefer to use the absorption coefficient. However, whereas one can calculate the absorption coefficient from the flow impedance if the configuration is known, one cannot calculate the flow impedance from the absorption coefficient. The starting point in the investigation was thus to determine the flow impedance of the given plate.

One possibility might be to use a set of relationships derived by Maa.[5] Maa's equations are the result of an analysis where the microperforated plate is treated as a lattice of short narrow identical tubes with a certain diameter and a length equal to the thickness of the plate. However, the concept of diameter and thickness cannot be attributed to the punched perforations in the Acustimet plates, since each hole has a unique geometry with sharp edges that protrude out of the surface of the plate. Accordingly, it was decided to determine the flow impedance experimentally.

A widely accepted method of measuring the flow impedance of a perforated plate has been devised by Ingard and Dear.[6] This simple method is known to give reliable results, but it suffers from one major drawback: It gives the flow impedance only at certain discrete frequencies. Ren and Jacobsen have further developed this method to give the flow impedance as a continuous function of the frequency.[7] A sample of the material is placed in an impedance tube driven by a loudspeaker at one end and terminated near anechoically at the other end. Two microphones are mounted on either side of the sample, with their diaphragm flush with the inside surface of the tube. The complex flow impedance is calculated from the measured transfer function between the two microphones. Reflections from the near-anechoic termination, and possible amplitude and phase mismatch of the two microphones are taken into account.[7]

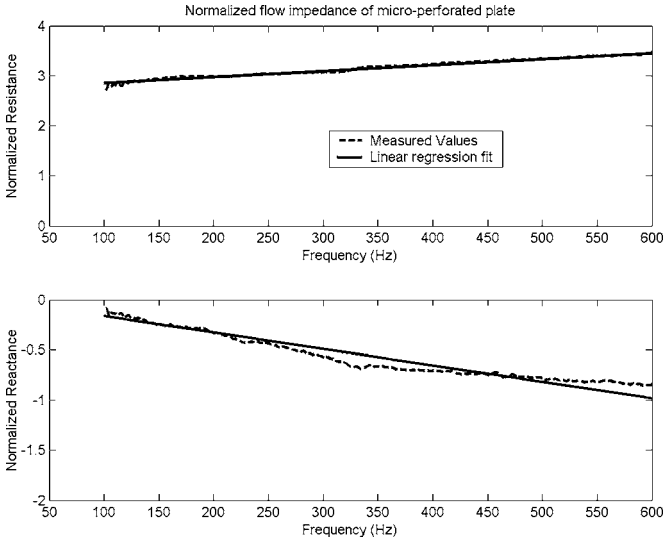


FIG. 3. Real and imaginary part of measured flow impedance of the Acustimet sample, and linear fits. The data are normalized with $\rho_0 c_0$.

A number of different samples were tested using this method, including conventional perforated plates with circular holes of submillimeter size. In general, the results agreed qualitatively and quantitatively well with Maa's theory, according to which the real part of the flow impedance depends weakly on the frequency, whereas the imaginary part is essentially masslike.[5] The method is rather sensitive to how the sample is mounted inside the tube, and leaks and structural vibrations in the sample can be difficult to eliminate completely. However, errors due to these effects dominate only at specific frequencies and a linear regression fit can be used to retrieve the general trend. Figure 3 shows the flow impedance (real and imaginary parts) and corresponding linear approximation of the Acustimet sample used for the experimental validation of the theoretical model described in Sec. II. The results are normalized by $\rho_0 c_0$. The holes in this 1-mm-thick sample can be considered to be slits approximately 3.5-mm long and 0.2-mm wide. The fractional open area was estimated at 2–3 %; this explains the relatively high impedance values.

B. Green's function in a cavity with a microperforated plate

A set of experiments were carried out in a rectangular cavity with dimensions of 2, 1.2, and 0.2 m; see Fig. 4. These dimensions correspond to the x, y, and z coordinates as defined in Fig. 2. The dimensions of the cavity were chosen such that below 850 Hz, the cavity acts as a two-dimensional space; and the frequencies at which the first modes occur are reasonably spaced apart. The cavity was constructed of 22-mm fiberboard panels screwed and glued together, except for the front vertical panel, which was removable so to allow easy access to the inside of the cavity.

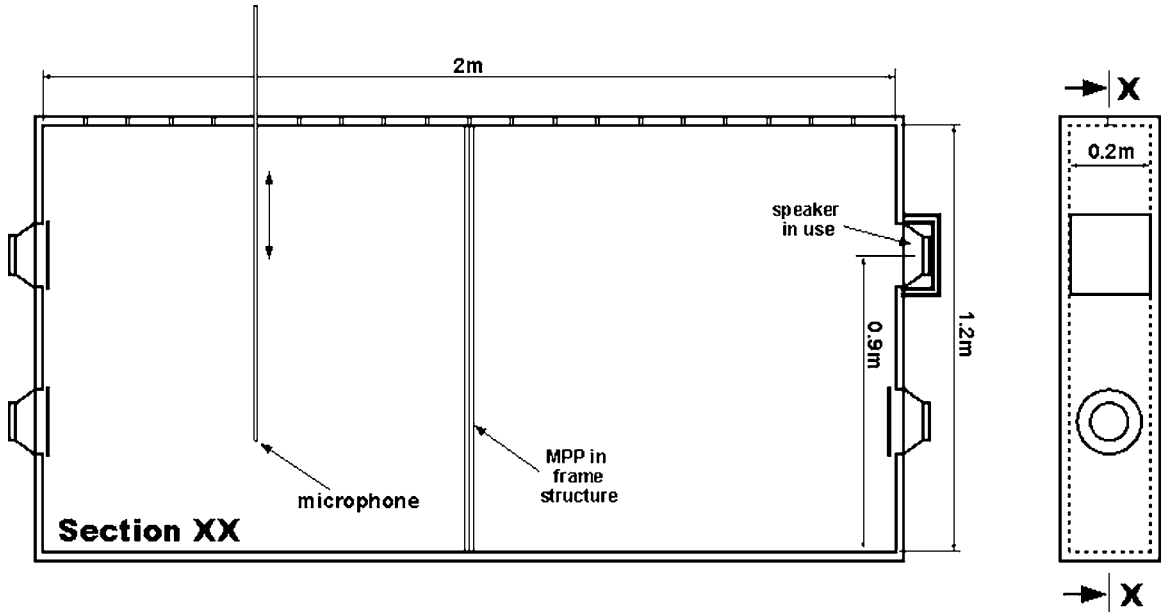


FIG. 4. The flat rectangular cavity used in the experiments. The volume velocity of the loudspeaker that drives the cavity is deduced from the sound pressure in a small box enclosing the back of the loudspeaker. Three unused loudspeakers from an earlier experiment are covered by aluminum plates.

Acoustic modes inside the cavity were excited using white noise generated by a loudspeaker with a diameter of 9 cm mounted on one of the vertical side panels. Pressure measurements inside the cavity were taken through holes in the top panel using an electret microphone with a diameter of 10 mm mounted on the tip of a 1.2-m-long rod. This setup made it possible to measure at a grid of points inside the cavity, so that given a sufficient amount of data points, the mode shape at a particular frequency could be reconstructed. Data was acquired using Brüel and Kjær's "Pulse" front-end and software, which was also used to carry out the fast Fourier transform (FFT) analysis. To get the Green's function the sound pressure data was normalized with the density of air and the volume velocity of the loudspeaker Q, estimated from the sound pressure p_c measured in a small box enclosing the back of the loudspeaker

$$p_c = -Q\frac{\rho_0 c_0^2}{i\omega V_c}, \tag{23}$$

where V_c is the volume of the small cavity, 1100 cm^3; in other words, the Green's function was estimated using the expression

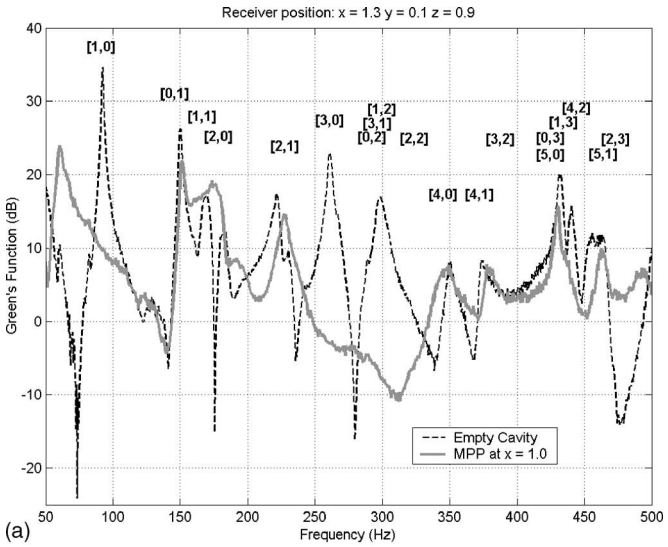
$$G(\mathbf{x},\mathbf{y};\omega) = \frac{p(\mathbf{x})}{-i\omega\rho_0 Q(\mathbf{y})} = -\frac{p(\mathbf{x})}{p_c}\left(\frac{c_0}{\omega}\right)^2\frac{1}{V_c}. \tag{24}$$

When a microperforated plate is mounted inside a cavity driven by a sound source, vibrations of the plate are likely to occur. Such vibrations obviously affect the relative velocity of the air oscillating through the perforations, resulting in an unpredictable damping behavior. To avoid such problems the microperforated plate of dimensions 1.2×0.2 m was clamped between two wooden frames. This arrangement raised the first fundamental structural frequency of the plate from about 60 Hz to above 500 Hz. Although this arrangement may have introduced some level of distortion in the sound field very close to the plate, the necessity of a stationary plate was given higher priority. In what follows the measured flow impedance of the plate has been corrected for the area covered by the supporting frame.

The sound pressure was measured at various locations inside the cavity. Three sets of measurements were taken: a reference measurement without the microperforated plate, and two with the plate mounted at two different positions, in the middle (at $d=1.0$ m, cf. Fig. 2) and close to one side (at $d=0.25$ m). With the plate in the middle it was expected to be easy to distinguish between affected and undamped modes. Modes with $l=1,3,5,\ldots$, should be significantly damped (maximum velocity through plate), while modes with l even should be practically undamped (velocity node at plate). For the third case, with the plate mounted close to one side, the constants i and j as defined in Sec. II B compute to 1 and 7, which implies that the plate has no effect on modes with $l=8,16,\ldots,$. These modes are not present below 500 Hz. It follows that all modes occurring with this setup should be damped with the exception of modes with $l=0$.

For more accurate comparisons between predictions and measurements, Eq. (10) was modified to take into account the finite size of the source (approximated by a square piston with the same area), and the inherent damping of the cavity, which turned out not to be negligible. The latter quantity was estimated from the reverberation time measured in one-third octave bands without the microperforated plate in the cavity.

The reverberation time was found to be approximately 0.5 s in most of the frequency range of concern—which is about one order of magnitude shorter than the reverberation time predicted from viscothermal losses assuming perfectly rigid walls.[8] (One-third octave bands were chosen because one cannot measure short reverberation times with fine frequency resolution.[9]) The corrected expression is



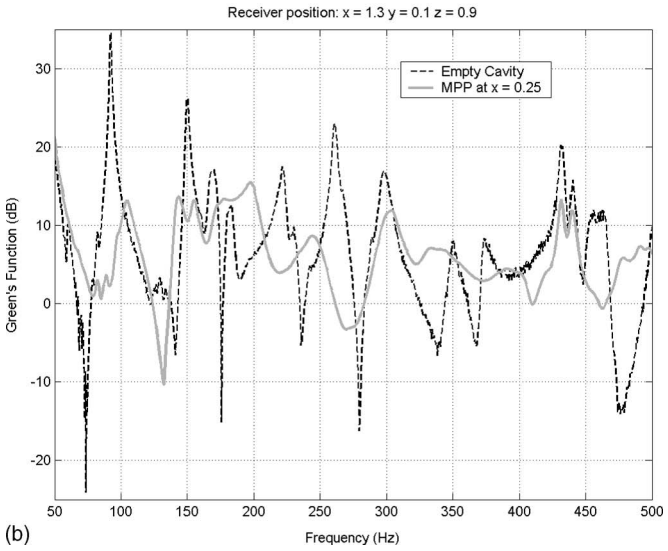


FIG. 5. Measured effect of a microperforated plate mounted (a) in the middle of the cavity, and (b) near a wall.

$$G(\mathbf{x},\mathbf{y};\omega) \approx -\sum_{n=1} \frac{1}{S_{\text{piston}}} \int_{\Xi} \xi_n(\mathbf{y}) \mathrm{d}^2\mathbf{y} \times \frac{\xi_n(\mathbf{x})}{V\left[k_0^2 - \kappa_n^2(\omega) + ik_0 \frac{1}{\tau_n c_0}\right]}, \tag{25}$$

where Ξ denotes the surface of the source with surface area S_{piston}, and τ_n is a modal time constant that takes account of the inherent damping of the enclosure.

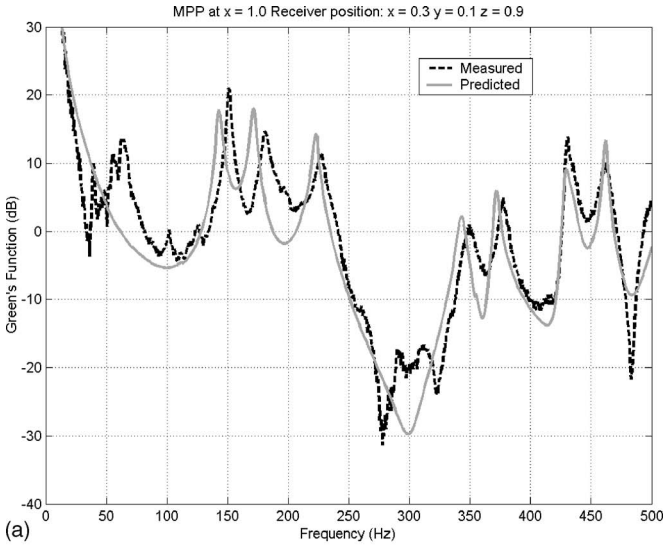
A comparison between the measured data in the three different scenarios is shown in Fig. 5. To facilitate mode identification, analytically calculated eigenfrequencies for an empty rigid-walled cavity with the same dimensions are given in Table I and indicated in Fig. 5(a). The particular receiver position used for these results was chosen because it picks up most of the acoustic modes in the given frequency range, but similar results have been obtained at a number of positions. One can immediately notice that the microperforated plate does provide significant damping for particular

TABLE I. Eigenfrequencies of the empty cavity.

l, m	0 (Hz)	1 (Hz)	2 (Hz)
0		143	286
1	86	167	298
2	172	223	333
3	257	294	385
4	343	372	446
5	429	452	

modes. In Fig. 5(a) modes $(1,0)$, $(1,2)$, $(1,3)$, $(3,0)$, and $(3,1)$ are no longer apparent, resulting in attenuations of as much as 25 dB, whereas modes with $l=0,2,4$ as expected are practically unchanged. With the plate close to the side, as shown in Fig. 5(b), the entire Green's function is significantly damped. This includes mode $(0,1)$ (with wave motion parallel to the plate), in disagreement with the theory.

Figure 6 compares experimental data with results predicted using the analytical model described in Sec. II. The



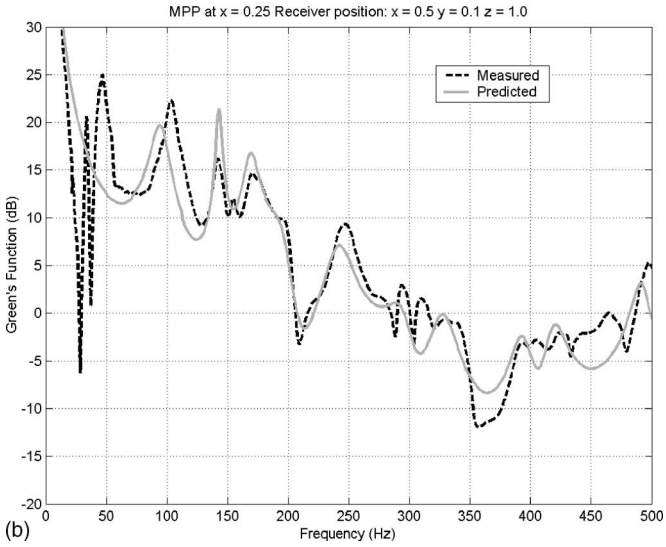


FIG. 6. Measured and predicted Green's function with the microperforated plate mounted (a) in the middle of the cavity, and (b) near a wall.

number of modes summed in the x, y, and z directions are 46, 20, and 3, respectively. Figure 6(a) shows the comparison for the plate in the middle, and Fig. 6(b) shows the case with the plate close to one of the sides of the enclosure. In general, a reasonably good match can be observed in both plots, and the same trends are observed at other locations inside the cavity (not shown).

C. Discussion

The experimental arrangement is not perfect. The most conspicuous difference between predictions and measurements is the small frequency shift between the curves. After extensive testing, this phenomenon has been attributed to vibroacoustic interactions between the sound field inside the cavity and the cavity walls. Such interactions are usually of a very complex nature, and any attempts to take them into consideration in the analytical model would give rise to unnecessary complications. This phenomenon also explains the peaks picked up by the measurements at around 50 Hz, a frequency that is significantly below the first acoustic mode of the cavity (cf. Table I). There are also discrepancies between the magnitudes of the peaks of some modes. In most cases these discrepancies can probably be attributed to the fact that the inherent damping of the cavity was measured in one-third octave bands. Thus if two modes were present in the same one-third octave band, an average decay rate was assumed. Moreover, the vibroacoustic losses of the cavity may have been affected by the introduction of the plate. Yet another source of uncertainty is that the wooden frame supporting the plate changes the geometry of the two subcavities. However, the geometry of the holes of the Acustimet plate is probably the most serious problem. Given their three-dimensional shape, which is almost like that of a grater, it seems reasonable to expect the Acustimet plates to provide damping even to modes with wave motion only parallel to the plate. In all probability this explains the damping of mode $(0,1)$, occurring approximately at 150 Hz. According to the model the $(0,1)$ mode of the empty cavity (with wave motion tangential to the plate) is simply not affected by the plate since, in steady state, the $(0,1)$ modes in the two subcavities match each other exactly so that there is no pressure drop across the plate and therefore no losses. In other word, in such cases the model is conservative and underestimates the damping, as demonstrated in Fig. 6(b).

There seems to be a certain similarity between the behavior of the Acustimet plate and bulk-reacting absorbers; in both cases a normal incidence measurement does not provide sufficient information for predicting the behavior for incidence at arbitrary angles. However, there is also a significant difference between these two cases. In a bulk-reacting absorber there is wave propagation in the material.[10] The effect of a bulk-reacting lining can be determined from equations expressing the boundary condition of the surface of the material (continuity in the pressure and the normal component of the particle velocity) if the characteristic impedance and propagation constant of the material are known.[10,11] By contrast, a strict analysis of the effect of the Acustimet plate seems to involve coupling the regions on either side of the plate through continuity in the pressure and a nonperpendicular component of the particle velocity (or introducing a tensorial flow impedance), and that would be exceedingly complicated.

Nevertheless, in spite of all these shortcomings, the experiments certainly confirm the practical utility of the theoretical approach presented in Sec. II.

IV. CONCLUSIONS

An analytical method for predicting the attenuation provided by a microperforated plate in a cavity has been presented. The microperforated plate is described in terms of its complex flow impedance. The method involves finding complex eigenvalues and eigenfunctions, and is based on solving an eigenvalue equation iteratively using the set of eigenvalues of the cavity with no perforated plate as starting guesses. Each of the resulting eigenvalues corresponds uniquely to an eigenvalue of the cavity without the plate, although the ordering may change.

The model has been validated experimentally by comparing predicted Green's functions with measurements with a microperforated plate mounted in a flat box and using experimentally determined values of the flow impedance of the plate. Very good agreement was obtained, except for one mode with wave motion only parallel to the perforated plate. Because of the peculiar geometry of the punched holes such modes are actually attenuated by the plate, but since the model takes account only of the normal component of the acoustic particle velocity this attenuation is underestimated. However, in general the substantial attenuation provided by the microperforated plate was predicted very well. The model makes it possible to optimize the position of the plate.

These findings show that microperforated plates can provide useful and predictable attenuation of boiler tones in heat-exchange cavities. Further research is underway to determine, in general, the optimal location and flow impedance of a perforated plate for a given cavity geometry, gas temperature, and gas pressure.

ACKNOWLEDGMENTS

The authors would like to thank Ralf Corin at Sontech NoiseControl for providing the microperforated plates used in this project, and also Aage Sonesson and Jørgen Rasmussen, Acoustic Technology, Ørsted-DTU, for their help with the practical aspects of the experimental work.

[1] R. D. Blevins, *Flow-Induced Vibrations* (Wiley, New York, 1990).
[2] G. M. Keith, "Flow-acoustic interactions in heat-exchange cavities. A summary of results and conclusions from recent analyses and experiences." Technical Report, Ødegaard & Danneskiold-Samsøe A/S, 2004.
[3] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[4] Sontech Noise Control, Airborne sound absorption: Acustimet. http://sontech.se
[5] D.-Y. Maa, "Potential of micro-perforated panel absorber," J. Acoust. Soc. Am. **104**, 2861–2866 (1998).
[6] K. U. Ingard and T. A. Dear, "Measurement of acoustic flow resistance," J. Acoust. Soc. Am. **103**, 567–572 (1985).
[7] M. Ren and F. Jacobsen, "A method of measuring the dynamic flow resistance and reactance of porous materials," Appl. Acoust. **39**, 265–276 (1993).

[8]L. Cremer and H. A. Müller, *Principles and Applications of Room Acoustics*, Vol. **2** (Applied Science Publishers, London, 1978). See Sec. IV.7.8 "Unavoidable sound absorption at a rigid wall."
[9]F. Jacobsen, "A note on acoustic decay measurements," J. Sound Vib. **115**, 163–170 (1987).
[10]F. P. Mechel and I. L. Vér, "Sound absorbing materials and sound absorbers," in *Noise and Vibration Engineering: Principles and Applications*, edited by L. L. Beranek and I. L. Vér (Wiley, New York, 1992), Chap. 8.
[11]R. A. Scott, "The propagation of sound between walls of porous materials," Proc. Phys. Soc. **58**, 358–368 (1946).

Rayleigh–Ritz approach for predicting the acoustic performance of lined rectangular plenum chambers

Hoi-Jeon Kim and Jeong-Guon Ih[a)]
Center for Noise and Vibration Control, Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Science Town, Taejon 305-701, Korea

(Received 20 September 2005; revised 18 July 2006; accepted 19 July 2006)

The purpose of this study was to predict the acoustic performance of a fully lined rectangular plenum chamber having inlet and outlet ports at arbitrary locations. Because no exact analytic solution exists for this reactive-resistive silencer configuration, numerical methods are only available for a three-dimensional analysis. The lined plenum chamber was modeled as a piston-driven rectangular tube without mean flow and the acoustic pressure in the lined chamber was obtained by superposing the acoustic pressures due to each harmonically fluctuating piston. Air pore and skeleton material of the porous liner was reduced to an equivalent medium; thus, its acoustic characteristic was given by bulk-reacting liner properties. A single weak variational statement, which satisfies the conditions of the oscillating piston and all necessary boundary conditions, was developed. The Rayleigh-Ritz method was employed as the numerical scheme for the derived variational statement. Using a transfer matrix and measured material properties, all possible types of lined plenum chambers were tested. Computed results were compared with the predicted transmission loss by the locally reacting liner model and experimental results. Transmission loss predicted by the Rayleigh–Ritz method using the bulk-reacting liner model agreed well with the measured one. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2336748]

PACS number(s): 43.20.Mv, 43.20.Bi [LLT] Pages: 1859–1870

I. INTRODUCTION

Lined rectangular plenum chambers which can be regarded as reactive-resistive silencers[1] are commonly used for attenuating mid- to high-frequency noise emitted from HVAC (heating, ventilating, and air conditioning) and intake/exhaust systems of commercial vehicles and heavy machinery. Rectangular plenums can be also used when the overall ductwork is to be made compact and simple. In comparison with typical dissipative silencers, such as the splitter silencer, they would have a relatively large amount of noise attenuation at low frequencies, which is benefited from the reactive action of the chamber. The analysis presented in this paper embodied a low frequency approach for a rectangular plenum by assuming the plane wave propagation at inlet/outlet ports.

The lined rectangular plenum chambers can have a disadvantageous feature in having a certain amount of pressure loss due to abrupt area expansion and contraction or flow reversal; however, this drawback might be insignificant in many actual cases. Also, for the same volume and major length of the chamber, a rectangular chamber includes many flat larger surfaces than a circular one, so a rectangular chamber is disadvantageous from the viewpoint of control of sound radiation and transmission. However, it can maximally and compactly utilize a given space or volume for silencers; it can also be fitted into a space forming a flush face with adjacent structural parts, which is beneficial both aerodynamically and aesthetically. Some commercial trucks employed lined rectangular silencers, perhaps for these reasons.

An approximate expression for the attenuation of the lined rectangular plenum chamber[2] exists, using the absorption coefficient of the lining material and geometric factors, such as the distance and angle between inlet and outlet ports. Due to the simplicity, such a formula can be used conveniently and quickly *in situ*. However, this expression is not suitable for predicting the spectral characteristics of the whole silencer system, in which several different types of dissipative silencers are combined. Moreover, the effects of three-dimensional wave action due to excitation from inlet and outlet ports at arbitrary locations and the subsequent high-order modes cannot be treated properly.

Cummings[3] obtained the transmission loss (TL) of a rectangular plenum chamber, lined all around with locally reacting material, using two theories. One was for the low frequency range by virtue of the mode-matching technique and the other used the ray acoustics model, which is only valid at high frequencies. The explicit form of the 2×2 transfer matrix was not developed in his work, thus, the incorporation of this model into other muffler elements for constructing a total system was somewhat inconvenient.

In considering the three-dimensional wave effects in the muffler element, several methods can be employed; the mode-matching technique,[4–8] the point source model,[9] or the piston-driven method.[10] The mode-matching technique has been used to analyze electrical fields,[4] and also to predict the sound transmission of expansion chamber mufflers.[6–8] However, this technique sometimes brought difficulties in actual implementation due to the mathematical complexity in matching the eigenfunctions at the interface with abrupt area

[a)]Author to whom correspondence should be addressed; electronic mail: j.g.Ih@kaist.ac.kr

change and the nonorthogonality of modal eigenfunctions, in particular, in the presence of mean flow.[6] On the other hand, the point source method[9] is easily applicable to complex-shaped mufflers, but, due to too many simplified assumptions, it may yield erroneous results at some frequencies. Kim *et al.*[11] derived a 2×2 transfer matrix of the rectangular plenum chamber, lined with locally reacting material, using the piston-driven chamber model. Excitation by a plane wave at the inlet and outlet was assumed for deriving the transfer matrix, so application of the method was limited to small inlet and outlet ports, or low frequencies, or some high order modes having nodal lines passing through the inlet or outlet port. Comparison of measured and predicted transmission loss curves showed good agreement at low frequencies, but the predicted frequencies of peaks and troughs, and their levels, differed from the measured data as the frequency increases. Large discrepancies at high frequencies were mainly caused by the locally reacting liner, disregarding the actual fact that the sound can propagate into the sound-absorbing material. This can be inferred from the fact that the predicted transmission loss becomes very close to the measured data when there is little chance of sound passing through the liner. Kirby and Lawrie[12] suggested a method which combines two-dimensional finite element method (FEM) and point collocation technique. This technique would be very useful in many potential application areas from the viewpoint of CPU expenditure. Besides it was reported that any problem was not caused even when the orthogonality condition does not exist, e.g., in the presence of mean flow in the airway, or the lining material is positioned at the corner.

Scott[13] considered the liner as a bulk-reacting material and adopted this concept in formulating the expression for propagation constants in two-dimensional lined rectangular and circular ducts. Ducts with infinite length were only considered in this study. Cummings and Chang[6] calculated the transmission loss for a finite length lined circular chamber by employing the mode-matching technique. Good agreement between measured and predicted transmission losses were observed, even though nonorthogonal eigenfunctions which exist in the presence of mean flow were used as weighting functions. The transfer matrix of a finite length circular silencer with the bulk-reacting liner model was also suggested by Peat.[14] However, the available frequency range was rather restricted due to the fundamental mode approach. Glav[8] considered the three-dimensional effect in the derivation of the transfer matrix for a dissipative silencer of arbitrary cross section by using the mode-matching method and the bulk-reacting liner model. The point-matching technique was employed to obtain propagation constants and eigenfunctions, which were expressed as the sum of Bessel functions. Panigrahi and Munjal[15] summarized and compared three approaches for analyzing lined circular ducts with the bulk-reacting liner model. A complete three-dimensional FEM, considering mean flow and complex duct shape, was suggested by Peat and Rathi;[16] but this full FEM solution demanded a large expenditure of CPU.

At the moment, there is no exact analytical solution when the lining material is positioned at the corner of rectangular ducts, for example, in dealing with four-sided lined ducts[17] or bar-type dissipative silencers.[18] At present, the acoustic performance of the aforementioned mufflers can only be assessed through numerical methods. For the numerical methods, the functional, which takes account of the governing equations and all boundary conditions, should be derived first. Astley and Cummings[17] derived a two-dimensional functional, which is adequate for four-sided lined ducts, using the finite element method. Farvacque[18] derived a similar functional for the bar-type silencer using the Rayleigh–Ritz method.

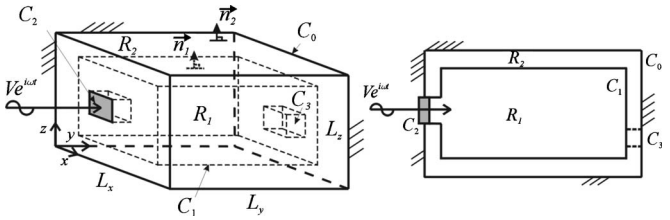


FIG. 1. Rectangular lined plenum chamber excited by a harmonically oscillating rigid piston positioned at the inlet port.

In this paper, a modified three-dimensional functional was proposed to derive the transfer matrices of a lined rectangular plenum chamber in the absence of mean flow. Transmission loss for all possible types of lined rectangular plenum chamber, under the assumption of the bulk-reacting lining material, could be calculated from the acquired transfer matrices. The Rayleigh–Ritz method was employed as a numerical scheme, thus, the calculation effort using the full finite element method could be reduced substantially.

However, it should be mentioned that the applicability of the present method will be restricted due to the oscillating rigid piston model at inlet and outlet ports and the inherent limitations of the Rayleigh–Ritz method. For example, the transmission loss at some frequencies may not converge within an acceptable error bound with a given limited CPU expenditure. Also, the present method requires a number of acoustic modes for calculating the transmission loss at a frequency far higher than the first cutoff frequency or for a large chamber having a dimension larger than several wavelengths, which may lead to the convergence problem. Subsequently, the maximum size of chamber and high frequency limit can only be determined after carrying out the convergence check with the Rayleigh–Ritz method and this restricts the use of the present method to a relatively small chamber which is involved with only several wavelengths for a dimension.

II. THEORETICAL MODEL

A. Governing equations and boundary conditions

For acoustic analysis on the lined rectangular plenum chamber, the rigid piston-driven model, which was successfully applied to the reactive rectangular plenum,[10] was adopted. When utilizing this model, the sound field caused by the rigid piston, which is fluctuating back and forth at the inlet and outlet ports, should be calculated for each port motion. The coordinates and geometry of the chamber are depicted in Fig. 1. The inside of a rectangular plenum chamber is lined with a sound-absorbing material. The regions occu-

pied by airway and lining materials are denoted by R_1 and R_2, respectively. The gas medium within the airway was assumed to be homogeneous and there was no mean flow inside of the chamber. The interface between airway and lining material is denoted as C_1. The rigid piston, representing the inlet port and denoted as C_2, was assumed to oscillate in a harmonic manner, i.e., varying with $e^{j\omega t}$. The outer surface plenum jacket and outlet port plane were assumed to be acoustically hard walls and are denoted by C_0 and C_3, respectively. Under these assumptions, the wave equation in the airway can be expressed as

$$\nabla^2 p_1 + k_0^2 p_1 = 0, \tag{1}$$

where p_1 is the acoustic pressure in the airway, k_0 the free-field wave number, and the three-dimensional Laplacian operator in a Cartesian coordinate system. The wave equation in the lining material is given by

$$\nabla^2 p_2 + k_a^2 p_2 = 0, \tag{2}$$

where p_2 is the acoustic pressure in the lining material and k_a the complex wave number incorporating sound attenuation in the lining material. On the outer surface C_0 and the outlet port plane C_3, the normal component of the particle displacement should be zero:

$$\nabla p_2 \cdot \mathbf{n}_2 = 0 \text{ on } C_0, \tag{3}$$

$$\nabla p_1 \cdot \mathbf{n}_1 = 0 \text{ on } C_3. \tag{4}$$

Here, $\mathbf{n}_1$ and $\mathbf{n}_2$ denote the normal outward unit vector on regions R_1 and R_2, respectively, as shown in Fig. 1. The continuity of the acoustic pressure and normal component of particle displacement should be satisfied at the interface between the airway and lining material (C_1) as follows:

$$p_1 = p_2, \tag{5}$$

$$\nabla p_1 \cdot \mathbf{n}_1 - (\rho_0/\rho_a) \nabla p_2 \cdot \mathbf{n}_1 = 0. \tag{6}$$

Here, ρ_0 is the density of air and ρ_a is the effective complex density of the lining material. The concept of effective complex density came from the equivalent medium model,[19] which considers only longitudinal wave propagation through the air within the pores of porous materials. The acoustic behavior of porous materials can be totally described by propagation constants and characteristic impedance. The effective complex density and bulk modulus can be defined by the aforementioned values as

$$\rho_{\text{effective}} = \frac{Zk}{\omega}, \tag{7}$$

$$K_{\text{effective}} = \frac{Z\omega}{k}, \tag{8}$$

where Z is the characteristic impedance, ω the angular frequency, and k the propagation constant. The normal component of acoustic velocity on the surface of harmonically oscillating, rigid piston is given by

$$\rho_0 j\omega V = -\nabla p_1 \mathbf{n}_1 \quad \text{on } C_2, \tag{9}$$

where V is the velocity of the rigid piston at the inlet port.

B. Bulk properties of the lining material

As previously mentioned, the lining material can be regarded as an equivalent medium.[19] Thus, the characteristic impedance and propagation constants are used for characterizing sound propagation through the sound-absorbing material completely. In the case of a fibrous material, the empirical formulas of Delany and Bazley[20] or those of Mechel[21] can be employed to obtain the aforementioned parameters using steady-flow resistivity. In general, it is known that these empirical formulas give quite acceptable values for a fibrous material. For polyurethane foam-type sound-absorbing material, i.e., in cellular foam structures, the impedance tube method, e.g., using multiple acoustic loads,[22] has been often used for the measurement of acoustic properties.

C. Derivation of the functional

An exact analytic solution of these coupled equations does not exist because the wave number in the lining material, which is positioned at the corner of the four-sided lined duct, cannot be defined.[17,18] Kakoty and Roy[23] derived an approximate analytic solution of the four-sided lined duct by neglecting the lining material positioned at the corners. At present, it seems that the numerical method is the only way to obtain the correct results.

1. Isotropic lining material

Astley and Cummings[17] suggested a single two-dimensional functional, combining acoustic wave equations and all required boundary conditions in the rectangular lined duct, to obtain the propagation constants of the lined duct. In this paper, this functional is modified to consider the three-dimensional sound field generated by the oscillating rigid pistons at the inlet and outlet ports in rectangular lined plenum chambers. The modified functional, satisfying all the wave equations and boundary conditions of a chamber lined with isotropic material, can be derived as follows:

$$\begin{aligned}\chi(p_1,p_2) = &\int_{R_1} [\nabla p_1 \cdot \nabla p_1 - k_0^2 p_1^2] dR_1 \\ &+ \left(\frac{\rho_0}{\rho_a}\right) \int_{R_2} [\nabla p_2 \cdot \nabla p_2 - k_a^2 p_2^2] dR_2 \\ &+ \int_{C_2} [2\rho_0 j\omega V p_1] dC_2. \end{aligned} \tag{10}$$

The first and second term of Eq. (10) is the three-dimensional extension of previous work[17] and the third term is added for the inclusion of the oscillating rigid piston model at the inlet or outlet ports. Let p_1 and p_2 be continuous functions within regions R_1 and R_2, respectively. Small variations $\varepsilon\eta_1$ and $\varepsilon\eta_2$ in p_1 and p_2, respectively, results in the following variational form of $\chi(p_1,p_2)$:

$$\delta\chi(p_1,p_2) = \chi(p_1+\varepsilon\eta_1, p_2+\varepsilon n_2) - \chi(p_1,p_2)$$
$$= 2\varepsilon\int_{R_2}[\nabla p_1\cdot\nabla\eta_1 - k_0^2 p_1\eta_1]dR_1$$
$$+ 2\varepsilon\left(\frac{\rho_0}{\rho_a}\right)\int_{R_2}[\nabla p_2\cdot\nabla\eta_2 - k_a^2 p_2\eta_2]dR_2$$
$$+ \int_{C_2}[2\rho_0 j\omega V\varepsilon\eta_1]dC_2. \quad (11)$$

Applying the divergence theorem to the first and second terms of Eq. (11), and utilizing the fact that η_1 and η_2 are identical on C_1, one obtains

$$\delta\chi(p_1,p_2) = -2\varepsilon\int_{R_2}[\nabla^2 p_1 + k_0^2 p_1]\eta_1 dR_1$$
$$- 2\varepsilon(\rho_0/\rho_a)\int_{R_2}[\nabla^2 p_2 + k_a^2 p_2]\eta_2 dR_2$$
$$+ 2\varepsilon(\rho_0/\rho_a)\int_{C_0}[\nabla p_2\cdot\mathbf{n}_2]\eta_2 dC_0$$
$$+ 2\varepsilon\int_{C_3}[\nabla p_1\cdot\mathbf{n}_1]\eta_1 dC_3 + 2\varepsilon\int_{C_1}[\nabla p_1\cdot\mathbf{n}_1$$
$$- (\rho_0/\rho_a)\nabla p_2\cdot\mathbf{n}_1]\eta_1 dC_1 + 2\varepsilon\int_{C_2}[\nabla p_1\cdot\mathbf{n}_1$$
$$+ \rho_0 j\omega V]\eta_1 dC_2 + O(\varepsilon^2). \quad (12)$$

Because Eq. (12) becomes a value less than $O(\varepsilon^2)$ for any values of η_1 and η_2, the equations and boundary conditions within all brackets should vanish and these are exactly equal to Eqs. (1)–(4), (6), and (9), respectively. The only constraint on the variation is that p_1 and p_2 should be identical on the interface C_1.

2. Orthotropic lining material

For many fibrous lining materials, such as glass fiber, their acoustic properties depend on the orientation of the fibers. In general, most fibers lie in planes parallel to the surface, so that such a fibrous liner can be regarded as an orthotropic material.[24] During the manufacturing process of the lined plenum, such an orthotropic sound-absorbing material can be placed in an arbitrary direction inside a chamber for fabrication convenience and cost. The allocation pattern depicted in Fig. 2 was chosen as an example to describe the derivation of the functional on this rectangular lined chamber. The region denoted by 1 is the airway and the other regions (from 2 to 7) are packed with the lining material. The normal direction of the lining material is set, pointing toward the center of the chamber.

The x-directional component of the momentum equation in the anisotropic material[25] in the absence of flow can be written by

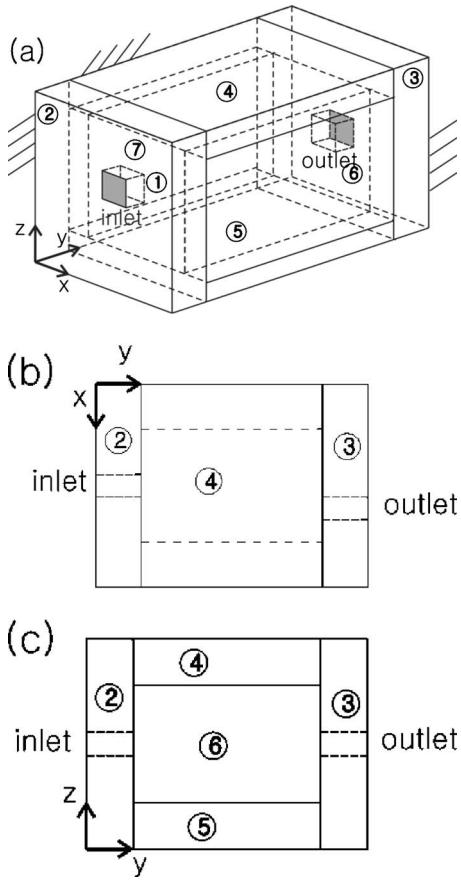


FIG. 2. Rectangular plenum chamber lined with an orthotropic material. (a) Oblique view, (b) x-y view, (c) y-z view.

$$j\omega\rho_x u_x = -\frac{\partial p}{\partial x}, \quad (13)$$

where ρ_x is the effective density of the anisotropic material in the x direction and u_x is the acoustic particle velocity in the x direction. The mass conservation equation in the anisotropic material can be expressed as[24]

$$K_x\frac{\partial u_x}{\partial x} + K_y\frac{\partial u_y}{dy} + K_z\frac{\partial u_z}{\partial z} + j\omega p = 0, \quad (14)$$

where K_x, K_y, and K_z denote the bulk modulus of the anisotropic material of each axis, respectively. The displacement continuity condition of the chamber, as shown in Fig. 2, cannot be fulfilled within a single weak formulation of the functional because the bulk moduli of each direction differ from each other. In the rigid frame model for sound-absorbing material, the bulk modulus and effective density are mainly determined by temperature variation and flow resistivity, respectively.[26] On the assumption that the difference in flow resistivity between each direction is larger than that of temperature variation, all bulk moduli can be assumed to be identical and it is now possible to construct a single weak formulation of the functional. However, a comparison of bulk moduli should be made to check for the appropriateness of this assumption before any further calculation. Combining Eqs. (13) and (14), and using the aforementioned assumption, the wave equation for the anisotropic material can be derived as

$$\frac{1}{\rho_x}\frac{\partial^2 p}{\partial x^2}+\frac{1}{\rho_y}\frac{\partial^2 p}{\partial y^2}+\frac{1}{\rho_z}\frac{\partial^2 p}{\partial z^2}+\frac{1}{\mathrm{K}_a}\omega^2 p=0, \tag{15}$$

where K_a is the equivalent bulk modulus of the lining material. The continuity conditions of acoustic pressure and displacement for two adjacent liners can be written as

$$p_l = p_m \quad (\text{for } \forall\, l,\ \forall\, m, \text{ and } l=m), \tag{16}$$

$$\frac{1}{\rho_l}\nabla_l \cdot \mathbf{n}_l \eta_l = \frac{1}{\rho_m}\nabla p_m \cdot \mathbf{n}_m \eta_m \quad (\text{for } \forall\, l,\ \forall\, m, \text{ and } l \neq m). \tag{17}$$

The zero displacement condition on the rigid outer wall can be described as

$$\nabla p_m \cdot \mathbf{n}_m = 0 \ (\text{for } \forall\, m). \tag{18}$$

The functional, comprised of the aforementioned wave equation in each region and all boundary conditions, can be defined as

$$\chi(p_1,\ldots,p_7)=\int_{R_1}[\nabla^{\rho_0}p_1\cdot\nabla p_1-\omega^2 p_1^2/\mathrm{K}_0]dR_1+\sum_{n=2}^{7}\int_{R_n}[\nabla^{\rho_n}p_n\cdot\nabla p_n-\omega^2 p_n^2/\mathrm{K}_n]dR_n+\int_{C_{\text{piston}}}2j\omega V p_1 dC_{\text{piston}}, \tag{19}$$

where

$$\nabla^{\rho_n}=\frac{1}{\rho_{n,x}}\frac{\partial}{\partial x}\mathbf{e}_x+\frac{1}{\rho_{n,y}}\frac{\partial}{\partial y}\mathbf{e}_y+\frac{1}{\rho_{n,z}}\frac{\partial}{\partial z}\mathbf{e}_z. \tag{20}$$

Here, $\rho_{n,x}$ is the effective density of the nth lining material in the x direction and $\mathbf{e}_x$ is the unit vector in the x direction. By obtaining the variational form of Eq. (19) through the same procedure for isotropic liners, one can obtain

$$\begin{aligned}\delta\chi(p_1,\ldots,p_7)=&-2\varepsilon\int_{R_1}\left[\frac{1}{\rho_0}\frac{\partial^2 p_1}{\partial x^2}+\frac{1}{\rho_0}\frac{\partial^2 p_1}{\partial y^2}+\frac{1}{\rho_0}\frac{\partial^2 p_1}{\partial z^2}+\frac{\omega^2}{\mathrm{K}_0}p_1\right]\eta_1 dR_1+2\varepsilon\int_{C_{\text{piston}}}j\omega V\eta_1 dC_{\text{piston}}-2\varepsilon\int_{R_2}\left[\frac{1}{\rho_{2,x}}\frac{\partial^2 p_2}{\partial x^2}\right.\\&\left.+\frac{1}{\rho_{2,y}}\frac{\partial^2 p_2}{\partial y^2}+\frac{1}{\rho_{2,z}}\frac{\partial^2 p_2}{\partial z^2}+\frac{\omega^2}{\mathrm{K}_a}p_1\right]\eta_1 dR_2-\cdots-2\varepsilon\int_{R_7}\left[\frac{1}{\rho_{7,x}}\frac{\partial^2 p_7}{\partial x^2}+\frac{1}{\rho_{7,y}}\frac{\partial^2 p_7}{\partial y^2}+\frac{1}{\rho_{7,z}}\frac{\partial^2 p_7}{\partial z^2}+\frac{\omega^2}{\mathrm{K}_a}p_7\right]\eta_7 dR_7\\&+2\varepsilon\int_{\partial R_1}[\nabla^{\rho_0}p_1\cdot\mathbf{n}_1]\eta_1 d\partial R_1+2\varepsilon\int_{\partial R_2}[\nabla^{\rho_2}p_2\cdot\mathbf{n}_2]\eta_2 d\partial R_2+\cdots+2\varepsilon\int_{\partial R_7}[\nabla^{\rho_7}p_7\cdot\mathbf{n}_7]\eta_7 d\partial R_7+O(\varepsilon^2).\end{aligned} \tag{21}$$

Each term in Eq. (21) satisfies the conditions in Eqs. (1), (9), (15), (17), and (18), respectively. The only constraint on this functional is that the pressure should be identical over the interface between two different media, as explained in Sec. II C 1.

III. NUMERICAL MODEL

A. Application of the Rayleigh–Ritz method

The sound field due to the oscillating motion of the rigid pistons at the inlet and outlet ports can be obtained by finding the stationary values of the derived functional. Numerical methods, such as the FEM[17] or Rayleigh–Ritz method (RR method, hereafter),[18,27,28] have been used to find the stationary values of the functional for a given eigenvalue. FEM is now prevalently used in many areas and it also has a good capability in the design of arbitrarily shaped dissipative silencers. However, a significant CPU expenditure and huge amount of preprocessing work are inevitable when dealing with a three-dimensional problem. The RR method converges faster than the FEM provided the selected admissible functions are close to the actual eigensolutions.[29] Preprocessing work, such as meshing, is not needed. One only needs to increase the number of admissible functions, instead of remeshing the chamber, when extending the high frequency limit. However, there is no clear criterion to select the suitable admissible functions and a convergence check should be available for the given problem. Fortunately, it is known that cosine functions can serve as good admissible functions[18] in the case of rectangular ducts. The RR method seems to be a very good method in the analysis of regular-shaped acoustical systems.

The shape of admissible functions for the entire region can be approximated as

$$p(x,y,z)=\sum_{l=1,m=1,n=1} P_{lmn}\Psi_{lmn}(x,y,z), \tag{22}$$

where $\Psi_{lmn}(x,y,z)$ is the admissible function for the acoustic pressure and P_{lmn} is the coefficient or weighting function for each admissible function. The subscripts on Ψ represent the number of admissible functions for each axis of the rectangular Cartesian coordinate. In the case of isotropic lining material, substitution of Eq. (22) into Eq. (10) yields

$$\chi = \int_{R_1} \left[\nabla\left(\sum_i P_i\Psi_i\right) \cdot \nabla\left(\sum_i P_i\Psi_i\right) - k_0^2\left(\sum_i P_i\Psi_i\right)^2 \right] dR_1 + \left(\frac{\rho_0}{\rho_a}\right) \int_{R_2} \left[\nabla\left(\sum_i P_i\Psi_i\right) \cdot \nabla\left(\sum_i P_i\Psi_i\right) - k_a^2\left(\sum_i P_i\Psi_i\right)^2 \right] dR_2 + \int_{C_2} \left[2\rho_0 j\omega V\left(\sum_i P_i\Psi_i\right) \right] dC_2. \tag{23}$$

Here, the index i denotes the combined index of l, m, n, as previously defined in Eq. (22). To find the stationary values of Eq. (23), it should be differentiated with P_i for all i and equated to zero. The differentiated functional can be constructed as

$$\bar{\mathrm{K}}P = \bar{F}, \tag{24a}$$

where

$$\bar{\mathrm{K}}_{ii'} = 2 \cdot \int_{R_1} \left[\frac{\partial\Psi_i}{\partial x}\frac{\partial\psi_{i'}}{\partial x} + \frac{\partial\psi_i}{\partial y}\frac{\partial\Psi_{i'}}{\partial y} + \frac{\partial\Psi_i}{\partial z}\frac{\partial\Psi_{i'}}{\partial z} - k_0^2\Psi_i\Psi_{i'} \right] dR_1 + 2 \cdot \left(\frac{\rho_0}{\rho_a}\right) \int_{R_2} \left[\frac{\partial\Psi_i}{\partial x}\frac{\partial\Psi_{i'}}{\partial x} + \frac{\partial\Psi_i}{\partial y}\frac{\partial\Psi_{i'}}{\partial y} + \frac{\partial\Psi_i}{\partial z}\frac{\partial\Psi_{i'}}{\partial z} - k_a^2\Psi_i\Psi_{i'} \right] dR_2, \tag{24b}$$

$$\bar{F}_i = -2\rho_0 j\omega V \int_{C_2} \Psi_i dC_2, \tag{24c}$$

$$P = [P_{000} \cdots P_{00N}\, P_{01N} \cdots P_{LMN}]^T. \tag{24d}$$

The rate of convergence of the RR process depends on the quality of the admissible functions and, in particular, how well linear combinations of these functions can approximate the actual eigenfunctions. However, because the actual eigenfunctions are not known *a priori*, an assessment of the choice of admissible functions can be made only after examination of the numerical results. Admissible functions should consist of linear combinations of functions, which at least satisfy the geometrical boundary conditions.[29] In this operation, the derivatives of the functions at the rigid outer wall should be zero for the condition of acoustically hard walls. Cosine functions are selected as one of the global admissible functions for this rectangular lined plenum chamber, as follows:

$$\Psi_{lmn}(x,y,z) = \cos\left(\frac{(l-1)\pi x}{L_x}\right)\cos\left(\frac{(m-1)\pi y}{L_y}\right) \times \cos\left(\frac{(n-1)\pi z}{L_z}\right). \tag{25}$$

These admissible functions automatically fulfill the pressure continuity condition across the interface between the airway and liner. Moreover, the orthogonality of the cosine functions will facilitate the ease of the application of the RR method. Substitution of Eq. (25) into Eq. (24) yields the sound pressure at an arbitrary point inside the chamber. Although the condition of velocity continuity between liner and airway can only be approximately satisfied by employing a number of continuous admissible functions, the solution converges to a value but with unavoidable errors in the region which is very near to the interface, as like the Gibb's phenomenon in Fourier series theory.

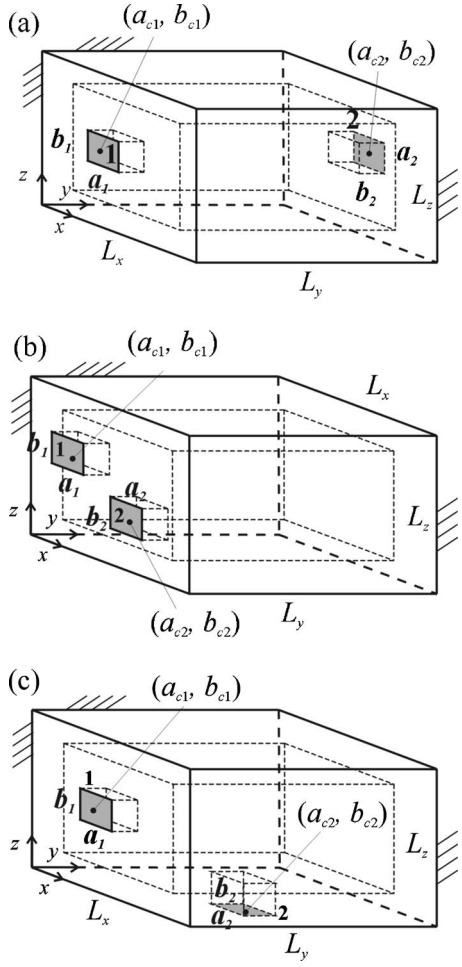


FIG. 3. Geometrical layout of three types of lined rectangular plenum chambers: (a) throughflow type; (b) reversal type; (c) end-in/side-out type.

B. Derivation of the transfer matrix and transmission loss

The transfer matrix of the plenum can be derived from the expression for the sound field, which is formed by the excitation at an individual part in inlet or outlet side.[10] All three possible types of lined chambers are illustrated in Fig. 3. The average sound pressure on a region i' with area $S_{i'}$ excited from the input piston i is given by

$$\bar{p}_{i'i} = \frac{1}{S_{i'}} \int \int_{S_{i'}} \left[\sum_n P_{n,i}\Psi(x,y,z) \right] ds, \tag{26}$$

where $P_{n,i}$ is the coefficient of admissible functions when the moving piston is positioned on region i with area S_i. The inlet and outlet ports are represented by 1 and 2, respec-

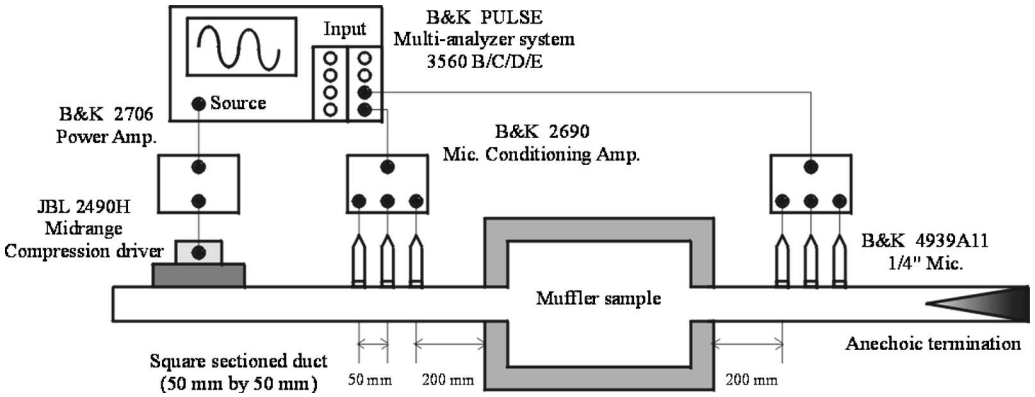


FIG. 4. Measurement setup for transmission loss, based on the three-microphone method.

tively, in Fig. 3. The total sound pressure acting on each piston is the sum of the acoustic pressure due to its own motion and the motion of the opposite piston as

$$\bar{P}_1 = \bar{p}_{11} + \bar{p}_{12}, \tag{27a}$$

$$\bar{P}_2 = \bar{p}_{21} + \bar{p}_{22}. \tag{27b}$$

Then, the transfer matrix can be given by

$$\begin{bmatrix} \bar{P}_1 \\ U_1 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} \bar{P}_2 \\ U_2 \end{bmatrix}, \tag{28}$$

where the volume velocity at the inlet and outlet ports are defined, respectively, as

$$U_1 = S_1 V_1, \tag{29a}$$

$$U_2 = S_2 V_2. \tag{29b}$$

By the assumption of unit velocity excitation, the four-pole parameters $T_{ii'}$ can be written as

$$T_{11} = \left. \frac{\bar{P}_1}{\bar{P}_2} \right|_{U_2=0} = \frac{\bar{p}_{11}}{\bar{p}_{21}}, \tag{30a}$$

$$T_{12} = \left. \frac{\bar{P}_1}{\bar{P}_2} \right|_{\bar{P}_2=0} = \frac{-\bar{p}_{11} \cdot (\bar{p}_{22}/\bar{p}_{21}) + \bar{p}_{12}}{S_2}, \tag{30b}$$

$$T_{21} = \left. \frac{U_1}{\bar{P}_2} \right|_{U_2=0} = \frac{S_1}{\bar{p}_{21}}, \tag{30c}$$

$$T_{22} = \left. \frac{U_1}{U_2} \right|_{\bar{P}_2=0} = \frac{-S_1 \bar{p}_{22}}{S_2 \bar{p}_{21}}. \tag{30d}$$

For anechoic termination with unequal inlet and outlet areas, the transmission loss can be given by

$$\mathrm{TL(dB)} = 20 \log_{10}\left(\frac{1}{2} \sqrt{\frac{S_1}{S_2}} \left| T_{11} + \frac{S_2}{\rho_0 c} T_{12} + \frac{\rho_0 c}{S_1} T_{21} + \frac{S_2}{S_1} T_{22} \right| \right). \tag{31}$$

IV. MEASUREMENTS

A. Experimental setup

Experiments were carried out to measure transmission loss of the lined chamber. The source sound was generated by a 4-in., 200-W midrange compression driver (JBL 2490H) mounted on an upstream side-wall, with a lower cut-off frequency of 200 Hz. A square acrylic duct, 50×50 mm^2 in internal section, 1.5 m in length, and 10 mm in thickness, was used for the upstream and downstream parts. The first higher mode cut-on frequency in the acrylic duct was about 3.4 kHz. A glass fiber wedge of 1.0 m in length was inserted in the same acrylic box of 50×50 mm^2 and it was used as the anechoic termination. The measured power reflection coefficient was less than 0.1 above 150 Hz. A rectangular plenum chamber, which had inner dimensions of $0.2 \times 0.275 \times 0.2$ m^3, was made from acrylic plates of 10 mm in thickness. All three kinds of the rectangular chambers, i.e., throughflow-, reversal- and end-in/side-out-type, were tested. In addition, the thickness and property of the lining material were also varied to examine their effects on acoustic performance. The measurement was based on the multiple microphone technique[30] in the absence of mean flow, as depicted in Fig. 4. Acoustic transfer functions were measured with flush-mounted 1/4-in. microphones (B&K 4135). The spacing between microphones was 50 mm, which limits the effective frequency range to about 3.4 kHz. A random white signal in the range 0–3 kHz was fed to a compression driver from the signal analyzer (B&K Pulse) and the measured acoustic signals were processed by the same analyzer.

B. Bulk properties of the lining material

Polyurethane (PU) foam and thermally fused polyester fibrous material were chosen as the test liners. The multi-termination method[22] for the measurement of characteristic impedance and propagation constants was adopted. Four types of end terminations, such as rigid end and semi-anechoic termination, and two air cavities (20 and 40 mm in depth) were employed to reduce random errors. A regression technique for experimental data[21] was used to calculate the required acoustic parameters of the fibrous materials using the measured flow resistivity. Experimentally, flow resistivity

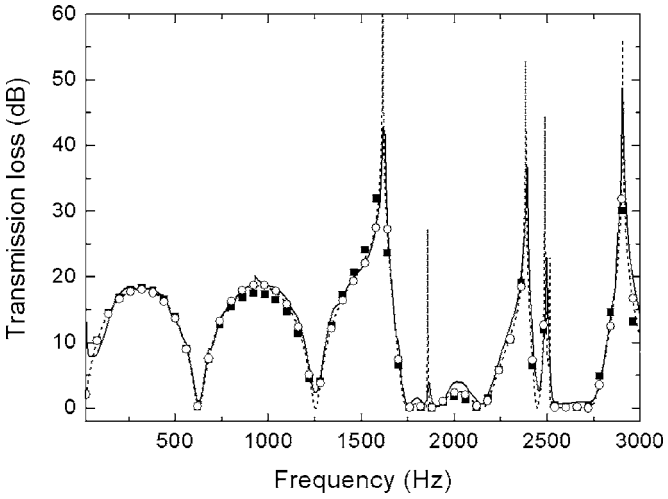


FIG. 5. Comparison of measured and calculated transmission losses for a rectangular throughflow chamber with rigid walls (L_x=L_z=0.2 m, L_y =0.275 m, a_1=a_2=b_1=b_2=0.05 m, centered inlet/outlet): (—) measured; (---) analytic solution (Ref. 10); (■) predicted by the RR method using 175 modes (5, 7, 5) for three directions; (○) predicted by the RR method using 1400 modes (10, 14, 10) for three directions.

of the specimen was measured by passing a compressed, regulated airflow through the sample along a straight circular tube of 40 mm in diameter. The sample was located at the inside of the duct having one side open to atmospheric pressure. The pressure drop across the sample was measured by a manometer (Furness Controls FCO12) and the airflow rate was measured by a flow meter (Flowmetrics FM-20N). Because the fibrous material was orthotropic,[24] the characteristic impedance and propagation constants were measured for both normal and transverse directions.

V. RESULTS AND DISCUSSIONS

Measured and predicted transmission losses of various lined rectangular plenum chambers were compared to validate the predictive method. The fluid medium was air and the temperature was maintained at 19 °C which was constant during all the conducted experiments.

First, a rectangular plenum chamber with acoustically rigid walls, i.e., without liners, was tested to check the applicability to the limit condition. When applying the RR method, the order of admissible functions in each direction was determined by the ratio of the plenum chamber dimensions. The maximum number of admissible functions in the y direction was 14. Figure 5 shows a comparison between predicted and measured results. It was found that the predicted result by the RR method converged, which also agreed very well with analytical and numerical results.

The convergence of the solution is very important for the RR method, because, in general, no exact analytic solution exists, such as in the case of a lined plenum chamber. When carrying out a convergence check with the RR method, one should be cautious in determining the number of admissible functions for each direction. Typically, it is known that the more admissible functions included, the more accurate the results. However, it should be borne in mind that

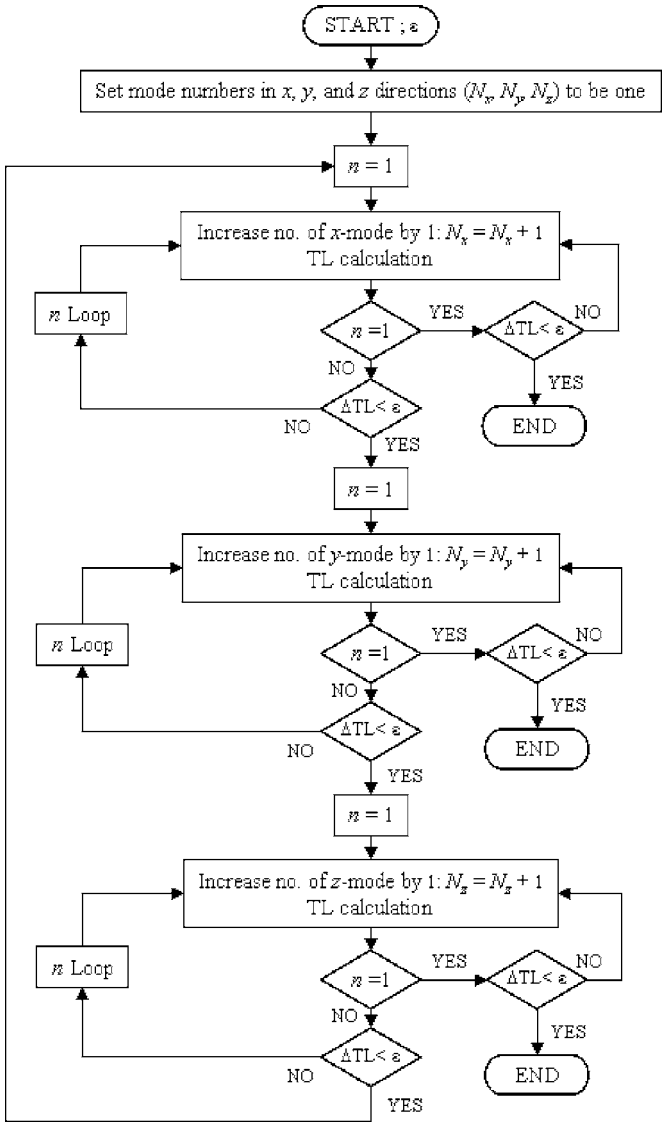


FIG. 6. Flow chart for convergence check of the RR method.

the benefit of the RR method will be reduced as the number of admissible functions increases due to the large expenditure of CPU.

In this paper, the convergence of solutions can be detected by the procedure in the flow chart depicted in Fig. 6. As can be seen in Fig. 6, the numbers of admissible functions in the x, y, and z directions should be increased, one by one, until an acceptable convergence criterion is satisfied. Here, the convergence criterion, ε, is set as 0.2 dB, which corresponds to the error in sound energy by less than 5%.

A. Throughflow type

1. Isotropic lining material

The present method was applied to a rectangular plenum chamber lined with an open-cell PU foam (density =30 kg/m^3) of varying thickness (5 and 20 mm). The geometry of the chamber was the same as in Fig. 5 for the rigid-walled chamber. First, a convergence check was performed by increasing the number of admissible functions until the computed TL value converges within the acceptable error

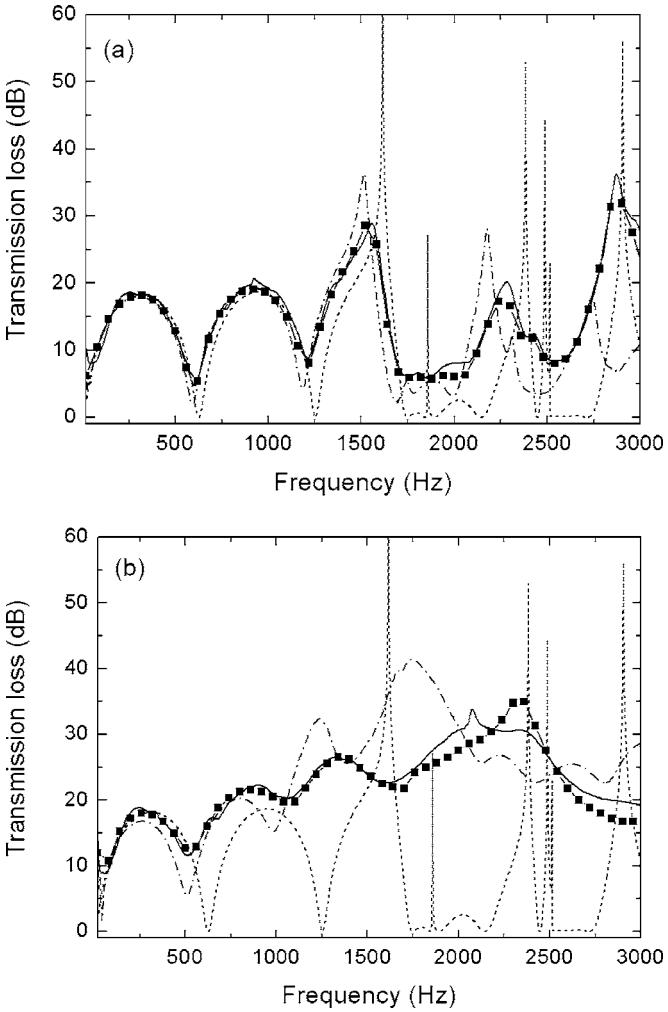


FIG. 7. Comparison of measured and predicted transmission losses for a throughflow-type lined rectangular chamber with the same sectional dimension as in Fig. 5: (—) measured; (---) predicted without liner; (–·–) predicted by locally reacting liner model (Ref. 11); (-■-) predicted by the RR method for bulk-reacting liner model. (a) PU foam (5 mm thick) and (b) PU foam (20 mm thick).

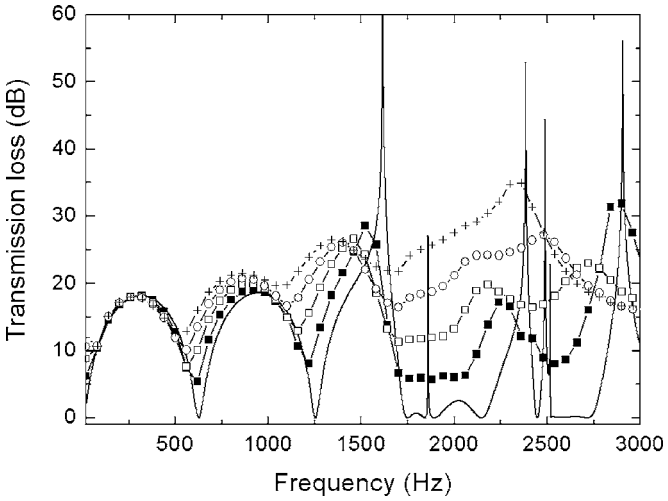


FIG. 8. Change in the predicted transmission loss by increasing the thickness of the PU foam liner. The geometry of plenum chamber is the same as for Fig. 5: (—) prediction without liner; (-■-) 5 mm; (-□-) 10 mm; (-○-) 15 mm; (-+-) 20 mm.

boundary of 0.2 dB, which corresponds to sound energy error by less than 5%. The required number of admissible functions was found to be 2873 (number of modes is 13, 17, and 13 at x, y, and z direction, respectively) for a 5-mm liner and 3887 (13, 23, and 13) for a 20-mm liner.

Figure 7 shows a comparison between measured and predicted transmission losses for the rectangular plenum chamber. Dash-dotted lines in Fig. 7 are the predicted results using the locally reacting model,[11] which strongly deviate from the measured TL curves, especially when the liner becomes thick. In contrast, when the RR method was applied by employing the bulk properties of the lining material, there was a very good agreement between measured and predicted TL curves, except for a peak between 1.8 and 2.2 kHz. This peak originates from the higher order mode effects, but it gets flattened and moves to the low frequency region due to the sound attenuation in lining material.[3] A trend was also observed where the difference between measured and predicted results increases slowly with an increase in frequency. However, if one recalls that the precise prediction of troughs in TL curves is very important in silencer design, the result in Fig. 7(b) can be regarded as quite acceptable in this sense.

When the thickness of the lining material is smaller than the width of the airway, the convergence rate of predicted transmission loss is higher because the cosine function, which is chosen as the admissible function, becomes the eigenfunctions of a chamber with acoustically hard walls. The effects of liner thickness on transmission loss were also examined in Fig. 8. With the same geometry as shown in Fig. 5 and PU foam as liner, the thickness of the liner varied between 0 and 20 mm in 5-mm step. The number of admissible functions was 3887 (13, 23, 13), which was the same number of functions required in the case of a 20-mm liner. The reactive action of the chamber was predominant below 500 Hz, so all curves showed similar trends regardless of liner thickness. One can still find the influence from the reactive wave action of the chamber until the third lobe of TL curve. The resistive action by the liner starts to be effective above 500 Hz. The curves become smooth and the amount of TL becomes higher by increasing the liner thickness at high frequencies. It is also noticeable that the peaks and troughs of the transmission loss curves have a tendency to shift to the low-frequency range as the thickness increases. This is due to the fact that the effective length of the chamber is increased due to the high sound attenuation at the wall, similar to the phenomenon occurring in an anechoic chamber.

2. Orthotropic lining material

When the fibrous material is fully lined inside the chamber, the liner can be considered an orthotropic material. The rectangular plenum chamber is lined as depicted in Fig. 2, where the normal direction of liner is directed toward the center of the chamber.

The physical and acoustical properties of thermally fused polyester fibrous liner materials are as follows: 60 kg/m3 in density, 13 mn in thickness; the fluid resistivity in the normal and transverse direction is 3478 and 1490 SI Rayl/m, respectively. The measurement of flow resistivity

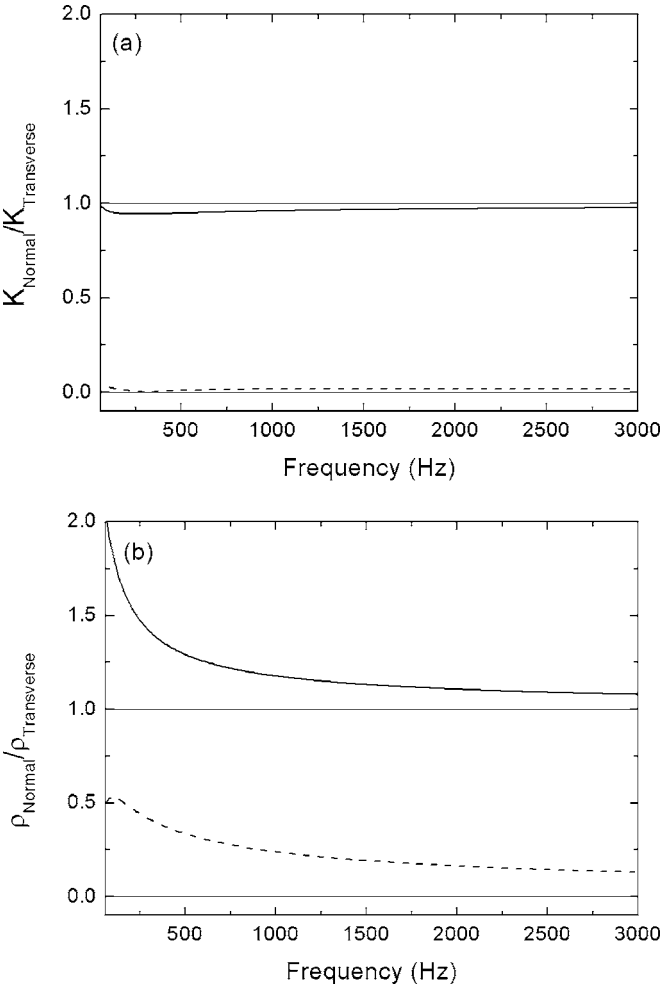


FIG. 9. Ratio of bulk modulus and effective density of normal direction to that of transverse direction for fibrous liner material: (—) absolute value of real part; (---) absolute value of imaginary part. (a) Bulk modulus and (b) effective density.

was carried out with the same method and device described in Sec. IV B, but the specimen was tested in normal and transverse directions.

The ratio of bulk modulus of normal direction to that of transverse direction was assumed to be unity in deriving Eq. (15), which was the single weak formulation of the functional. To validate this assumption, the effective density and bulk modulus ratio was calculated, as shown in Fig. 9. Effective density and bulk moduli were calculated from Eqs. (7) and (8). It was found that the difference was less than 5% by assuming the bulk modulus ratio to be unity within the frequency range of interest. As can be seen in Fig. 9(b), the effective density for each direction is significantly different from each other.

Figure 10 is the resultant transmission loss curves for a plenum chamber lined with fibrous material. From convergence testing, as illustrated in Fig. 6, the number of admissible functions was determined to be 2783 (11, 23, 11 for the x, y, and z directions, respectively). The assumption of bulk modulus can be validated if two predicted TLs, using each bulk modulus of normal and transverse direction, agree within acceptable error limits. It was found that two predicted results differ from the measured data by about

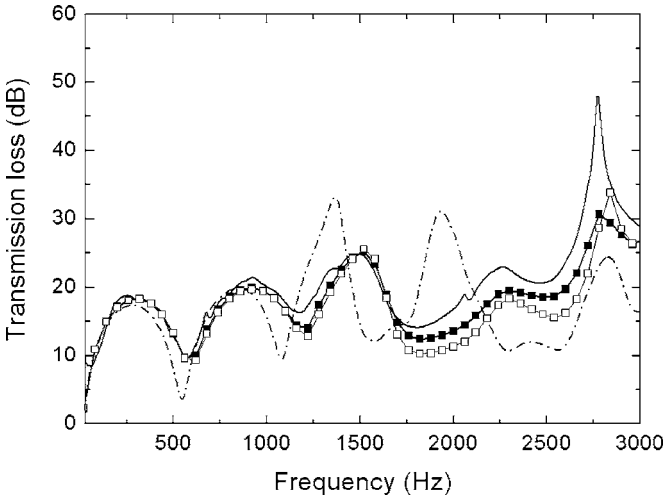


FIG. 10. Comparison of measured and predicted transmission losses for a throughflow-type rectangular plenum chamber, the inside of which is fully lined with 13 mm fibrous material (same sectional dimensions as for Fig. 5): (—) measured; (–·–) predicted by locally reacting liner model; (-■-) predicted by the RR method for bulk-reacting liner model with bulk modulus K_{normal}; (-□-) predicted by the RR method for bulk-reacting liner model with bulk modulus $K_{transverse}$.

3–4 dB over 1.7 kHz. Unfortunately, this discrepancy means that our assumption on the bulk modulus was not totally true over 1.7 kHz. However, the overall trends and locations of troughs and peaks could still be predicted very closely at all frequency ranges of interest. In contrast, the TL curve predicted by the assumption of locally reacting liner deviated largely from the measured transmission loss, in particular, at troughs.

B. Other types of lined plenum chambers

The present method can be applied to other types of plenum chambers, such as the reversal- and end-in/side-out-type chamber. PU foam of 10 mm in thickness and fibrous material of 13 mm in thickness were used as the liner for testing these configurations. Numbers of admissible functions for converged results are listed in Table I. In Figs. 11(a) and 11(b), the results of the reversal types with the two aforementioned liners are plotted. In the case of the PU foam liner, predicted transmission loss agrees very well with measured data, except for a discrepancy at 2 kHz. This is mainly due to the fact that the inlet and outlet ports are close and located at the corner. In such a case, the number of high-order modes generated at each port does not decay satisfactorily, so that a lot of modal functions are needed to describe

TABLE I. Mode number of admissible functions needed for a converged result of the RR method in the case of a reversal or end-in/side-out type plenum. N_x, N_y, and N_z denote the number of admissible functions in the x, y, and z directions, respectively.

Plenum type	Liner	N_x	N_y,	N_z
Reversal	PU foam 10 mm	9	9	9
	Polyester fiber 13 mm	7	9	7
End-in/side-out	PU foam 10 mm	9	11	9
	Polyester fiber 13 mm	13	13	13

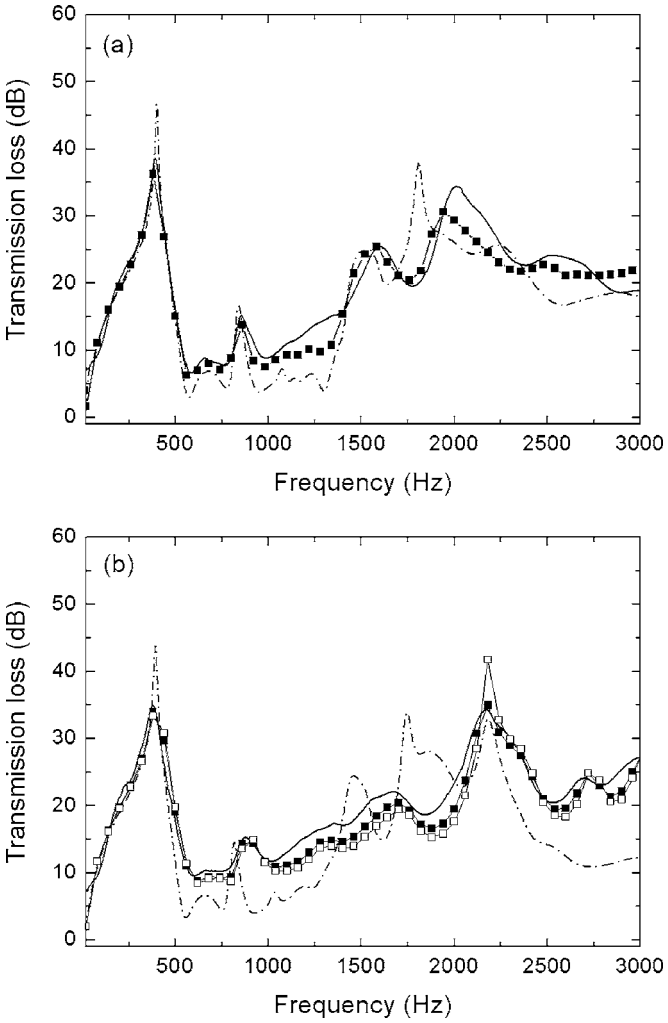


FIG. 11. Comparison of measured and predicted transmission losses for a reversal-type lined chamber ($L_x=L_z=0.2$ m, $L_y=0.275$ m, $a_1=a_2=b_1=b_2=0.05$ m, $a_{c1}=b_{c1}=0.055$ m, $a_{c2}=b_{c2}=0.145$ m): (—) measured; (-·-) predicted by locally reacting liner model; (-■-) predicted by the RR method for bulk-reacting liner model with K_{normal}; (-□-) predicted by the RR method for bulk-reacting liner model with $K_{\text{transverse}}$. (a) PU foam (10 mm thick) and (b) fibrous material (13 mm thick).

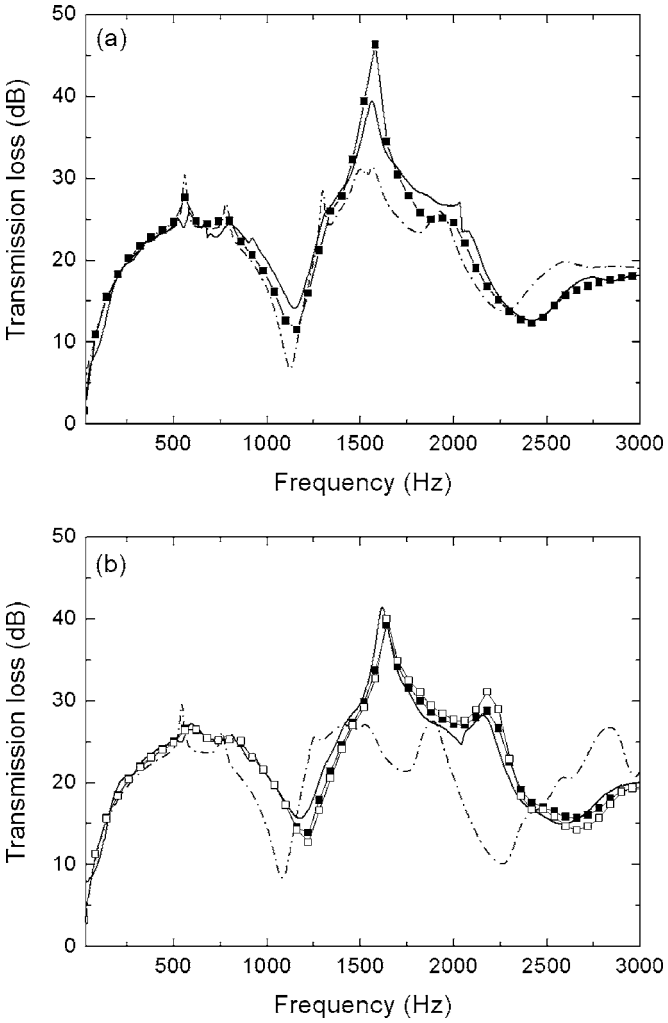


FIG. 12. Comparison of measured and predicted transmission losses for an end-in/side-out type lined chamber ($L_x=L_z=0.2$ m, $L_y=0.275$ m, $a_1=a_2=b_1=b_2=0.05$ m, $a_{c1}=b_{c1}=0.1$, $a_{c2}=0.1$, $b_{c2}=0.1375$): (—) measured; (-·-) predicted by locally reacting liner model; (-■-) predicted by the RR method for bulk-reacting liner model with K_{normal}; (-□-) predicted by the RR method for bulk-reacting liner model with $K_{\text{transverse}}$. (a) PU foam (10 mm thick) and (b) fibrous material (13 mm thick).

the sound field.[31] Consequently, cosine functions, chosen as admissible functions, cannot serve as real eigenfunctions very successfully at peaks. Similar to the previous results explained in Sec. V A 2, valid frequency ranges also exist when applying the present method to the case of a reversal-type chamber, as shown in Fig. 11(b). In the case of end-in/side-out-type chamber, as shown in Fig. 12, predicted TL curves by the present RR method for bulk-reacting liner model match very well with measured data at all frequency ranges of interest. The discrepancies between measured and predicted results, based on the locally reacting liner model, are prominent in chambers lined with 13-mm-thick fibrous material. Also, the low frequency shift of TL troughs is apparent for chambers lined with 10-mm-thick PU foam around 1.2 kHz.

In comparison with all three types of chambers, one can find that the reversal-type chamber behaves like a quarter-wavelength tube at low frequencies.[31] Therefore, the TL curve has a very sharp peak below the cut-on frequency of the first cross mode. It can also be observed that the end-in/side-out configuration is the most efficient in silencing the low-frequency range of 0–1.2 kHz, as depicted in Figs. 8 and 10–12. This is due to the elimination of the first longitudinal mode by placing the outlet ports at the nodal plane of that mode.[32,33]

VI. CONCLUSIONS

In this study, the 2×2 transfer matrix was derived for the prediction of acoustic performance in lined rectangular plenum chambers. The lined rectangular plenum chamber was modeled on a chamber excited by oscillating pistons that represent the inlet and outlet ports. Both isotropic and orthotropic sound-absorbing materials were considered as liners. To obtain the stationary value of the modified single weak variational functional, the Rayleigh–Ritz method was adopted as the numerical scheme to reduce the preprocessing operation. By changing the location of the rigid oscillating pistons, all three possible types of lined chambers, such as

throughflow, reversal, and end-in/side-out types, were studied and their corresponding transfer matrices were obtained. Good agreements were observed between measured and predicted transmission losses for different types of chambers and lining materials. In comparison with the absorbent model, based on the locally reacting liner assumption, the present Rayleigh–Ritz method, using the bulk-reacting liner model, could predict the TL very accurately, especially the locations of TL troughs. The suggested single weak variational statement can be used effectively for deriving the transfer matrix of other regular cross-sectional shapes of a lined plenum chamber excited by the plane wave if the admissible functions are chosen properly.

However, there exist some difficulties in the convergence and choice of admissible functions when dealing with large-sized or severely irregular-shaped dissipative silencers. When applying the present method to silencers with a dimension on the order of several meters or corresponding to a large number of wavelengths, the slow convergence rate would be a serious problem in the computation and the given CPU expenditure would be exceeded. In this case, no merit in using the present method can be expected. Consequently, a full finite element modeling will be a practical method for this case, although a large CPU expenditure would be inevitable yet.

ACKNOWLEDGMENTS

The authors are grateful to Professor. M. Åbom and Professor H. Bodén at KTH, Sweden, for their useful comments. This study was partially supported by the BK21 project.

[1]D. E. Baxa, *Noise Control in Internal Combustion Engines* (Wiley, New York, 1982), Chap. 5.
[2]R. J. Wells, "Acoustical plenum chambers," Noise Control **4**, 9–15 (1958).
[3]A. Cummings, "The attenuation of lined plenum chambers in ducts. I. Theoretical Models," J. Sound Vib. **61**, 347–373 (1978).
[4]R. Mittra and S. W. Lee, *Analytical Technique in the Theory of Guided Waves* (Macmillan, New York, 1971), Chap. 2.
[5]A. Cummings, "Sound transmission in a folded annular duct," J. Sound Vib. **41**, 375–379 (1975).
[6]A. Cummings and I.-J. Chang, "Sound attenuation of a finite length dissipative flow duct silencer with internal mean flow in the absorbent," J. Sound Vib. **127**, 1–17 (1988).
[7]M. Åbom, "Derivation of four-pole parameters including higher order mode effects for expansion chamber mufflers with extended inlet and outlet," J. Sound Vib. **137**, 403–418 (1990).
[8]R. Glav, "The transfer matrix of a dissipative silencer of arbitrary cross-section," J. Sound Vib. **236**, 575–594 (2000).
[9]J. Kim and W. Soedel, "Development of a general procedure to formulate four pole parameters by modal expansion and its application to three-dimensional cavities," J. Vibr. Acoust. **112**, 452–459 (1990).
[10]J.-G. Ih, "The reactive attenuation of rectangular plenum chambers," J. Sound Vib. **157**, 93–122 (1992).
[11]H.-J. Kim, J.-G. Ih, and C.-M. Park, "Acoustic performance of the lined rectangular plenum chamber," Proceedings Inter-Noise 2003, Jeju, Korea, pp. 856–861.
[12]R. Kirby and J. B. Lawrie, "A point collocation approach to modelling large dissipative silencers," J. Sound Vib. **286**, 313–339 (2005).
[13]R. A. Scott, "The propagation of sound between walls of porous material," Proc. Phys. Soc. **58**, 358–368 (1946).
[14]K. S. Peat, "A transfer matrix for an absorption silencer element," J. Sound Vib. **146**, 353–360 (1991).
[15]S. N. Panigrahi and M. L. Munjal, "Comparison of various methods for analyzing lined circular ducts," J. Sound Vib. **285**, 905–923 (2005).
[16]K. S. Peat and K. L. Rathi, "A finite element analysis of the convected wave motion in dissipative silencers," J. Sound Vib. **184**, 529–545 (1995).
[17]R. J. Astley and A. Cummings, "A finite element scheme for attenuation in ducts lined with porous material: Comparison with experiment," J. Sound Vib. **116**, 239–263 (1987).
[18]B. Farvacque, "Modelling of Large Dissipative Silencers," M.S. thesis, KTH, 2003.
[19]L. L. Beranek, "Acoustical properties of homogeneous, isotropic rigid tiles and flexible blankets," J. Acoust. Soc. Am. **19**, 556–568 (1947).
[20]M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).
[21]F. P. Mechel, "Design charts for sound absorber layers," J. Acoust. Soc. Am. **83**, 1002–1013 (1988).
[22]J.-G. Ih, J.-H. Lee, and Y.-I. Kwon, "On the precision measurement of bulk acoustic properties of absorbing materials in an impedance tube," in Proceedings of the International Symposium on Room Acoustics: Design and Science 2004 (CD ROM), Hyogo, Japan.
[23]S. K. Kakoty and V. K. Roy, "Bulk reaction modeling of ducts with and without mean flow," J. Acoust. Soc. Am. **112**, 75–83 (2002).
[24]J. F. Allard, *Propagation of Sound in Porous Media—Modelling Sound Absorbing Materials* (Elsevier, New York, 1993), Chap. 3.
[25]K. U. Ingard, *Notes on Sound Absorption Technology* (Noise Control Foundation, New York, 1994), Chap. 4.
[26]A. Cummings and I.-J. Chang, "Acoustic propagation in porous media with internal mean flow," J. Sound Vib. **114**, 565–581 (1987).
[27]A. Cummings and R. J. Astley, "The effects of flanking transmission on sound attenuation in lined ducts," J. Sound Vib. **179**, 617–646 (1995).
[28]A. Cummings, "A segmented Rayleigh-Ritz method for predicting sound transmission in a dissipative exhaust silencer of arbitrary cross-section," J. Sound Vib. **187**, 23–37 (1995).
[29]L. Meirovitch and M. K. Kwak, "Convergence of the classical Rayleigh-Ritz method and the finite element method," AIAA J. **28**, 1509–1516 (1990).
[30]S.-H. Jang and J.-G. Ih, "On the multiple microphone method for measuring in-duct acoustic properties in the presence of mean flow," J. Acoust. Soc. Am. **103**, 1520–1526 (1998).
[31]J.-G. Ih and B.-H. Lee, "Theoretical prediction of transmission loss of circular reversing chamber mufflers," J. Sound Vib. **112**, 261–272 (1987).
[32]J.-G. Ih and J.-S. Lee, "Low frequency characteristics of unlined end-in/side-out rectangular plenum chambers," Noise Control Eng. J. **40**, 179–185 (1993).
[33]J.-G. Ih and B.-H. Lee, "Analysis of higher order mode effects in the circular expansion chamber with mean flow," J. Acoust. Soc. Am. **77**, 1377–1388 (1985).

Scattering of the fundamental torsional mode by an axisymmetric layer inside a pipe

J. Ma, F. Simonetti, and M. J. S. Lowe[a)]
Department of Mechanical Engineering, Imperial College London, London SW7 2AZ, United Kingdom

(Received 9 February 2006; revised 4 July 2006; accepted 19 July 2006)

The scattering of the fundamental guided torsional mode by a local axisymmetrical layer coated inside a pipe, referred to as a bilayered pipe, is studied. In a prescribed frequency range, the number of torsional modes which can propagate in an empty pipe is increased by the presence of the coating layer, including new cutoff frequencies that depend on the layer thickness and shear acoustic properties. This principle suggests the potential for detecting, and perhaps characterizing layers inside pipes, which may be exploited in two ways using either remote or local measurements. A remote measurement may be employed to measure the reflection from the entry point of the layer inside the pipe. Such a measurement shows that the reflection coefficient spectrum exhibits periodic maxima that occur at the cutoff frequencies of the torsional modes in the bilayered pipe. On the other hand, when the location of the coating layer is accessible, the local guided wave measurement can be made to measure the dispersion curves of the torsional modes in the bilayered pipe by applying a reassigned spectrogram analysis. The idea is investigated through theoretical analysis and finite element modeling and is validated by experimental measurements. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2336750]

PACS number(s): 43.20.Mv, 43.35.Zc [LLT] Pages: 1871–1880

I. INTRODUCTION

Detecting and characterizing contents inside pipes is required by many industries. For example, it is important for the process industry to monitor the accumulation of sludge in pipes in order to prevent the formation of blockages. In this context, it would be valuable to be able to quantify the material and geometry properties of the sludge layer noninvasively. Existing techniques using ultrasonic bulk waves sent through the pipe wall thickness have the limitation that the sludge position needs to be known *a priori* and the area to be inspected needs to be accessible.

Guided ultrasonic waves propagating along the pipe may be an attractive method to do this, because of their capability of inspecting a long length from a single point,[1,2] which could fulfill the need for remote detection and even characterization. The basic idea is that the presence of the content inside the pipe will influence the propagation of guided waves compared to that in a free pipe. Some properties of the content may then be extracted by studying the change of guided wave behavior. In the case of sludge detection, the problem becomes very complicated due to the irregular shape and properties of the sludge layers that can be encountered in practice. However, in order to assess the feasibility and potential of such an approach the simplified case of an axisymmetrical layer inside the pipe can be considered (Fig. 1). This study, based on such a "bilayered pipe" model, has been referred to in order to understand the fundamental physics and provide an insight into the possibility of characterizing the sludge using guided ultrasonic waves.

Models of guided wave propagation in filled waveguides, relating to various applications, have been the subject of numerous studies. Kumar[3] examined dispersion of axially symmetrical waves in empty and fluid-filled cylindrical shells. Lafleur and Shields[4] discussed the influence of the pipe material on the first two longitudinal modes in a liquid-filled pipe, considering the low frequency range only. Sinha and co-workers[5,6] addressed the case of axial-symmetric guided wave propagation in pipes with fluid on the inside or the outside of the pipes. The case of a pipe filled with a viscous liquid has been theoretically analyzed by Elvira.[7] Aristegui *et al.*[8] reported a series of experimental measurements of guided wave propagation in pipes filled and surrounded by different media, including pipes filled with inviscid water and viscous liquid. Vollmann *et al.*[9] have theoretically and experimentally[10] studied the guided waves propagating in a cylindrical shell containing a viscoelastic medium. Simonetti and Cawley[11] developed a new approach to characterize highly attenuative viscoelastic materials by filling the material into a pipe and measuring the dispersion curves of guided modes in such a waveguide.

It has been generally observed in the above-mentioned works that in a prescribed frequency range, the number of guided wave modes which can propagate in an empty pipe is increased by the presence of the filling content, which is exhibited by the generation of new cutoff frequencies.[7–11] The physical mechanism behind the appearance of the new cutoff frequencies in the bilayer structure has been discussed by Simonetti[12] and Simonetti and Cawley.[13] For example, let us consider the case of torsional waves propagating in a bilayered pipe, consisting of an aluminum pipe (16 mm inner diameter and 1 mm wall thickness) coated with a 6-mm-thick epoxy layer inside (Fig. 1). Figure 2(c) shows the group velocity dispersion curves of the torsional waves which can propagate in the pipe. They were calculated using the modeling software DISPERSE developed at Imperial Col-

[a)]Electronic mail: m.lowe@imperial.ac.uk

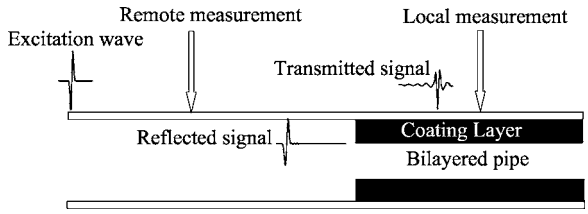


FIG. 1. Schematic of the model investigated in the paper.

lege. The software has been developed to model guided wave propagation in layered system and more details can be found in Refs. 14 and 15. The input parameters to DISPERSE are the material and dimensions of the pipe and the coating layer which are summarized in Table I. The modes of the bilayered pipe are labeled T_1, T_2, T_3, T_4. For comparison, the torsional mode in the free pipe labeled as T(0,1) is also given. This shows that in the same frequency range there are many more modes occurring in the bilayered pipe compared to the single T(0,1) mode in the free pipe. Note that although only one torsional mode can propagate in the free pipe in the frequency range considered in Fig. 2(a), higher order mode will occur at higher frequencies. With the occurrence of the new modes, new cutoff frequencies are generated accordingly. Also, the modes of the bilayered pipe are very dispersive, unlike the T(0,1) mode of the free pipe which always keeps a constant velocity value for all frequencies.

In this paper, we develop two measurement ideas to characterize the layer inside the pipe, by making use of these new cutoff frequencies of torsional modes in the bilayered pipe. The model studied here for both approaches consists of a pipe locally coated with an epoxy layer inside, which is a simplified model of sludge in a pipe (Fig. 1). The T(0,1) mode is excited in the free pipe region; this mode will propagate along the free pipe until it reaches the location where the bilayered pipe region starts. Due to the acoustic impedance change, the torsional wave will be scattered: part of the T(0,1) mode will be reflected back to the free pipe, while part will be transmitted into the bilayered pipe region and mode converted into the bilayered pipe modes.

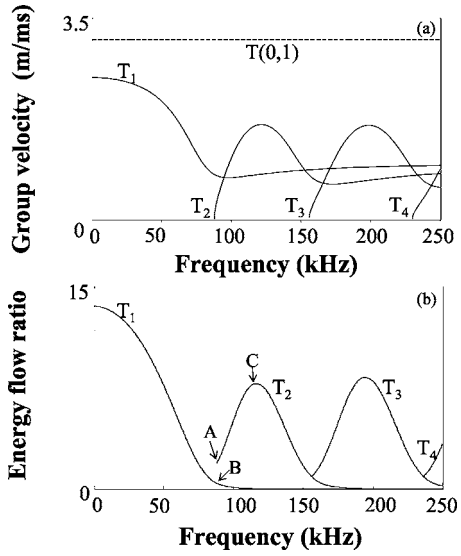


FIG. 2. (a) Group velocity dispersion curves of torsional modes in an aluminum pipe (inner diameter 16 mm and thickness 1 mm) coated with a 6-mm-thick epoxy layer inside. Material properties are given in Table I. For comparison, the torsional mode in a pipe without coating is also shown, by the dashed line. (b) Energy flow ratio.

TABLE I. Material properties of the aluminum pipe and epoxy layer used for DISPERSE calculation and FE modeling.

	Bulk longitudinal velocity (m/s)	Bulk shear velocity (m/s)	Density (kg/m^3)
Aluminum	6320	3130	2700
Epoxy	2610	1000	1170

The first measurement is carried out on the free pipe region, measuring the reflection of the T(0,1) mode from the entry point of the layer inside the pipe (Fig. 1). The characteristic of the reflection coefficient spectrum is associated with the cutoff behavior of torsional modes in the bilayered pipe region. This measurement is aimed to be applicable to those circumstances when the content inside the pipe cannot be located or the region of interest is inaccessible. When the cutoff frequencies have been measured, the thickness of the coating layer can be obtained, provided that the bulk shear velocity has been known beforehand.

The second measurement is performed at a local position of the bilayered pipe, where some energy of the free pipe mode T(0,1) has been converted into the bilayered pipe modes (Fig. 1). A signal processing technique, the reassigned spectrogram, is applied on the transmitted signal to distinguish the multiple modes and to reveal the group velocity dispersion curves. The extraction of the coating properties can be achieved from the measured cutoffs as before or by best fitting analytical dispersion curves to the measured ones. This second method is more reliable in the presence of high noise levels.

The investigation starts with an introduction of the propagation of torsional waves in bilayered pipes, followed by a detailed analysis of their energy flow distribution (Sec. II). The physical insight gained from this analysis enables the design of the two guided wave measurement concepts to characterize the coating layer inside the pipe. These are demonstrated first by finite element modeling in Sec. III. Then experiments are presented to validate the finite element modeling in Sec. IV. General discussion on the applicability of presented measurements follows in Sec. V.

II. TORSIONAL WAVES IN THE BILAYERED PIPE

In order to understand the nature of the scattered field of the T(0,1) mode from the discontinuity where the free pipe meets the bilayered pipe, the characteristics of the torsional waves propagating in the bilayered pipe need to be studied.

Let us consider a hollow, elastic, and isotropic cylinder coated with an axisymmetric elastic layer inside. The cylindrical coordinates (r,θ,z) are appropriate here to represent the geometry of the system (r,θ,z denote the radial, angular, and axial position, respectively).

The Fourier transformed wave equation describing the propagation of distortional waves in a homogeneous medium is[16]

$$\nabla^2\mathbf{H} + \frac{\omega^2}{C_s^2}\mathbf{H} = 0, \tag{1}$$

where $\mathbf{H}$ is the vector potential which is parallel to z, ω is the angular frequency, and $C_s = \sqrt{\mu/\rho}$ is the shear acoustic velocity of the medium (μ and ρ represent the shear modulus and density of the medium). The generic solution to Eq. (1) will be of the form

$$\mathbf{H} = [AJ_0(k_r r) + BY_0(k_r r)]e^{i\xi z}, \tag{2}$$

where A and B are arbitrary constants, k_r and ξ are the projections of the wave vector along the r and z directions, and J_0 and Y_0 are the Bessel function of the first and second kind. The displacement field can be obtained by applying the curl operator to H, thus

$$U_\theta = [CJ_1(k_r r) + DY_1(k_r r)]e^{i\xi z}, \tag{3}$$

where $C = -Ak_r$ and $D = -Bk_r$, the displacement field being tangential. The phase velocity of a guided mode is defined as $C_{\text{ph}} = \omega/\xi$. The nonzero components of the stress tensor are therefore

$$\tau_{r\theta} = \mu\left(\frac{\partial U_\theta}{\partial r} - \frac{U_\theta}{r}\right), \tag{4}$$

$$\tau_{z\theta} = \mu\left(\frac{\partial U_\theta}{\partial z}\right), \tag{5}$$

The displacement field associated with a torsional mode propagating in a bilayered pipe can be expressed according Eq. (3) where the constants C and D are different in the pipe wall and the internal layer, leading to four unknown constants. The unknown can be determined by solving the characteristic equation obtained by imposing suitable boundary conditions at the interface. In particular, we assume a good bond state between the pipe and the coating layer, which means the continuity of the displacement and the stress component $\tau_{r\theta}$ at the interface. Also the zero traction condition is imposed on both the free surfaces of the pipe and the internal layer.[17] Moreover, if the pipe is completely filled, the constant D has to vanish in order to remove the singularity of Y_1 at $r = 0$.

These equations and conditions are implemented in the software DISPERSE. Thus, the solution to the characteristic equation was obtained by using the software DISPERSE.

Theoretical investigations of guided wave propagation in layered structures have been carried out for a long time. Of particular significance here is a theoretical study on the propagation of shear horizontal (SH) waves in a plate coated with a viscoelastic layer, which is a counterpart to our model, but flat rather than cylindrical.[13] In the bilayered plate case, the bilayer modes originate from the interaction between the modes of the free plate and those of the coating layer, if it were rigidly clamped at the bilayer interface. The cutoff frequencies of the SH modes of the bilayer were found to correspond to the cutoff frequencies of the clamped-free coating layer, which only depend on the thickness and bulk shear velocity of the coating layer.[13] They are in direct proportion to the bulk shear velocity and inverse proportion to the thickness of the layer and can be approximated as

$$F_n \simeq \frac{C_s}{4d}(2N - 1), \tag{6}$$

where $N \in \{1, 2, 3 \ldots\}$, C_s and d are the bulk shear velocity and thickness of the coating layer, respectively. As a result, if the cutoff frequencies can be measured and the velocity of the coating is known, the thickness of the layer can be obtained by inverting Eq. (6).

This principle is also valid here for the case of torsional modes in the bilayered pipe. However, Eq. (6) is no longer an exact expression of the cutoff frequencies in cylindrical structures, but it still indicates a qualitative relationship between the cutoff frequencies and the internal layer. Here, one needs to solve the secular equation numerically as explained later.

The energy flow distribution is an important parameter that can help understanding of the scattering of the torsional wave by the internal coating layer. The energy flow of the guided wave is the rate at which it propagates energy along a particular direction. Specifically for torsional modes, the energy flow density in the axial direction (z) at any radial position r can be expressed as[16]

$$I_z(r) = \frac{\omega^2 \mu |U_\theta(r)|^2}{2C_{\text{ph}}}. \tag{7}$$

For the bilayered pipe, total time-averaged axial energy flow in the pipe wall (E_p) and in the coating layer (E_l) can be obtained by integrating Eq. (7) over the pipe wall thickness

$$E_p = \frac{\pi\omega^2}{C_{\text{ph}}}\int_a^b \mu_1 |U_{\theta 1}|^2 r dr, \tag{8}$$

and the coating thickness

$$E_l = \frac{\pi\omega^2}{C_{\text{ph}}}\int_c^a \mu_2 |U_{\theta 2}|^2 r dr. \tag{9}$$

Here, a and b represent the inner and outer radius of the pipe and c is the inner radius of the coating layer. The subscripts 1 and 2 refer to the pipe and the layer, respectively. In order to quantify the distribution of propagating energy between the pipe and the layer, we define a parameter EFR (energy flow ratio) as the ratio of the energy flow in the pipe to that in the layer

$$\text{EFR} = \frac{E_p}{E_l} = \frac{\int_a^b \mu_1 |U_{\theta 1}|^2 r dr}{\int_c^a \mu_2 |U_{\theta 2}|^2 r dr}. \tag{10}$$

The EFR spectra calculated for the modes shown in Fig. 2(a) are given in Fig. 2(b), their significance being addressed in the next section.

III. SCATTERING OF THE T(0,1) MODE AT THE DISCONTINUITY

Consider a torsional wave T(0,1) which is incident from the free end of a pipe (Fig. 1). At the point where a coating layer starts inside the pipe, the incident guided wave is partly

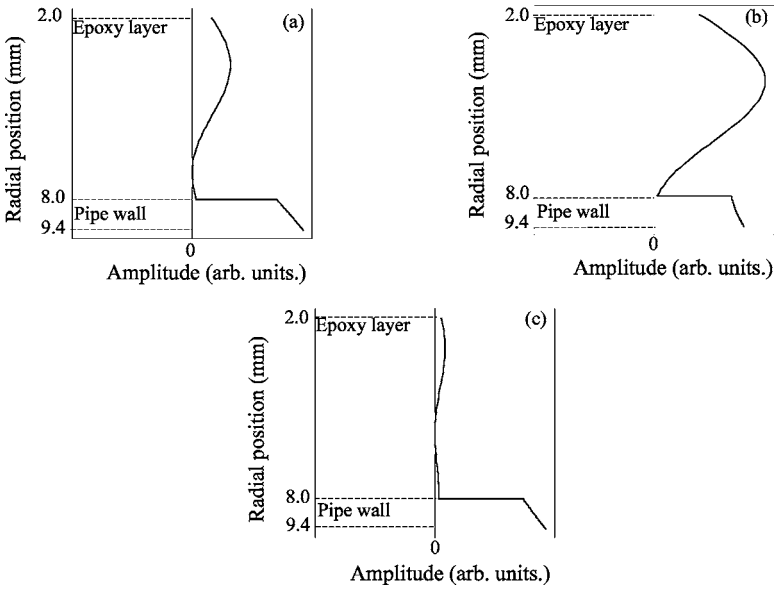


FIG. 3. Energy flow density mode shapes in Fig. 2(b). (a) Mode shape of point A on the T_2 mode. (b) Mode shape of point B on the T_1 mode. (c) Mode shape of point C on the T_2 mode.

reflected into the free pipe mode T(0,1), and also partly transmitted into the multiple torsional modes in the bilayered pipe. It also needs to be mentioned that, since the layer is axially symmetric, no mode conversion into longitudinal or flexural modes occurs. Since both the reflected and the transmitted signals are caused by the presence of the layer, their characteristics should depend on the layer properties and the nature of the guided wave behavior. However, due to the different nature of the reflected and the transmitted signals, different methods will be used to process them.

A. Reflected signal: The remote measurement

In a bulk analysis of a plane wave incident at the interface between two different media, we would analyze the reflection and transmission by studying the change of impedance between the two media. The situation here for guided waves is more complex, but it is the extent to which the modes in the free pipe match those in the bilayered pipe, which determines the strength of the reflection, according to the modal analysis theory.[16] In particular, if the energy flow mode shape of T(0,1) in the free pipe matches well that of a bilayered mode within the pipe wall, the mode is transmitted with little energy being reflected. On the other hand, if the matching is poor, a strong reflection occurs. The extent of the mode matching can be assessed by studying the EFR spectrum.

Let us consider the T(0,1) mode incident at the cutoff frequency of T_2 (Fig. 2). The EFR of T_2 is always larger than that of T_1 [as it can be deduced from Figs. 3(a) and 3(b), which show the normalized energy flow distribution through the thickness of the bilayered pipe], which means that T_2 tends to carry more energy in the pipe wall than T_1. Therefore, it is expected that T(0,1) in the free pipe will transmit more energy into T_2 rather than T_1. However, it can be observed that as the frequency increases, the EFR of T_2 increases up to a maximum value labeled C in Fig. 2(b), where the amount of energy traveling in the pipe wall is maximum [see Fig. 3(c)]. As a result, as the frequency increases from the cutoff frequency of T_2 to C, T(0,1) tends to transmit more energy into T_2; conversely the reflected energy tends to decrease by the energy conservation law. If the frequency increases further, the EFR of T_2 starts decreasing together with the transmitted energy from T(0,1), since the matching of the energy flow mode shape deteriorates. When the frequency reaches the cutoff of T_3 a transition occurs and the energy of T(0,1) starts to be mostly transmitted into T_3, since it will have a larger EFR than T_1 and T_2 afterwards. The same phenomenon will repeat periodically as the frequency increases. It can be concluded that the frequencies where maximum of EFR occurs will approximate the minimum of the reflection coefficient, whereas the maximum of the reflection coefficient should occur at the frequencies close to the cutoff frequencies of the bilayered pipe modes.

To validate these predictions, a finite element (FE) modeling simulation was performed to study the scattering of torsional waves in such a structure, using the FE software FINEL, developed at Imperial College.[18] The details of the models are outlined in the Appendix. The material properties used for the FE modeling are the same as those used for the previous calculation and are listed in Table I.

As illustrated in Fig. 1, the monitoring point of the reflected signal is set at one point in the free pipe region indicated as the remote measurement. The time domain signal from the FE simulation at this monitoring point is shown in Fig. 4. The time trace shows the incident wave on its way to the bilayered pipe region, and the reflection from the entry point of the epoxy layer; the long temporal duration of the reflected signal indicates the frequency dependence of the reflection coefficient spectrum. The reflection coefficient spectrum, calculated by dividing the magnitude of Fourier transform of the reflected signal by that of the incident signal, is plotted in Fig. 5, which exhibits periodic peaks and troughs. The roughness of the curves is due to the window effect on the time domain signal. The cutoff frequencies of the bilayered pipe modes (T_2, T_3, T_4) are marked as F_2, F_3, F_4 in the frequency axis for comparison. It can be seen that the peaks of the reflection coefficient spectrum occur almost exactly at the cutoff frequencies of the bilayered

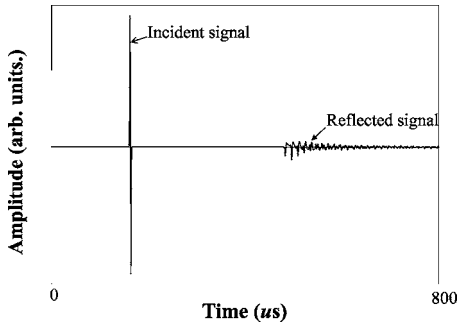


FIG. 4. Time trace, simulated by FE modeling, of the scattering of T(0,1) in an aluminum pipe partially coated with a 6-mm-thick epoxy layer inside, showing the incident and reflected signals received at the remote measurement position indicated in Fig. 1.

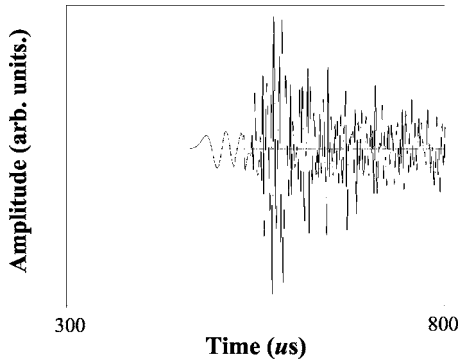


FIG. 6. Time trace, simulated by FE modeling, of the transmitted signal for T(0,1) in the pipe partially coated with 6-mm-thick epoxy layer inside.

modes, which confirms our above-presented analysis. Since these cutoff frequencies are, for an assigned pipe, uniquely determined by the epoxy layer thickness and bulk shear velocity, the shear acoustic properties or thickness of the coating layer can be characterized by measuring the peak positions of the reflection coefficient spectrum.

B. Transmitted signal: The local measurement

Due to the coexistence of multiple modes and their dispersive nature, the transmitted signal received at a location in the bilayered pipe (Fig. 1) is too complicated to be interpreted directly, as shown by the simulated transmitted signal in Fig. 6. Some researchers have resorted to advanced signal processing techniques to separate multiple modes and have thus tried to measure the high order modes in a filled pipe structure. Vollmann *et al.* showed that it is possible to measure the high order modes by means of a spectrum estimation method to operate on multiple equally spaced wave forms which are sampled along the filled pipe.[9,10] Unfortunately, the need for exact, spatially sampled data makes this technique very time consuming. Kwun *et al.* experimentally investigated the dispersion of the longitudinal wave modes for a pipe that was completely filled with a liquid.[19] By using the spectrogram of the short-time Fourier transform (STFT), they observed that the dispersion curve for the longitudinal mode was divided into approximately equally spaced regions separated by cutoff-type behavior. However, it was also observed by the authors that due to the poor resolution, it is somewhat inaccurate to determine the cutoff frequencies from the spectrogram. Recently, time-frequency representations such as the spectrogram, the scalogram, and the Wigner-Ville method have shown promise for the interpretation of such signals.[20] The time-frequency representations analyze signals by quantitatively resolving changes in frequency content, as a function of time. These representations suffer from the uncertainty principle, making it impossible to simultaneously have perfect resolution in both time and frequency. However, researchers at Georgia Institute of Technology have shown that the reassigned spectrogram (the reassigned energy density spectrum of the STFT) is capable of distinguishing multiple Lamb wave modes from only a single signal, and giving excellent representation of its dispersion curves.[20] The work by this group has used the reassigned spectrogram to represent the dispersion curves of free plates and plates with cracks.[20,21]

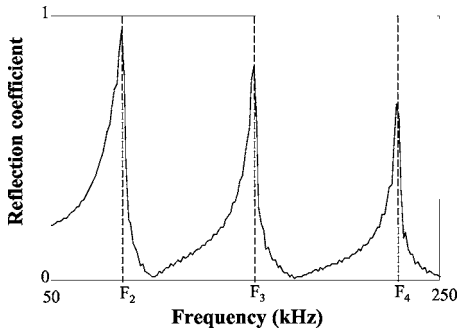


FIG. 5. Reflection coefficient amplitude spectrum of T(0,1) in the aluminum pipe incident at the section where it is partially coated inside with a 6-mm-thick epoxy layer (FE modeling). F_2, F_3, F_4 are cut-off frequencies of the bilayered pipe.

In this paper, we apply the latter technique to the local measurement of the transmitted signal in the bilayered pipe region. By applying a wide band pulse signal for excitation, all the bilayered modes in a frequency range of interest can be generated. The receiver is positioned in the bilayered region to monitor the transmitted signal shown in Fig. 6. Only one measured time-domain signal is needed. Then, to transform the signal into the time-frequency domain, instead of considering the Fourier transform of the entire signal at once, we use STFT through a reassigned procedure, which effectively refines the time-frequency resolution.[20,22]

The contour plot of the square root of the reassigned spectrogram of the signal shown in Fig. 6 is given in Fig. 7. It reveals for each mode and at each frequency, the arrival time of the transmitted signal which is determined by its group velocity. As we can see here, the reassigned spectrogram provides a very clear representation of the individual modes of the bilayered pipe. The cutoff frequencies (marked as $F_2, F_3, F_4, \ldots$) of the bilayered modes can be estimated from frequencies which correspond to the largest arrival time as the group velocity tends to zero at the cutoff frequencies. The accuracy of the dispersion curves represented from the reassigned spectrogram can be assessed by comparison with the analytical results from DISPERSE, shown in dashed curves. DISPERSE provides the group velocity of each mode, which can then be converted into arrival time for a certain propagation distance. The comparison shows excellent

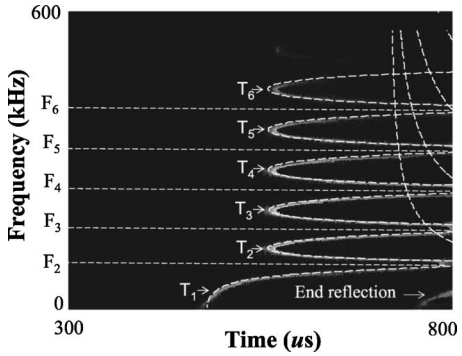


FIG. 7. Comparison of reassigned spectrogram analysis and analytical calculation (white dashed curves) for the transmitted signal in the bilayer part of a pipe coated inside with a 6-mm-thick epoxy layer (FE signal).

agreement. A little reflected signal from the end of the bilayered pipe can also be observed from the reassigned spectrogram, at the right bottom corner of Fig. 7.

IV. EXPERIMENTS

Experiments have also been carried out to validate the analysis and FE calculations. The general setup for the experiments is shown in Fig. 8. Two samples of aluminum pipes (16.5 mm inner diameter and 1.4 mm wall thickness) with partial coating inside were manufactured. The first was a 1.25-m-long pipe, 0.55 m of which was coated inside with a 6-mm thickness epoxy resin layer to represent the model studied in this paper. Due to application difficulty, the epoxy layer built inside the pipe was not perfectly symmetrical all along its length. The second sample consisted of the same pipe, but the epoxy resin completely filled the pipe for 0.3 m from one end. This sample aimed to simulate the case of ultimate blockage inside the pipe. The epoxy was cured inside the pipe so as to meet the theoretical assumption of good bond condition between the layer and the pipe. The epoxy resin employed is a commercial adhesive, Araldite 2020, whose acoustic properties were determined by the authors using conventional ultrasonic time of flight method[23] and are listed in Table II.

The torsional mode T(0,1) was excited by four pieces of shear-vibration piezoelectric (PZT) elements. These four piezoelectric elements were mounted axisymmetrically on the external surface of the pipe to generate the torsional vibration as shown in Fig. 8. The piezoelectric elements were

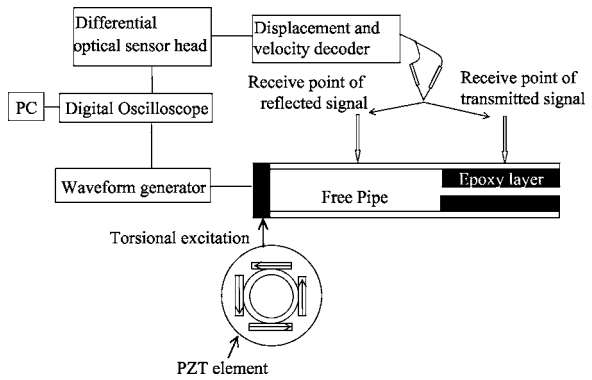


FIG. 8. Schematic of the experimental setup.

TABLE II. Material properties of the aluminum pipe and epoxy layer used for FE calculations to compare with the remote measurement results. The bulk shear velocity of epoxy is determined by an independent measurement.

	Bulk longitudinal velocity (m/s)	Bulk shear velocity (m/s)	Density (kg/m^3)
Aluminum	6320	3130	2700
Epoxy	2610	1010	1170

excited by a Hanning windowed toneburst generated by a custom-made wave-form generator and amplifier. T(0,1) is the only axisymmetric propagating mode in the free pipe with significant circumferential displacement below the cutoff frequency of T(0,2), which will occur at 1.16 MHz for the tested pipe. Therefore, only T(0,1) will be generated below 1.16 MHz, provided the excitation is perfectly axially symmetric. A single element attached to the pipe surface would excite both torsional modes (axisymmetrical modes) and flexural modes (nonaxisymmetrical modes). However, if a series of elements are attached axisymmetrically around the pipe to form a ring, it should be possible to suppress the flexural modes to a certain extent.[24] The use of four balanced transducers is sufficient to suppress the excitation of only two asymmetric flexural modes F(1,1) and F(1,2) below 110 kHz. The torsional vibration was detected by a laser interferometer (sensor head: Polytec OFV 512, controller: Polytec OFV 3001) operating in differential mode. The tangential displacement was measured by focusing the two beams at 30° with respect to the radial direction and by orienting the beams in the plane perpendicular to the axis of the pipe.

Different procedures for the excitation and reception of torsional modes were used according to the signals to be measured. To measure the reflected signal, a five cycle excitation signal in a Hanning window was used to narrow the frequency band into the range in which the current setup could guarantee only the T(0,1) mode being excited without flexural modes. The receiver was set at a position halfway between the free end of the pipe and the start point of the bilayered pipe, so both the incident and the reflected signals would be received. To further remove the influence of the flexural mode, measurements were carried out at four equally spaced points around the circumference of the pipe, but at the same axial location. By adding the four sampled wave forms, we can obtain both the incident and reflected signal only containing a single T(0,1) mode, since the flexural modes which have a sinusoidal displacement distributed along the pipe circumference are canceled by circumferential averaging.

For the transmitted signal, we aimed to excite multiple axially symmetric modes in the bilayered pipe. A two cycle excitation signal with central frequency 150 kHz in a Hanning window was used, which covers a wide frequency band from 40 up to 300 kHz. Since four PZT elements are not enough to suppress flexural modes of order higher than two, the excitation of T(0,1) is accompanied by weakly excited higher order flexural modes. The transmitted signal was received by the laser at a position in the bilayered region. Only one signal wave form is needed for the signal processing.

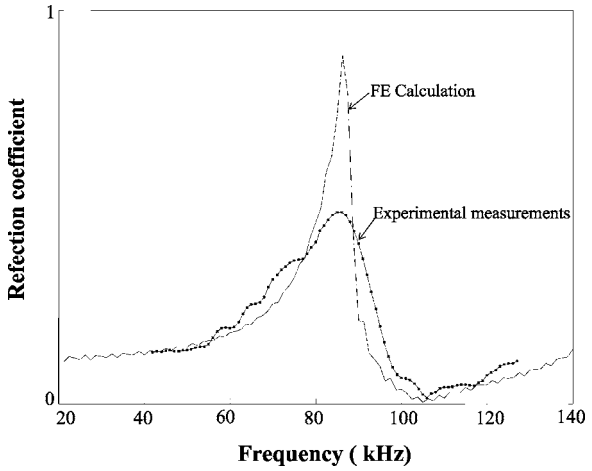


FIG. 9. Reflection coefficient amplitude of the T(0,1) mode from the aluminum pipe coated inside with a 6-mm-thick epoxy layer, obtained by experimental measurement (line with black squares) and FE calculation (black solid line).

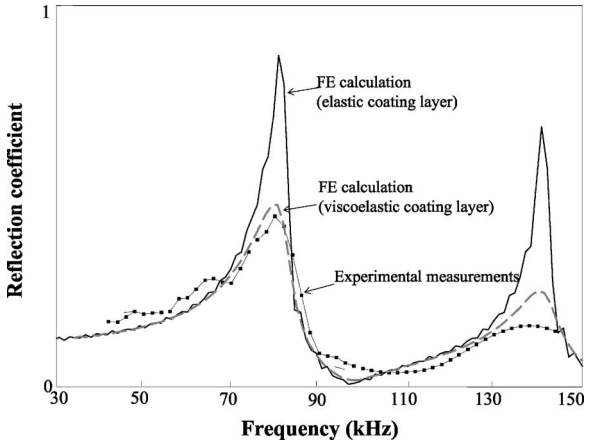


FIG. 10. Reflection coefficient amplitude of the T(0,1) mode from the aluminum pipe completely filled with epoxy over part of its length, obtained by experimental measurement (line with black squares), FE calculation assuming an elastic coating (black solid line), and FE calculation assuming a viscoelastic coating (gray dashed line).

A. Reflected coefficient spectrum: Experimental results

The reflection coefficient is calculated by dividing the frequency spectrum of the measured reflected signal by that of the incident signal. The reflection coefficient amplitude spectrum obtained from the pipe partially coated on the inside surface with a 6-mm thickness epoxy layer is plotted in Fig. 9. The measured curve is obtained as a superposition of measurements performed at different central frequencies, ranging from 50 to 110 kHz. For comparison, the results of a FE calculation performed with the material properties given in Table II are also shown. We can see that the measured reflection coefficient spectrum displays the characteristic peak and trough, which follow the trend of the FE calculation well. The most important feature of the reflection coefficient spectrum is that the first measured peak occurs at the position of the FE prediction (86 kHz), which is very close to the predicted cutoff frequency (83 kHz) of the T_2 mode of the corresponding bilayered pipe.

However, it is clear that the magnitude of the measured reflection coefficient spectrum is not as sharp as that of the FE prediction, especially at the frequencies around the peak position. The disagreement is largely due to the material damping of the epoxy. So far the FE modeling considered the epoxy to be an elastic material; however, the epoxy is known to behave more like a viscoelastic material. Since the reflection peak is due to the presence of standing waves in the coating layer, material absorption weakens the resonance phenomenon, so resulting in a lower reflection coefficient.[13]

Figure 10 shows the reflection coefficient spectrum obtained for the pipe sample, completely filled with epoxy over part of its length. The measured reflection spectrum shows very similar characteristics to those shown in Fig. 9, but the position of the first peak position shifts to a slightly lower frequency (80.5 kHz) due to the increase of thickness of the filling epoxy. For this configuration, Professor M. Castaings at the University of Bordeaux I has provided FE simulations which incorporate the material damping of epoxy assuming a damping coefficient of 0.13 Np/wavelength given in the literature.[25] The details of this technique have been submitted for publication in a separate paper by Castaings and Bacon.[26] For reference, the FE modeling without considering the material damping of the epoxy is also provided. These results confirm that material damping reduces the magnitude of the reflection coefficient at the cutoff frequencies, and the agreement between the FE prediction and the experimental results is very good.

B. Reassigned spectrogram analysis of the transmitted signal: Experimental results

The transmitted signal was measured in the bilayered region of the pipe and only one time domain signal was needed. The reassigned spectrogram was applied to the measured transmitted signals to represent the group velocity dispersion curves of the torsional modes in the bilayered pipe. The signal was first measured on the sample of pipe locally coated with 6-mm thickness layer, at a bilayer pipe position 0.15 m away from where the coating layer starts in the pipe. The contour plot of the reassigned spectrogram analysis is shown in Fig. 11, where the dashed curves represent the analytical curves calculated with DISPERSE using the material properties obtained from the remote measurement of the same sample (listed in Table II). While the T_2 and the T_4

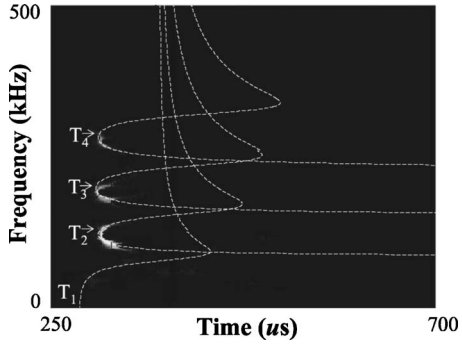


FIG. 11. Reassigned spectrogram analysis of measured transmitted signal from the aluminum pipe coated inside with a 6-mm-thick epoxy layer, and analytical calculation (white dashed lines) by DISPERSE.

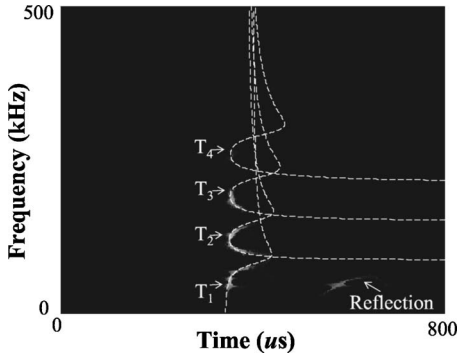


FIG. 12. Reassigned spectrogram analysis of measured transmitted signal and best fit curves (white dashed lines) calculated by DISPERSE for the aluminum pipe locally filled with epoxy.

modes agree well with the analytical curves, the dispersion curve of the T_3 mode is influenced by some additional modes, which were not present in the FE simulations (Fig. 7). These additional modes are flexural bilayered pipe modes (asymmetrical modes) caused by the asymmetrical thickness of the coating layer inside the pipe. Since the dispersion curves of some flexural modes are close to that of the T_3 mode (these are not shown for clarity in Fig. 7), they cannot be well separated in either time domain or frequency domain, in which case the reassigned spectrogram is incapable of distinguishing them.[20] Also, the mode structure of the T_1 mode is destroyed by the asymmetrical layer, and little energy is transmitted into this mode.

The transmitted signal was also measured on the sample of pipe locally filled with epoxy, at a position 0.09 m away from where the epoxy filling starts. The result of the reassigned spectrogram analysis is shown in Fig. 12. In this case the layer is perfectly symmetrical, so the dispersion curves of the first four torsional modes of the bilayered pipe and their cutoff frequencies can easily be identified and no flexural modes are present. The white dashed curves are the analytical curves calculated from DISPERSE. The mode arriving later is the reflection of the transmitted signal from the end of the bilayered pipe.

It should be observed that the cutoff frequencies cannot be estimated as in the case of the FE prediction shown in Fig. 7. The lack of definition is due to the large attenuation of torsional modes at their cutoff frequencies caused by the material damping of the epoxy filling.[12] Moreover, no dispersion curves can be extracted above 220 kHz, since material damping increases with frequency. A more effective method is to best fit the analytical dispersion curves obtained from DISPERSE to the measured ones, with respect to the thickness or the bulk shear velocity of the coating layer (dashed curves). The measured dispersions of high order modes are highly sensitive to the change of these two parameters of the coating. To show this, we estimated the thickness of the filling epoxy through the best fit procedure, using the bulk shear velocity of epoxy obtained from the remote measurement (listed in Table II). The thickness of the epoxy filling obtained in this way is 8.2 ± 0.1 mm. This is within 2% of the pipe inner radii (8.25 mm).

V. DISCUSSION

The motivation for this work is to provide an insight into the possibility of detecting and characterizing sludges and blockages inside pipes by guided ultrasonic waves. The presented principles and the measurement techniques which have been demonstrated for an idealized bilayered pipe may also be useful for various related applications. For example, the local measurement technique could be beneficial to measure the properties of fluids inside pipes as is needed in many online processing technologies and this is the topic of a separate paper.[27] For the sludge detection case, there are some issues which need to be discussed as they may affect these techniques in different ways.

First, the irregular shape of contents will certainly complicate the model studied in this paper. The reflected signal is caused by the acoustic impedance change at the entry part of the coating layer. So it should be expected that the remote measurement will be subject to the shape of the entry part of the coating layer. In particular, the more gentle the transition between the free pipe and the coated pipe, the lower the reflection coefficient will be. On the other hand, this will benefit the local measurement, since more energy is transmitted into the bilayered pipe. The variable thickness of the coating layer will affect the local measurement, since the guided wave propagating in local parts of the bilayered pipe region with different thicknesses will follow different dispersion changes. Depending on the thickness of each local part of the bilayered pipe and the corresponding length, the local measurement may represent an overall dispersion change of the signal measured in the bilayered pipe. Also, if the coating layer is asymmetrical, which is most likely for a sludge layer to form in practice, some flexural modes (asymmetrical modes) will appear in addition to the expected torsional modes (symmetrical modes). Some of these effects have already been observed in the local measurement of the sample of pipe with a slightly asymmetrical coating layer (Fig. 11). A study is proceeding to investigate the significance of these effects.

Second, the work described in this paper assumes that the coating layer is properly bonded to the pipe, however, there may be such circumstances in practice in which the bond may be poor, or may even be a slip contact. The deterioration of bond quality gives rise to the decreasing of the mode coupling between the free pipe and the coating layer. Nevertheless, on the positive side, an initial experimental study by the authors has found that the presence of fluid may be beneficial, when using the longitudinal mode with nonnegligible radial displacement. This has been observed experimentally on a sample with a poor bond condition by measuring the reflection coefficient for the cases in which the contact was dry or when it is wet. The presence of water enables the transfer of energy from the pipe to the layer thanks to pressure waves excited by the radial motion of the pipe wall and transmitted into the layer.

Lastly, high material damping of a coating layer can compromise the applicability of the remote measurement, since the sharp reflection peaks at the cutoff frequencies tends to disappear as the material damping increases.

VI. CONCLUSION

The work in this paper has studied the scattering of the fundamental torsional mode by an axisymmetric coating layer inside a pipe. One of the stated objectives was to develop some general methods to detect and characterize the properties of a coating layer inside a pipe using torsional modes. The presence of the layer increases the number of torsional modes, which leads to new cutoff frequencies that depend on the layer thickness and its bulk shear velocity. Therefore, by measuring the cutoff frequencies, it is possible to determine the thickness of the coating layer, if its bulk shear velocity is known, and vice versa. Two methods have been developed.

The first method measures the spectrum of the reflection coefficient of the fundamental torsional mode incident on the coated part of the pipe. FE calculations have shown that the reflection coefficient spectrum exhibits periodic peaks, which just occur at the cutoff frequencies of the bilayered pipe modes. The phenomenon has been found to be due to the fact that at the cutoff frequencies of the bilayered pipe modes, the energy tends to flow in the coating primarily, while little remains in the pipe wall. This causes a large mismatch between the modes of the free pipe and those of the bilayered pipe region, so resulting in a strong reflection. This phenomenon is confirmed by experimental measurements, which showed a very good agreement with predictions. This method has the advantage that the measurement could be carried out remotely from the coated region of the pipe and the properties of the coating layer can be obtained without necessarily knowing its location.

A second method is based on the measurement of the torsional waves transmitted to the bilayered pipe region. A time-frequency technique called the reassigned spectrogram has been employed to analyze signals containing multiple dispersive modes. The reassigned spectrogram has been applied to both FE simulated signals and measured ones and proved to be very effective in extracting the group velocity dispersion curves of the bilayered pipe modes. The measurement is very efficient, since only a single time domain signal is needed.

ACKNOWLEDGMENTS

The authors are grateful to the Engineering and Physical Science Research Council (EPSRC) for funding this work. They are also grateful to Professor M. Castaings of the University of Bordeaux I, France, for performing the FE simulations of the case in which the damping of the epoxy material was taken into account.

APPENDIX

The FE modeling studies in this paper were carried out using the software FINEL, developed at Imperial College.[18] FINEL performs efficient modeling of elastic wave propagation using a time-marching procedure, and so produces simulations of the signals which would be seen experimentally. The model is axially symmetric, representing a radial-axial section through the pipe and thus enabling two-dimensional axially symmetric elements to be used. The aluminum pipe with 16 mm inner diameter and 1 mm wall thickness is 1.75 m long, 0.75 m of which, from the far end, is coated with 6-mm thickness epoxy layer inside (the configuration is illustrated in Fig. 1). The material properties of the pipe and epoxy layer remain the same as those used for the calculation of the dispersion curves shown in Fig. 2 and are listed in Table I. The elements were defined to be perfectly square in shape with the size of 0.2 mm. Thus, there are 5 elements through the thickness of the pipe and 30 elements through the thickness of the epoxy layer. A one-cycle pulse excitation of 150 kHz central frequency is applied at the free end of the pipe. Because the excitation has no significant energy above the cutoff frequency of T(0,2), the only propagating mode excited by this configuration is the T(0,1) mode. One has to make sure that the reflection from the entry point is not disturbed by any other reflections from the end of the pipe by choosing a sufficiently long coated length. The results of the models are obtained by monitoring tangential displacements at the outside surface of the pipe. Two monitoring points are set for measuring the reflected signal and transmitted signal, respectively: the monitoring point of the remote measurement is set at the half-way point between the free end of the pipe and the entry point of the epoxy layer, so that both the incident and the reflected signals could be measured; the monitoring point of the local measurement is located 0.4 m away after the layer starts in the pipe.

[1] D. N. Alleyne, B. Pavlakovic, M. J. S. Lowe, and P. Cawley, "Rapid, long range inspection of chemical plant pipework using guided waves," Insight **43**, 93–96,101 (2001).

[2] D. N. Alleyne and P. Cawley, "Long range propagation of Lamb waves in chemical plant pipework," Mater. Eval. **55**, 504–508 (1997).

[3] R. Kumar, "Dispersion of axially symmetric waves in empty and fluid-filled cylindrical shells," Acustica **27**, 317–329 (1972).

[4] L. D. Lafleur and F. D. Shields, "Low-frequency propagation modes in a liquid-filled elastic tube waveguide," J. Acoust. Soc. Am. **97**, 1435–1445 (1995).

[5] B. K. Sinha, T. J. Plona, S. Kostek, and S.-K. Chang, "Axisymmetric wave propagation in fluid-loaded cylindrical shells. I. Theory," J. Acoust. Soc. Am. **92**, 1132–1143 (1992).

[6] T. J. Plona, B. K. Sinha, S. Kostek, and S.-K. Chang, "Axisymmetric wave propagation in fluid-loaded cylindrical shells. II. Theory versus experiments," J. Acoust. Soc. Am. **92**, 1144–1155 (1992).

[7] L. Elvira-Segura, "Acoustic wave dispersion in a cylindrical elastic tube filled with a viscous liquid," Ultrasonics **37**, 537–547 (2000).

[8] C. Aristegui, M. J. Lowe, and P. Cawley, "Guided waves in fluid-filled pipes surrounded by different fluids," Ultrasonics **39**, 367–375 (2001).

[9] J. Vollmann and J. Dual, "High-resolution analysis of the complex wave spectrum in a cylindrical shell containing a viscoelastic medium. I. Theory and numerical results," J. Acoust. Soc. Am. **102**, 896–908 (1997).

[10] J. Vollmann, R. Breu, and J. Dual, "High-resolution analysis of the complex wave spectrum in a cylindrical shell containing a viscoelastic medium. II. Experimental results versus theory," J. Acoust. Soc. Am. **102**, 909–920 (1997).

[11] F. Simonetti and P. Cawley, "A guided wave technique for the characterisation of highly attenuative viscoelastic materials," J. Acoust. Soc. Am. **114**, 158–165 (2003).

[12] F. Simonetti, "Lamb wave propagation in elastic plates coated with viscoelastic materials," J. Acoust. Soc. Am. **115**, 2041–2053 (2004).

[13] F. Simonetti and P. Cawley, "On the nature of shear horizontal wave propagation in elastic plates coated with viscoelastic materials," Proc. R. Soc. London, Ser. A **460**, 2197–2221 (2004).

[14] B. N. Pavlakovic, M. J. S. Lowe, D. N. Alleyne, and P. Cawley, "DISPERSE: A general purpose program for creating dispersion curves," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), Vol. **16**, pp. 185–192.

[15] B. N. Pavlakovic and M. J. S. Lowe, "A general purpose approach to

calculating the longitudinal and flexural modes of multi-layered, embedded, transversely isotropic cylinders," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (Plenum, New York, 1999), Vol. **18**, pp. 239–246.
[16]B. A. Auld, *Acoustic Fields and Waves in Solids*, 2nd ed. (Krieger, Malabar, FL, 1990), Vol. **2**.
[17]M. J. S. Lowe, "Matrix techniques for modelling ultrasonic waves in multilayered media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 525–542 (1995).
[18]D. Hitchings, "FE77 User Manual," Tech. Rept., Imperial College, 1994.
[19]H. Kwun, K. Bartels, and C. Dynes, "Dispersion of longitudinal waves propagating in liquid-filled cylindrical shells," J. Acoust. Soc. Am. **105**, 2601–2611 (1999).
[20]M. Niethammer, L. J. Jacobs, J. Qu, and J. Jarzynski, "Time-frequency representations of Lamb waves," J. Acoust. Soc. Am. **109**, 1841–1847 (2001).
[21]R. Benz, M. Niethammer, S. Hurlebaus, and L. J. Jacobs, "Localization of notches with Lamb waves," J. Acoust. Soc. Am. **114**, 677–685 (2003).
[22]M. Niethammer, L. J. Jacobs, J. Qu, and J. Jarzynski, "Time-frequency representation of Lamb waves using the reassigned spectrogram," J. Acoust. Soc. Am. **107**, L19–L24 (2000).
[23]A. S. Birks, R. E. Green, and P. Mclntire, *Nondestructive Testing Handbook*, 2nd ed. (American Society of Nondestructive Testing, 1991), Vol. **7**.
[24]D. N. Alleyne and P. Cawley, "The excitation of Lamb waves in pipes using dry-coupled piezoelectric transducers," J. Nondestruct. Eval. **15**, 11–20 (1996).
[25]R. J. Freemantle and R. E. Challis, "Combined compression and shear wave ultrasonic measurements on curing adhesive," Meas. Sci. Technol. **9**, 1291–1302 (1998).
[26]M. Castaings and C. Bacon, "Finite element modelling of torsional wave modes along pipes with absorbing materials," J. Acoust. Soc. Am. **119**, 3741–3751 (2006).
[27]J. Ma, M. J. S. Lowe, and F. Simonetti, "Measurement of the properties of fluids inside pipes using guided longitudinal waves," IEEE Trans. Ultrason. Ferroelectr. Freq. Control (submitted).

Resonance frequency shift saturation in land mine burial simulation experiments

W. C. Kirkpatrick Alberts II,[a] James M. Sabatier, and Roger Waxler
National Center for Physical Acoustics, University of Mississippi, University, Mississippi, 38677

(Received 27 March 2006; revised 27 July 2006; accepted 28 July 2006)

In the field of acoustic land mine detection, the mechanical resonant behavior of land mines has proven to be a feature key to this detection method's success. Land mines are often interred at depths ranging from several millimeters to tens of centimeters. Thus, it is necessary to understand the effect of burial on the land mine's resonances. Theoretical works on burial's effects on land mine resonance frequency have reported varying results. In this work, burial simulation experiments in both water and sand using a land mine and a simulant have been performed to study the effect "burial" has on the resonance frequencies of a land mine. Saturation of the resonance frequency shift, i.e., the resonance frequency becomes constant after shifting due to added mass, is observed in both burial simulations. The experimental observations are qualitatively explained by recent theoretical work on the subject. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2338294]

PACS number(s): 43.20.Tb, 43.40.Dx, 43.40.Rj [KA] Pages: 1881–1886

I. INTRODUCTION

Acoustic methods of land mine detection have recently emerged as viable techniques for detecting land mines buried in soil.[1–3] In some of these experiments,[1,4,5] broadband noise, typically from 80 to 300 Hz, is used to insonify the soil surface. The incident sound excites the vibration of the soil matrix and any land mines buried in the insonified area. The induced soil velocity amplitude is measured with a scanning laser Doppler vibrometer (LDV). A buried land mine is detected when an on/off-target velocity contrast is observed in a localized area of the interrogated region through several user-defined frequency bands.[1,4,5] It has been suggested that the observed velocity contrast is due to the resonance behavior of the buried land mine.[2,3]

In order to operate, many land mines are designed to mechanically deform in response to an applied force. This deformation typically takes place in some type of plate supported by a spring, which transfers the force on the plate to the detonator.[6] The spring can be a simple air cavity or a mechanical device directly connected to the plate. The resulting mechanical system can exhibit resonances that increase the on/off target velocity contrast as mentioned above. Recent experimental observations[7–9] and theoretical efforts[3,10,11] have demonstrated and attempted to model, respectively, a complicated dependence of the land mine resonance on burial depth in the simplest case, that of a uniform soil.

Experimental efforts by Donskoy *et al.*[3] and Zagrai *et al.*[7,10] have shown a resonance depth dependence that cannot be described by simple mass loading. Their observations, as well as those of Sabatier and Korman,[8,9] show, with greater burial depth, a tendency for the land mine to exhibit a decrease in resonance frequency at shallow depths followed by an increase in resonance frequency at larger depths.

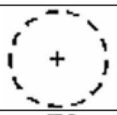
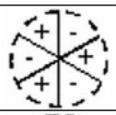
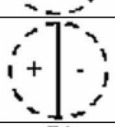
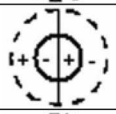
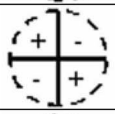
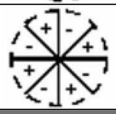
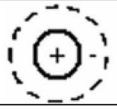
To describe the soil-land mine system, one-dimensional (1D) lumped element models have been proposed by Donskoy *et al.*[3] and Zagrai *et al.*[7,10] Models of this type are well suited to systems where the insonifying wavelengths are large with respect to the dimensions of the system and, therefore, fit well with acoustic land mine detection.[3,7] In its simplest form, this model includes an effective mass, spring, and dashpot representing the land mine. Coupled directly to the land mine mass are a spring and a dashpot representing the compressional stiffness and resistance of the soil. The mass of the soil column above the land mine rests above the soil compressional elements and couples to ground through a spring and a dashpot representing the shear stiffness and resistance of the soil. This model predicts a monotonic decrease in resonance frequency with increasing burial depth.

A more recent modification by Zagrai *et al.* in Sec. III b of Ref. 10 allows the lumped element model to predict the observed upward trend in resonance frequency at large depths. By considering the soil above the land mine to act as an effective plate, the authors show a cubic depth dependence for the effective shear stiffness of the soil. This cubic depth dependence allows the land mine resonance to increase in frequency at increasing burial depth. The modifications to the lumped element model allow the model to closely fit their observations.

A recent theoretical effort by Velea *et al.*[11] gives significantly different predictions for the land mine's resonance shift with depth. In this three-dimensional (3D) model, the soil is modeled as an effective fluid. The scattered field around the land mine is calculated to determine the particle velocity of the fluid at the surface. As the depth of the land mine is increased, this model predicts a cessation of the resonance frequency shift at shallow depths; the resonance frequency becomes constant at shallow burial depth. This effect will be referred to here as saturation. This saturation can be explained by the land mine's beginning to see the fluid layer above the pressure plate as an infinite medium. At sufficient

[a]Author to whom correspondence should be addressed; electronic mail: kirk.alberts@arl.army.mil

TABLE I. Flush-buried mode shapes and frequencies measured on the VS 1.6 in both water and sand.

Mode (m,n)	Shape	Freq. (Hz) Water	Freq. (Hz) Sand	Mode (m,n)	Shape	Freq. (Hz) Water	Freq. (Hz) Sand
(0,0)		281	351	(3,0)		770	808
(1,0)		120	426	(1,1)		925	958
(2,0)		390	570	(4,0)			
(0,1)		485	660				

burial depth, the land mine's own radiation pattern into the fluid begins to outweigh the effect of greater mass loading with increasing depth.

In this paper, burial simulation experiments will be described. The simulations were performed in both dry, unconsolidated sand and water to observe resonance frequency behavior in both shear and nonshear supporting media. An Italian VS 1.6, plastic, anti-tank land mine was used during burial simulation in water and in sand. Burial simulation in sand also used an acoustic land mine simulant. The results of these experiments will be shown with an attempt to explain those results by utilizing the models described above.

II. EXPERIMENTAL PROCEDURES AND RESULTS

As part of a study conducted on the Italian VS 1.6 land mine, burial of the land mine was simulated in both water and sand. An experimental modal analysis was performed prior to the burial simulations in order to characterize the modal behavior of the VS 1.6. The modal response of the land mine was excited using an electromechanical shaker and the surface velocity of the pressure plate was measured using a LDV. The results of the modal analysis of the VS 1.6 flush buried in water and in sand are shown in Table I. In Table I, the modes are grouped according to the integers m and n, which refer to the number of nodal diameters and the number of nodal circles, respectively. The reader will note that, in the water results, the frequency of the first symmetric mode is higher than that of the first asymmetric mode. This and the subsequent reordering in sand have been presented in Ref. 12, will be addressed in detail in a companion paper, and are not of significance to this work. In this paper, modes will be referred to by a parenthetical containing the integers m and n separated by a comma; for example, (2,1) refers to a mode with two nodal diameters and one nodal circle. Similar modes have been observed by Zagrai *et al.*[7,10] in land mines and simulants. In both burial simulation experiments, the dependence of the modal frequencies on increasing burial depth was tracked to as great a depth as possible. Burial simulation in water will be discussed first.

A. VS 1.6 water immersion

Before beginning the experiment, the land mine was allowed to soak in water for a period of five days in an attempt to saturate the plastic. This was done since the land mine's pressure plate was believed to be constructed of a type of Nylon, which softens with higher moisture content.[13,14] After soaking, the land mine was placed in an open-topped 25 cm acrylic cube. Water was then added to the cube until it reached a level even with the top of the land mine's pressure plate, a depth of 9.1 cm. This is referred to as the flush-buried or flush condition. The vibration of the pressure plate was then excited by an electromechanical shaker in light contact with the pressure plate via a 3-mm-diam. aluminum rod. A single cycle of a 1 kHz signal was sent to the shaker from a function generator every 2 s. Simultaneously, a LDV was used to measure the velocity of the plate surface in a frequency range from 0 to 1.6 kHz with a spatial resolution of 1.22 points/cm^2. This experimental configuration is depicted in Fig. 1. Mode frequencies of the pressure plate were

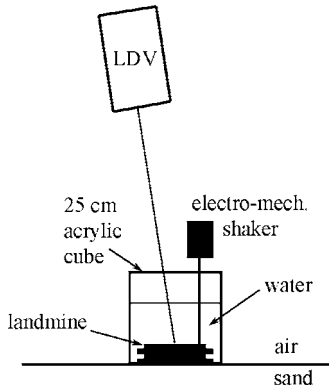


FIG. 1. Experimental configuration used during burial simulation in water.

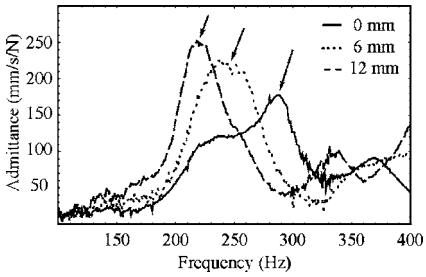


FIG. 2. Mechanical admittance of the (0,0) mode of the VS 1.6 with increasing depth in water, measured at the center of the pressure plate.

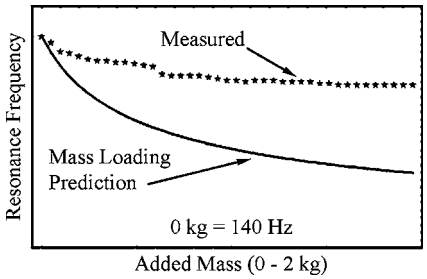


FIG. 4. Depth dependence of the (1,0) mode of the VS 1.6 as measured (dotted) and treating the water layers as uniform additions of mass (solid).

recorded as water was added above the pressure plate in 3 mm increments, corresponding to 48.8 g of water uniformly distributed across the surface of the plate, until a depth of 12 cm was reached. With the exception of the (4,0) mode, all modes persisted as the depth was increased. The (4,0) mode was not excited well by the shaker and the mode's absence was not deemed significant. It should be noted that the LDV was positioned so that it was several degrees away from the water surface normal. This attempts to minimize the water's effects on the vibration measurement of the land mine, which tends to add some uncertainty in the measured velocity without affecting the measured frequency.[15] The modes of the plate were measured with the LDV at each water increment.

Results of the burial simulation in water appear in Figs. 2 and 3. Plotted in Fig. 2 are admittance curves of the pressure plate for three submersion depths in water. The curves for the 3 and 9 mm submersion depths have been left out to avoid confusion. The reader should note the large admittance amplitude. The large peak in the 0 mm depth curve at 280 Hz corresponds to the (0,0) mode of the pressure plate, indicated by an arrow. This mode clearly shifts downward in frequency as the submersion depth is increased, indicated by arrows. Figure 2 illustrates the ease with which high admittance modes are tracked as the submersion depth increases. None of the peaks appear to be attributed to vertical modes in the water container since the half-wave resonance, calculated for the greatest water height in the box, 21.1 cm, is roughly at 3.5 kHz. A half-wave resonance in the water column is expected because the acrylic container rested on sand, which has a specific acoustic impedance roughly an order of magnitude less than that of the water. Assuming compressional wave speeds of 100 m/s in sand and 1500 m/s in water and densities of 1600 kg/m^3 in sand and 1000 kg/m^3 in water, the sand below the container approaches a pressure release boundary.

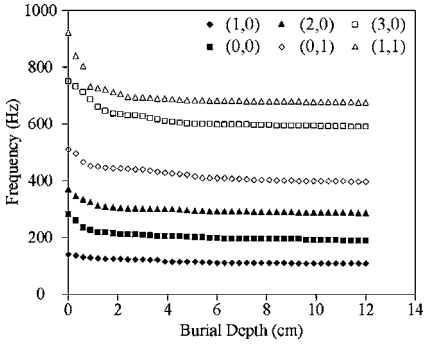


FIG. 3. Modal resonance frequency depth dependence curves showing saturation at greater depths.

Plotted in Fig. 3 are the resonance frequencies of several modes of the VS 1.6 versus submersion depth. It should be noted in Fig. 3 that as the height of the water above the plate increases, the modal frequencies cease their downward shift in frequency. Of interest is the rapid saturation of the frequency shift at a rather shallow depth of approximately 3 cm.

Figure 4 shows a plot of the resonance frequency shift of the (1,0) mode of the VS 1.6. The solid line in Fig. 4 is calculated from the solution to the equation of motion for a thin plate carrying a mass uniformly distributed across the surface of the plate; i.e., equivalent to a one-dimensional treatment of the mass load. Immediately, it is noted that the measured and calculated curves rapidly separate and, as observed in Fig. 3, the measured curve saturates while the uniform loading curve decreases with increasing immersion depth. Calculated curves for the other modes show behavior similar to that shown in Fig. 4 for the (1,0) mode.

There are some difficulties in making a direct link between the results presented here and the predictions by Velea *et al.*;[11] such a link is also not the intention of this work. The difficulties arise from the small size of the acrylic container and the parameters of the fluid in Velea *et al.* and those of the water. The effective fluid in Velea *et al.*[11] has a compressional wave speed of 160 m/s and a density of 1400 kg/m^3 where the experiment presented here was conducted in water having a compressional wave speed of roughly 1500 m/s and a density of 1000 kg/m^3. The large difference between the parameters precludes any direct link between the results of Velea *et al.* and those presented here.

The small size of the container implies that a 1D description of the mass loading of the plate should suffice since the container is too small for it to be considered infinite as compared to the size of the land mine. However, Fig. 4 demonstrates that such a 1D description does not work. Although the saturation predicted by Velea *et al.*[11] does not explicitly describe that observed here, the authors' description coupled with the small size of the container illuminates an explanation of the observed behavior. At shallow immersion depths, the resonance frequencies of the land mine are affected by a one-dimensional process; the pressure plate of the land mine sees the layer of water as a mass uniformly distributed across the plate and the resonance frequencies of the plate shift

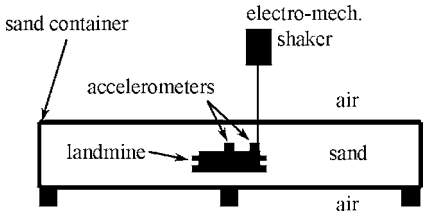


FIG. 5. Experimental configuration used during burial simulation in sand.

accordingly. As the immersion depth increases, a three-dimensional effect begins to dominate; the volume of water in the container adds an effective stiffness to the pressure plate. The observed saturation is due to an interpolation between the two effects.

B. VS 1.6 sand burial

In an attempt to obtain results for a medium that supports shear, the water experiments were repeated in dry, unconsolidated sand contained in a 96 by 146 by 29 cm wooden box. Before beginning the sand burial simulation, the land mine was allowed to dry. The shaker was again used to excite the vibration of the pressure plate, although the tip of the rod was glued to the pressure plate, and the LDV was replaced with two small accelerometers; the first placed 1 cm from the excitation point and the second placed at the center of the land mine. Figure 5 depicts the configuration used for this sand burial simulation experiment. The placement of the accelerometers allows the tracking of both symmetric and asymmetric modes. Ten centimeters of sand was then excavated from the box and the land mine was buried flush with the sand in the center of the box. The vibration of the pressure plate was then excited and the mode frequencies recorded. Sand was then added to the box in half-centimeter increments, corresponding to a mass of 130 g uniformly distributed over the pressure plate, until the excavated 10 cm of sand was replaced. The frequency response of each accelerometer was recorded at each increment.

Figure 6 shows the frequency response of the center of the pressure plate during the addition of the first 2 cm of sand. The broad peak in the zero depth response at approximately 300 Hz corresponds to the first symmetric mode of the pressure plate. As the height of sand above the pressure plate is increased, the first symmetric mode shifts downward in frequency. However, another large, broad peak at 430 Hz begins to overwhelm the first symmetric mode at a sand height of 2 cm. This large peak is believed to be due to

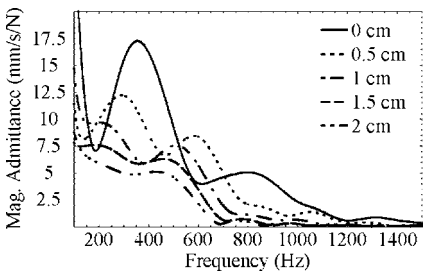


FIG. 6. Mechanical admittance of the center of the VS 1.6 with increasing depth in sand.

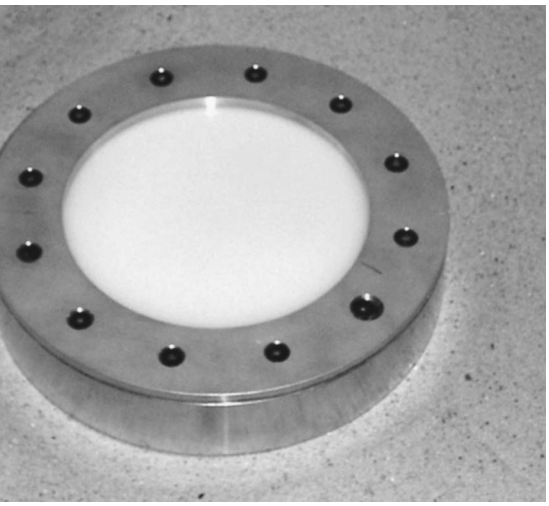


FIG. 7. Simulant as used during the sand burial simulation experiments.

resonance in the sand over the height of the sand container. By considering the frequency of 430 Hz to be a half-wave resonance in the sand column and taking the height of the column to be 17 cm, then a vertical compressional wave speed of approximately 150 m/s is obtained for the speed of sound in the sand. This appears to be a reasonable vertical wave speed since horizontal time-of-flight measurements in the sand container yield horizontal compressional wave speeds ranging from less than 100 to 200 m/s at depths of approximately 7–22 cm, respectively. Half-wave resonance is expected in this case because the bottom of the sand container is raised roughly 15 cm from the floor.

At sand heights greater than 2.5 cm, the modes within the sand container completely hide the response of the pressure plate as measured by the accelerometers. With the inability to track the modes during the burial simulation experiment in sand on the VS 1.6, the experiment was repeated using an acoustic land mine simulant with a greater mechanical response.

C. Simulant sand burial

The simulant constructed for this experiment was created to reasonably match the first symmetric frequency of the flush-buried VS 1.6 by placing a clamped plate above a backing volume of roughly the same size as that of the VS 1.6. The simulant is depicted in the Fig. 7. Figure 8 gives some insight into the reasons for attempting to repeat the

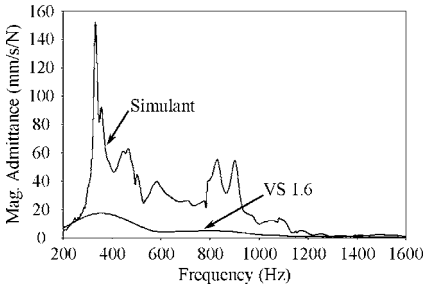


FIG. 8. Magnitude of the mechanical admittance at the center of the flush-buried VS 1.6 and Simulant.

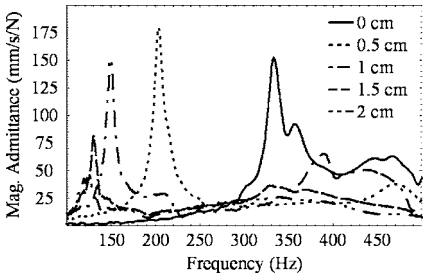


FIG. 9. Mechanical admittance of the center of the simulant with increasing depth.

experiment with the simulant. First the modes of the simulant are similarly located in frequency to those of the flush buried VS 1.6. Second and most important is the fact that the modes of the simulant are of higher Q than those of the VS 1.6. This does not allow the modes in the sand container to overwhelm those of the simulant.

Again, ten centimeters of sand was excavated from the container and the simulant was buried with its pressure plate flush with the sand. The shaker excitation and accelerometer measurements were repeated at the same half-centimeter increments as done with the VS 1.6. Figure 9 shows the response of the simulant pressure plate as the height of the sand above the pressure plate is increased to 2 cm. Immediately it is observed that the first symmetric mode of the pressure plate can be readily tracked as the height of the sand is increased. The first symmetric mode can be tracked to the full 10 cm sand height of the experiment. Figure 10 shows the change in the first symmetric modal frequency with increasing sand height relative to the frequency at the flush condition. Initially the frequency of the first symmetric mode shifts in a manner similar to the shift observed when adding mass to a simple harmonic oscillator. As the sand height increases, however, the frequency stops decreasing and begins to increase. When a sand height of approximately 4 cm is reached, the frequency shift ceases and a saturation phenomenon is observed similar to that observed in the water experiments. Plotted in Fig. 10 is a theoretical fit to the data utilizing an equation reported by Zagrai *et al.*,[10] which is

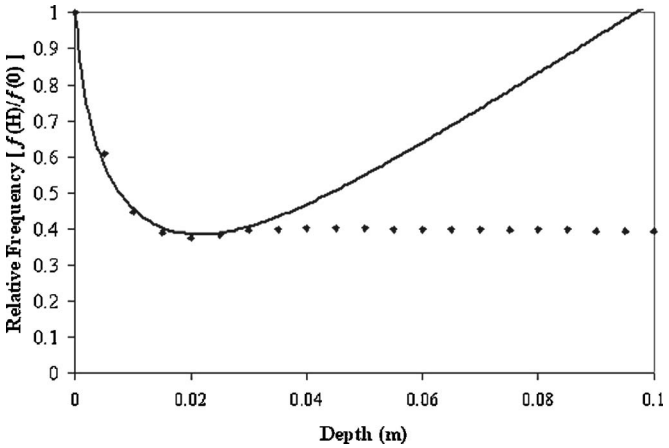


FIG. 10. Plot showing the resonance frequency shift of the first symmetric mode of the simulant (squares) versus increasing sand height. Also plotted is a theoretical curve (solid line) produced from Zagrai *et al.* (See Ref. 10) using typical parameter values for dry sand.

derived by considering the soil to act as an effective elastic plate directly coupled to the pressure plate. If the frequency equation reported by Zagrai *et al.* is normalized by the frequency of the pressure plate in the flush condition, the following frequency ratio is realized:

$$\frac{f(H)}{f(0)} = \sqrt{\frac{Y_m h_m^3 + Y_s H^3}{Y_m h_m^3}} \sqrt{\frac{M_m}{M_m + \rho_s H}}. \tag{1}$$

In Eq. (1), H refers to the burial depth of the mine, Y_s to an effective Young's modulus for the soil, ρ_s to the density of the soil, Y_m to the Young's modulus of the pressure plate, h_m to the thickness of the pressure plate, and M_m to the mass per unit area of the pressure plate. To create the fit curve in Fig. 10, the following parameters were used for the soil: Young's modulus of 8 MPa; soil density, in this case for dry sand, of 1600 kg/m^3. Parameters used for the simulant plate are a Young's modulus of 6.4 GPa, a density of 1245 kg/m^3, and a thickness of 3.2 mm. The value of Young's modulus for the plate is effective since the simulant is constructed with a backing volume that acts to stiffen the plate. Plate density and thickness are also slightly different from the material specifications. It is thought that these differences might arise from the polyamide's ability to absorb moisture.[13]

Equation (1) provides a very good fit, at depths less than 4 cm, to the data taken during the sand burial simulation (see Fig. 10). At depths larger than 4 cm, shallow in terms of land mine detection, the plate approximation of Eq. (1) breaks down, implying that the soil-land mine system can no longer be described one dimensionally. Experimental results reported by Zagrai *et al.*,[10] Korman and Sabatier,[9] and Turner *et al.*[16] all show the behavior predicted by Eq. (1). Upon increased burial depth, an initial decrease in resonance frequency was followed by an increase in resonance frequency at greater depths. In all cases, however, the experimental configuration consisted of a plate or land mine with a terminating pipe above. Into this pipe, layers of soil are distributed over the plate, thus increasing its burial depth. The pipe above the plate also has the effect of limiting the plate's radiation to one dimension. As such, the three-dimensional saturation effect observed on the land mine simulant is not observed in Refs. 9, 10, and 16.

There are further differences between the experiments presented here and those in the literature and in the field. In the literature, as well as in the field, the land mine is often driven by loudspeakers or shakers insonifying the soil surface and the soil surface vibration is then measured.[2,7–10] Here the pressure plates of the land mine and the simulant were driven directly using a thin rod connecting the plate to a shaker while an accelerometer or LDV is used to directly measure the vibration of the plate, not the soil surface. A recent work by Korman *et al.*[17] demonstrates, by simultaneously measuring the plate and soil surface responses, that the vibrations, at resonance, of the plate and the soil surface are nearly identical. This has the implication that the plate vibration measurement methods described here are justified. It also follows that the insonification method is justified be-

cause the land mine's modes control the soil surface vibration, so directly exciting the modes and exciting them through the soil are equivalent.

In the field, the ground is truly an infinite 3D medium and excitation and measurement are done through the soil surface. Further, the soil is inhomogeneous on a scale similar to that of the land mine, the soil is consolidated by weathering processes, and the soil is subject to wide variations in moisture content over relatively short time scales. These are all reasons to attempt to study the depth dependence of a land mine's resonance frequencies in a controlled laboratory environment. By moving into the laboratory, all of the previous unknowns can be reasonably controlled in order to isolate the depth dependent resonance frequency. However, laboratory containers introduce their own problems, particularly modes within the container. This has been addressed here by the high Q of the land mine simulant.

III. CONCLUDING REMARKS

Resonance frequency shift saturation has been observed in burial simulation experiments performed in both water and sand. Qualitatively, this can be explained by the land mine being affected by the three-dimensional nature of its burial medium.

When submerged in water, this saturation effect occurs at shallow depth after the land mine exhibits resonance frequency shifts similar to those expected from simple mass loading theory. The experimental configuration used in the water experiment, specifically the size of the acrylic cube, does not allow the saturation description reported by Velea *et al.*[11] to be applied to the observations of the water experiment. Despite the small cube, a transition between a one-dimensional mass loading effect and a three-dimensional effect of the water stiffness qualitatively explains the observations.

At very shallow burial depths in sand, the land mine simulant exhibits behavior observed in essentially one-dimensional experiments[7–10] and explained theoretically by Zagrai *et al.*[10] At the most shallow burial depths, less than 1.5 cm, the simulant's resonance first shifts downward in frequency due to simple mass loading. Upon deeper burial, this shift is followed by an increase in resonance frequency, suggested to be due to a cubic dependence of shear stiffness on burial depth [10]. The observed resonance frequency shift then deviates from the coupled plate prediction of Zagrai *et al.* at a depth of roughly 3 cm after which the upward shift slows and finally saturates near a burial depth of 4 cm. At depths ranging from 0 to 3 cm, the coupled plate model, with reasonable parameters for sand, yields a good theoretical fit to the data before the saturation is observed at depths still within the range of typical land mine emplacement.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Army Research, Development, and Engineering Command, Communications-Electronics Research, Development, and Engineering Center, Night Vision and Electronic Sensors Directorate under Contract No. DAAB15-02-C-0024 and the Department of the Navy, Office of Naval Research, under Grant No. N00014-02-1-0878. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors. The authors would like to thank Alexander Ekimov for discussions regarding soil shear stiffness and the two reviewers for their helpful comments.

[1] N. Xiang and J. M. Sabatier, "An experimental study on antipersonnel land mine detection using acoustic-to-seismic coupling," J. Acoust. Soc. Am. **113**, 1333–1341 (2003).

[2] W. R. Scott, Jr., and J. S. Martin, "Experimental investigation of the acousto-electromagnetic sensor for locating land mines," Proc. SPIE **3710**, 204–214 (1999).

[3] D. Donskoy, A. Ekimov, N. Sedunov, and M. Tsionskiy, "Nonlinear seismo-acoustic land mine detection and discrimination," J. Acoust. Soc. Am. **111**, 2705–2714 (2002).

[4] J. M. Sabatier and N. Xiang, "Laser-doppler based acoustic-to-seismic detection of buried mines," Proc. SPIE **3710**, 215–222 (1999).

[5] J. M. Sabatier and N. Xiang, "An investigation of acoustic-to-seismic coupling to detect buried antitank land mines," IEEE Trans. Geosci. Remote Sens. **39**, 1146–1154 (2001).

[6] C. King, editor, *Jane's Mines and Mine Clearance 2001–2002*, Jane's Information Group, Inc., Alexandria, VA (2001).

[7] A. Zagrai, D. Donskoy, and A. Ekimov, "Resonance vibrations of buried land mines," Proc. SPIE **5415**, 21–29 (2004).

[8] J. M. Sabatier and M. S. Korman, "Nonlinear tuning curve vibration response of a buried land mine," Proc. SPIE **5089**, 476–486 (2003).

[9] M. S. Korman and J. M. Sabatier, "Nonlinear acoustic techniques for land mine detection," J. Acoust. Soc. Am. **116**, 3354–3369 (2004).

[10] A. Zagrai, D. Donskoy, and A. Ekimov, "Structural vibrations of buried land mines," J. Acoust. Soc. Am. **118**, 3619–3628 (2005).

[11] D. Velea, R. Waxler, and J. M. Sabatier, "An effective fluid model for land mine detection using acoustic to seismic coupling," J. Acoust. Soc. Am. **115**, 1993–2002 (2004).

[12] W. C. K. Alberts, II, R. Waxler, and J. M. Sabatier, "A study of the acoustic behavior of a plastic, blast-hardened, anti-tank land mine," J. Acoust. Soc. Am. **118**, 2022 (2005).

[13] A. Freeman, S. C. Mantell, and J. H. Davidson, "Mechanical performance of polysulfone, polybutylene, and polyamide 6/6 in hot chlorinated water," Sol. Energy **79**, 624–637 (2005).

[14] C.-C. Pai, R.-J. Jeng, S. J. Grossman, and J.-C. Huang, "Effects of moisture on thermal and mechanical properties of Nylon-6,6," Adv. Polym. Technol. **9**, 157–163 (1989).

[15] R. Marsili, L. Pizzoni, and G. Rossi, "Uncertainty analysis of vibration measurements of tools inside fluids by laser Doppler techniques," Proc. SPIE **3411**, 92–100 (1998).

[16] J. A. Turner, L. Yang, and W. Kang, "Effect of particle size on the vibration of plates loaded with granular material," J. Acoust. Soc. Am. **118**, 2022 (2005).

[17] M. S. Korman, J. M. Sabatier, K. E. Pauls, and S. A. Genis, "Nonlinear acoustic land mine detection: comparison of off-target soil background and on-target soil-mine nonlinear effects," Proc. SPIE **6217**, 62170Y-1–62170Y-8 (2006).

Volumetric acoustic vector intensity imager

Earl G. Williams,[a] Nicolas Valdivia, and Peter C. Herdic[b]
Acoustics Division, Naval Research Laboratory, Washington DC 20375

Jacob Klos
NASA Langley Research Center, Hampton, Virginia 23681

(Received 1 March 2006; revised 21 July 2006; accepted 21 July 2006)

A new measurement system, consisting of a mobile array of 50 microphones that form a spherical surface of radius 0.2 m, that images the acoustic intensity vector throughout a large volume is discussed. A simultaneous measurement of the pressure field across all the microphones provides time-domain holograms. Spherical harmonic expansions are used to convert the measured pressure into a volumetric vector intensity field on a grid of points ranging from the origin to a maximum radius of 0.4 m. Displays of the volumetric intensity image are used to locate noise sources outside the volume. There is no restriction on the type of noise source that can be studied. An experiment inside a Boeing 757 aircraft in flight successfully tested the ability of the array to locate flow-noise-excited sources on the fuselage. Reference transducers located on suspected noise source locations can also be used to increase the ability of this device to separate and identify multiple noise sources at a given frequency by using the theory of partial field decompositions. The frequency range of operation is 0 to 1400 Hz. This device is ideal for the diagnostic analysis of noise sources in commercial and military transportation vehicles in air, on land, and underwater.
[DOI: 10.1121/1.2336762]

PACS number(s): 43.20.Ye, 43.40.At, 43.60.Pt, 43.60.Sx, 43.60.Lq [SFW] Pages: 1887–1897

I. INTRODUCTION

The measurement of acoustic intensity has long tempted the acoustician with its promises of localization and quantification of unknown noise sources and the determination of the sound power radiated from a machine.[1,2] The acoustic intensity probe using two microphones is now a main stay of many measurements in industry and government. More sophisticated probes are appearing such as the three-axis intensity probe that provides the intensity vector at a point.[3]

Spherical arrays of microphones/hydrophones, that are the footing of this paper, have long been used in acoustics. Spherical harmonic decompositions of the measured field almost always form the basis behind the theory in these array systems and two early implementations were significant in acoustics.[4,5] A very recent paper is a very good source of information about sound-field analysis using spherical arrays.[6] Radar antenna measurements using spherical arrays have a rich history, relying heavily on spherical harmonic decompositions.[7,8] Almost all of the research and development has been aimed at predicting the far field from knowledge of the near field representing mathematically a well-posed forward problem. The research introduced here we believe presents for the first time a volumetric and holographic projection of the intensity vector encompassing the volume in the interior of a spherical array of microphones (well-posed forward problem) as well as the volume just outside the sphere (ill-posed inverse problem). The imaged intensity represents a spatially smoothed version of the active intensity and thus differs dramatically from the the popular single axis and multiple axis vector intensity probes that provide high measurement accuracy and no spatial smoothing. The volumetric intensity imager aims at more qualitative results, sacrificing high accuracy to image the intensity vector field at hundreds of points throughout a sizable volume, instantaneously. Furthermore this device is augmented in its ability to isolate individual noise sources and map their intensity fields by using state-of-the-art signal processing.

We present in Sec. II a brief overview of basic theory of operation and design and in Sec. III we provide analytic formulas for the reconstruction error of the pressure field based on a plane wave source. A numerical experiment with a point monopole source outside the volume is presented in Sec. IV to articulate the errors of the intensity vector reconstructions. The front end signal processing used to isolate individual noise sources is described in Sec. V and application to an in-flight experiment inside a Boeing 757 aircraft is presented in Sec. VI.

II. RECONSTRUCTION EQUATIONS

A hallmark of near-field acoustical holography (NAH) is the reconstruction of the acoustic field in a volume, which we call a "volumetric reconstruction," from information obtained on a surface. Spherical NAH provides the most ideal formulation for volumetric reconstructions, both in simplicity of theory and ease of application. Consider a spherical reconstruction volume $\mathcal{V}$ represented by spherical coordinates $\mathbf{r}=(r,\theta,\phi)\in\mathcal{V}$ of extent defined by $0\leqslant r\leqslant r_{\max}$ which is source free (homogeneous wave equation applies). The array of microphones assumed constructed to have negligible scattering is located at $r=a<r_{\max}$. The acoustic pressure

[a]Electronic mail: earl.williams@nrl.navy.mil
[b]Also with SFA Inc., Largo, MD 20774.

$p_\infty(\mathbf{r},\omega)$ may be represented anywhere in $\mathcal{V}$ by an expansion in terms of orthogonal spherical harmonics $Y_n^m(\theta,\phi)$ and spherical Bessel functions[9] j_n:

$$p_\infty(\mathbf{r},\omega) = \lim_{N\to\infty} p_N(\mathbf{r},\omega),$$

where

$$p_N(\mathbf{r},\omega) \equiv \sum_{n=0}^{N} \frac{j_n(kr)}{j_n(ka)} \sum_{m=-n}^{n} P_{mn}(a,\omega) Y_n^m(\theta,\phi), \tag{1}$$

with $k=\omega/c$. The unknowns P_{mn} in this equation are called the Fourier coefficients. The components of the velocity vector are given in terms of these unknown Fourier coefficients:[9]

$$v_\theta(\mathbf{r},\omega) = \frac{1}{i\omega\rho} \sum_{n=0}^{N} \frac{j_n(kr)}{r j_n(ka)} \sum_{m=-n}^{n} P_{mn} \frac{\partial Y_n^m(\theta,\phi)}{\partial\theta},$$

$$v_\phi(\mathbf{r},\omega) = \frac{1}{i\omega\rho} \sum_{n=0}^{N} \frac{j_n(kr)}{r j_n(ka)} \sum_{m=-n}^{n} P_{mn} \frac{imY_n^m(\theta,\phi)}{\sin\theta}, \tag{2}$$

$$v_R(\mathbf{r},\omega) = \frac{1}{i\rho c} \sum_{n=0}^{N} \frac{j_n'(kr)}{j_n(ka)} \sum_{m=-n}^{n} P_{mn} Y_n^m(\theta,\phi),$$

where the equalities hold strictly only in the limit as $N\to\infty$. Note that these expressions use only $j_n(kr)$ which is finite at the origin, as opposed to $h_n(kr)$ that must be used if the sources are located at $r<a$.[4,5] Finally, the active intensity vector $\vec{I}$ in spherical coordinates is then determined by the usual expression using unit vectors $\hat{e}$:

$$\vec{I}(\mathbf{r},\omega) = \frac{1}{2}\Re[p_\infty^*(v_\theta\hat{e}_\theta + v_\phi\hat{e}_\phi + v_R\hat{e}_R)]. \tag{3}$$

For the volumetric acoustic intensity imager (VAIM) the intensity is computed on a cubic lattice of points in $\mathcal{V}$ and displayed in three-dimensional plots, as will be shown in Sec. IV.

The unknown Fourier coefficients $P_{mn}(a,\omega)$ are determined[9] by integration of the pressure field at $r=a$ over a sphere:

$$P_{mn}(a,\omega) \equiv \int\int p_\infty(a,\theta,\phi,\omega) Y_n^m(\theta,\phi)^* d\Omega, \tag{4}$$

with $d\Omega \equiv \sin\theta d\theta d\phi$ and where p_∞ is derived from a temporal Fourier transform of the measured pressure in the usual way. The spherical array is designed so that the microphones are located at the quadrature points (θ_j,ϕ_j), $j=1,\ldots,50$, of an efficient algorithm to compute the surface integration in Eq. (4). One such efficient numerical quadrature algorithm is given by Lebedev,[10,11] who provides a set of quadrature algorithms for optimum quadratures on a spherical surface for a range of microphone densities from 38 to 890. These algorithms are optimum by providing an exact integration of products of spherical harmonics up to a given sum of orders. As we will see in the following this errorless integration is critical. The 50 element algorithm used in this paper integrates with no error products of spherical harmonics (say of order n' and n) up to $n+n'\leqslant 11$. Thus under this condition

$$\sum_{j=1}^{50} w_j Y_n^m(\theta_j,\phi_j) Y_{n'}^{m'}(\theta_j,\phi_j) = \delta_{mm'}\delta_{nn'},$$

where δ is the Kronecker delta and w_j are the quadrature weights. Even when $n=n'=6$ most of the orthogonality still remains, so the quadrature algorithm breaks down "gracefully." Lebedev's algorithms are invariant with respect to octahedral symmetry, that is, the microphone locations on the spherical cap subtending one of the eight faces of the octahedron are identical (after rotation) on the other seven faces. Using the quadrature algorithm Eq. (4) is approximated by $\hat{P}_{mn}$,

$$\hat{P}_{mn}(a,\omega) = \sum_{j=1}^{50} w_j p_\infty(a,\theta_j,\phi_j,\omega) Y_n^m(\theta_j,\phi_j)^*. \tag{5}$$

Since from Eq. (1)

$$p_\infty(a,\theta_j,\phi_j,\omega) = \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} P_{m'n'} Y_{n'}^{m'}(\theta_j,\phi_j)$$

the algorithm forces

$$\hat{P}_{mn} = P_{m'n'}\delta_{mm'}\delta_{nn'} \quad \text{if } n+n' \leqslant 11. \tag{6}$$

Thus we are guaranteed that $\hat{P}_{mn}=P_{m'n'}$ as long as n or $n'\leqslant 5$, i.e., the first 36 Fourier coefficients are determined without any integration error. For the computation of the pressure and velocity vector in Eqs. (1) and (2) we choose to use only these accurately computed Fourier coefficients and thus these sums are truncated to $N\leqslant 5$. We found that there was no gain in accuracy if we used the $n=6$ approximated Fourier coefficients.

The quadrature weights and locations in Cartesian coordinates are easily derived using Lebedev's parameters listed under 11.1 in Ref. 11 and are not reproduced here for brevity. However, we note that the locations put 12 microphones on each of the Cartesian coordinate planes. Thus spatial aliasing of the acoustic wavelength occurs at 1.64 kHz so that at least two samples per wavelength occur below 1400 Hz, the upper frequency of operation.

As stated in the abstract, the VAIM provides a spatially smoothed version of the active intensity, an inevitable effect of the finite limit in the summation over n. It is straightforward to show that the relationship between the reconstructed field p_N and the measured field p_∞ can be expressed as

$$p_N(a,\theta',\phi') = \int\int p_\infty(a,\theta,\phi) \times \left[\sum_{n=0}^{N}\sum_{m=-n}^{n} Y_n^m(\theta,\phi)^* Y_n^m(\theta',\phi')\right] d\Omega.$$

The smoothing effect of the term in square brackets in the above-presented equation is brought out by the completeness relation for spherical harmonics,[9]

$$\sum_{n=0}^{\infty}\sum_{m=-n}^{n} Y_n^m(\theta,\phi)^* Y_n^m(\theta',\phi') = \delta(\phi-\phi') \times \delta(\cos\theta - \cos\theta'),$$

which provides a $\sin(Nx)/(\pi x)$ like behavior, becoming a delta function in space as $N\to\infty$. This smoothing carries over to the velocity and intensity vector computations. As to the errors that result, we turn to the following two sections.

III. ERROR ANALYSIS AND CONSIDERATION OF NOISE

One can derive analytical formulas of the error in the reconstruction of the pressure and velocity vector fields as a function of r if one assumes that the source is a plane wave and that the noise $\epsilon(\theta,\phi)$ in the microphones is Gaussian and spatially incoherent with variance $\sigma^2=E[|\epsilon|^2]$, where E is the ensemble average. The derivation is presented in Ref. 9 for the normal velocity error, and the development for the pressure error is identical and thus is not presented here. We present results only for the pressure case, as the velocity cases do not present any additional insight. Since these formulas are based on a plane wave source, they do not include the error due to the presence of evanescent waves in the source field, waves that typically exist in the near field of sources. However, the formulas do provide valuable insight. It is assumed in this derivation that the Fourier coefficients are determined without error by the quadrature algorithm, as long as $N\leqslant 5$, as discussed earlier.

We define the root mean square error $\mathcal{E}$ at a radius r using integration of the recon structed field over a sphere of radius r as

$$\mathcal{E}(r) \equiv \left(\frac{\langle E[|p_\infty(\mathbf{r}) - p_N^\delta(\mathbf{r})|^2]\rangle}{\langle |p_\infty(\mathbf{r})|^2\rangle}\right)^{1/2}, \tag{7}$$

where $\langle\bullet\rangle \equiv (1/4\pi)\int\int \bullet d\Omega$. In this equation p_∞ is the exact pressure with no noise and $p_N^\delta(\mathbf{r})$ is defined, following Eq. (1), as

$$p_N^\delta \equiv \sum_{n=0}^{N} \frac{j_n(kr)}{j_n(ka)} \sum_{m=-n}^{n} \hat{P}_{mn}^\delta Y_n^m(\theta,\phi), \tag{8}$$

constructed from the measured Fourier coefficients $\hat{P}_{mn}^\delta$ including noise, following Eq. (5):

$$\hat{P}_{mn}^\delta = \sum_{j=1}^{50} w_j p^\delta(a,\theta_j,\phi_j) Y_n^m(\theta_j,\phi_j)^*, \tag{9}$$

where $p^\delta(a,\theta_j,\phi_j) \equiv p_\infty(a,\theta_j,\phi_j) + \epsilon(\theta_j,\phi_j)$ is the pressure measured at the jth microphone location. Certainly as $\sigma\to 0$ $p_N^\delta \to p_N$.

Following Chap. 7 of Ref. 9 to evaluate Eq. (7) we find, given a plane wave incident *at any angle*, that the root mean square error at r is

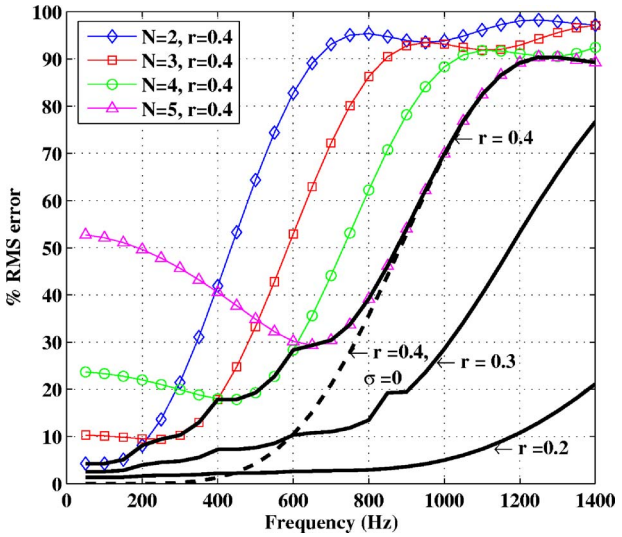


FIG. 1. (Color online) Total error from Eq. (10) for 30 dB SNR, $r=0.4$ m, and values of $N=2,3,4,5$. The solid curve labeled $r=0.4$ follows the values of N with the minimum error, and thus represents the actual error after regularization (choosing optimum value of N).

$$\mathcal{E}(r) = \left(1 - \sum_{n=0}^{N}(2n+1)j_n(kr)^2 + \frac{\sigma^2}{16\pi\langle|p_\infty|^2\rangle}\sum_{n=0}^{N}(2n+1)\left(\frac{j_n(kr)}{j_n(ka)}\right)^2\right)^{1/2}, \tag{10}$$

where we call the first two terms the "base system error" and the last (third) term the "noise error." In the derivation of Eq. (10) the 16π in the denominator results from an additional approximation given by $\Sigma_{j=1}^{50} w_j^2 |Y_n^m(\theta_j,\phi_j)|^2 \approx 1/4$, not in Ref. 9.

We study Eq. (10) to understand the basic errors that arise in the reconstructions. Note that in the base system error term that $\Sigma_{n=0}^{\infty}(2n+1)j_n(kr)^2=1$, so that the base system error diminishes to zero as N increases. Since N is limited due to Eq. (6), the base system error can only be reduced by increasing the number of microphones in the array. For example, based on Lebedev's formulas[11] we would need 86 microphones to get to $N=7$ and 170 to get to $N=10$. Note also the remarkable result that this error does not depend on the radius of the array; any size array will encounter the same base system error when reconstructing the field at a radius r. The upper frequency limit of the VAIM is determined by the base system error, as will show in the following.

The noise error term in Eq. (10) is inversely proportional to the signal-to-noise ratio (SNR) $\langle|p_\infty|^2\rangle/\sigma^2$ and increases with r. What is not clear is the critical fact that the noise error can be reduced by decreasing the limit of the summation N at the low frequencies. This fact leads directly to a regularization filter in n which we discuss in the next section.

Regularization filter. Figure 1 is a plot of $\mathcal{E}(r)$ of Eq. (10) for a SNR of 30 dB and various cutoffs N. In the results to follow $a=0.2$ m the radius of the VAIM array. The four curves corresponding to the legend represent the error when the reconstruction radius $r=r_{\max}=0.4$ and when N is limited to the values indicated in the legend. We can see that at the

lowest frequencies $N=2$ provides the least error up to 200 Hz, switching to $N=3$ then to $N=4$ at 400 Hz and finally to $N=5$ at and above 600 Hz. The solid line labeled $r=0.4$ m follows the values of N with the minimum error. The variable cutoff in the n summation for minimum error in Eq. (10) represents a regularization filter (with no taper in n) that controls the $(kr)^n$ behavior of $j_n(kr)$ in its small argument domain, viz. $kr<n$. Filters of this kind are critical to the success of NAH.[12]

Two solid lines below the $r=0.4$ line in the figure represent the minimum error for two other reconstruction radii, $r=0.3$ m and $r=0.2$ m as labeled and show that the errors diminish significantly for smaller reconstruction radii. Although not shown the results at these two radii have the same transition frequencies in N for minimum errors as the $r=0.4$ m case noted earlier, namely 200, 400, and 600 Hz. Thus, for example, $N=2$ provides the smallest error in the range 0–200 Hz for *all three* radii.

Finally, the dashed line in Fig. 1 labeled $r=0.4$ m, $\sigma=0$ is a calculation from Eq. (10) of only the base system error at a reconstruction radius of 0.4 m and $N=5$. Since this curve follows the last legend curve (that includes noise) it is clear that the base system error dominates the high frequency regime (above 800 Hz in this figure). This fact sets the upper frequency limit of operation of the VAIM as well as the maximum reconstruction radius. Adding more sensors will decrease this error, as discussed earlier, and thus extend the high frequency limit of operation. Futhermore with respect to regularization, we note the important deduction that no regularization is needed above about 800 Hz since the noise error is overshadowed by the base system error here. Contrarily, at the low frequencies the difference between the aforementioned two curves shows that the error is controlled by the noise error term of Eq. (10), not by the number of microphones in the array.

The above-presented simple regularization scheme provides the potential to determine an *a priori* regularization scheme based on two variables, the frequency and the SNR, as will be discussed further in Sec. IV A. We now turn to examine volumetric intensity reconstructions and the error associated with them.

IV. NUMERICAL EXPERIMENT WITH A POINT MONOPOLE SOURCE

We choose to model intensity errors using point sources instead of plane waves as was done earlier. The point source is an attempt to more correctly model the errors when the spherical array is placed close to vibrating surfaces. We choose a point source located outside the array in otherwise free space at distance of 1 m from its origin. Its angular orientation with respect to the coordinate system of the array is irrelevant due to the important fact that *reconstructions are invariant to rotations of the spherical array*. This result follows from the faithful computation of the Fourier coefficients Eq. (9). One cannot derive simple formulas, like those noted earlier for pressure reconstructions, to estimate the errors associated with reconstruction of the intensity vector field in the volume $r\leqslant 0.4$. Thus we resort to computer simulations in which the pressure field at each of the 50 microphones, to which is added a predetermined level of spatially uncorrelated Gaussian noise, is computed. The exact intensity field without noise, computed at each of the reconstruction points in the volume, is known analytically since the point source is in a free field. This result is compared with the reconstruction of the vector field from the pressure hologram with added noise using Eqs. (1)–(4) (spherical NAH) and errors are determined. The level of random noise is set by the SNR. The signal is defined as the rms pressure level without noise "measured" by the 50 microphones and the noise is the square root of the variance of the added noise.

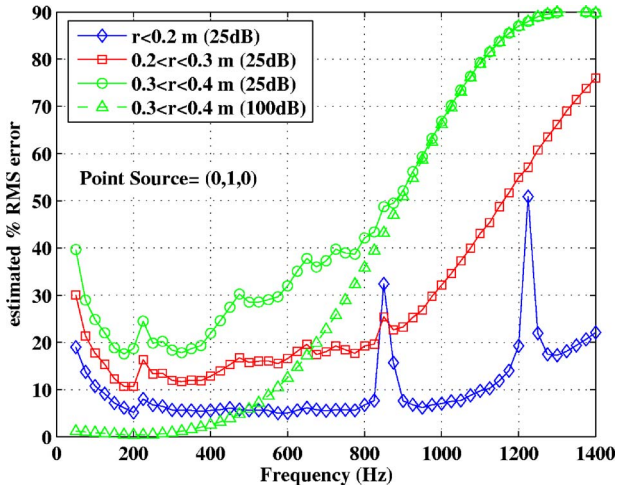


FIG. 2. (Color online) Estimated root mean square intensity errors in different reconstruction volumes with added noise (SNR=25 dB), for a point source at 1.0 m from origin.

A. Estimated errors for intensity vector reconstruction and an *a priori* regularization scheme

The reconstruction error for the intensity vector is defined by

$$\mathcal{E}=\frac{\|\vec{I}_{\mathrm{ex}}(\mathbf{r})-\vec{I}_{\mathrm{nah}}(\mathbf{r})\|_2}{\|\vec{I}_{\mathrm{ex}}(\mathbf{r})\|_2}, \tag{11}$$

where $\vec{I}_{\mathrm{ex}}$ is the exact intensity computed for a point source and $\vec{I}_{\mathrm{NAH}}$ is the vector intensity reconstructed from the pressure with noise at the microphones and $\|\bullet\|_2$ is the L2 norm over a specified volume.

Figure 2 shows the results for the intensity error in three different reconstruction volumes defined by $r\leqslant 0.2$, $0.2<r\leqslant 0.3$, and $0.3<r\leqslant 0.4$, using an ensemble of numerical experiments with Gaussian noise added to the microphones with a SNR level of 25 dB. The plots represent the minimum error determined by varying the values of N from 1 to 5; a manual regularization approach that determines the optimum filter. As with the pressure errors discussed earlier the error increases with reconstruction radius as well as with frequency. The fourth curve in the legend is for a 100 dB SNR corresponding closely to a no noise case, for comparison to Fig. 1. The two narrow peaks in the first legend curve arise from the zeros of $j_0(ka)$ and $j_1(ka)$ at 857 and 1225 Hz, respectively, that arise in the denominator of Eqs. (1) and (2).

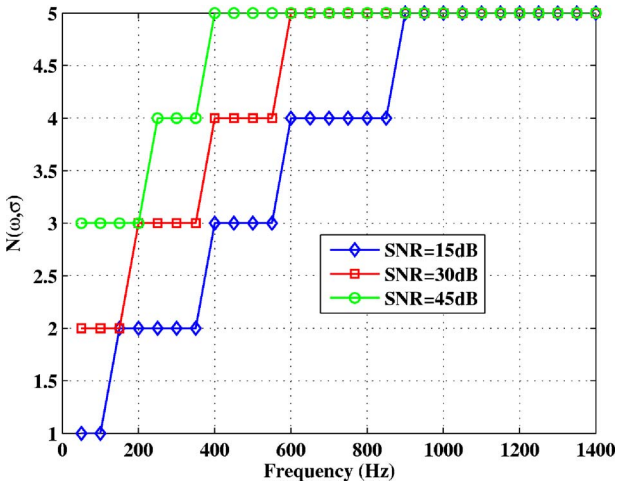


FIG. 3. (Color online) Optimum regularization filter for a point source located 1 m from the array center as a function of signal to noise ratio. The first three legend curves shown in Fig. 2 used a regularization curve similar to the curve with the square symbols (30 dB SNR).

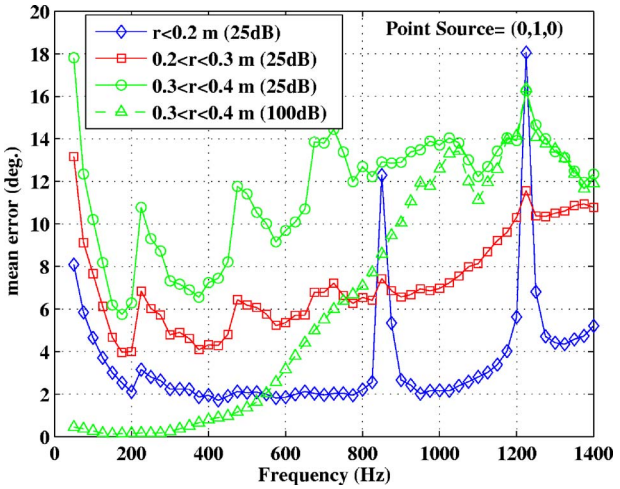


FIG. 4. (Color online) Mean error in the angle of the reconstructed intensity vector for a point source at 1 m and a SNR 30 dB. The average error is calculated in the three reconstruction volumes indicated. The sudden jumps in the curves occur when the value of N changes.

To have a finite solution, the Fourier coefficients must be zero for $n=0$ and $n=1$ at these frequencies, respectively. In other words the solution of the Helmholtz equation for the interior of a sphere is $j_n(kr)Y_n^m(\theta,\phi)$ which exhibits a node at the problem frequencies. However, these Fourier coefficients are nonzero due to noise in the measurement, so the solution becomes infinite. This problem is eliminated by choosing an optimum filter that excludes the $n=0$ term in a 10 Hz band about the 857 Hz zero, and the $n=1$ term in a 10 Hz band centered at 1225 Hz. Although this filtering eliminates nonevanescent fields and increases the error as the figure shows, a compromise must be struck to handle the indeterminacy of the reconstruction problem at the zeros of the denominator.

This optimum filter (minimum error) that was used to compute Fig. 2 is shown in Fig. 3 for three different values of SNR. It is quite remarkable that the break points for the values of N for the 30 dB curve are at 200, 400, and 600 Hz, the same as determined for the plane wave source and the pressure reconstruction presented in Fig. 1. This fact is important in justifying an *a priori* regularization filter that is fixed for the experiments discussed in Sec. VI A, a modification that speeds the processing of the VAIM and reflects our aim at real-time applications.

Returning to Fig. 2 it is clear that the intensity errors become very large in the outer reconstruction volume above 800 Hz. Although this might appear unacceptable (we will see in the following what the intensity field looks like with large error) it is important to consider a different view of the reconstruction accuracy, the error in the angle of the intensity vector. Errors in angle will be more misleading than amplitude errors as to the location of a concentrated noise source outside the array and conversely small angle errors will favor source ID even in the face of large magnitude errors. Figure 4 shows the mean of the error in the angle between the reconstructed intensity vector and the exact result for the same three reconstruction volumes presented in Fig. 2. The error angle α between the reconstructed and the exact vector was computed using the dot product relation, $\cos(\alpha)=\vec{I}_{\mathrm{ex}}\cdot\vec{I}_{\mathrm{NAH}}/(|\vec{I}_{\mathrm{ex}}||\vec{I}_{\mathrm{NAH}}|)$. Note that $\alpha\geqslant 0$. Comparisons were made only for intensity vectors of magnitude within 1/10 of the maximum magnitude as these are the vectors which are visible on a linear display. This figure shows the significant fact that the error in direction of the reconstructed intensity vectors in the volume is quite small.

B. Volumetric intensity reconstruction

Examples of the reconstructed intensity fields compared with the exact results for a point source located on the y axis 1 m from the origin are shown in the next set of figures, Figs. 5–7, or three different frequencies. The SNR is 25 dB and the intensity is plotted in the volume $r\leqslant 0.4$. In these figures the cones point to the direction of the intensity vector, and the length (and width) of the cone is proportional to the linear magnitude of the intensity. The center of the base of the cone is a point in a cubic lattice specifying the locations of the intensity vectors. The lattice spacing in each direction is 0.08 m. The 50 elements (small circles) of the measurement sphere are superimposed in each plot for reference. Results in Figs. 5–7 are for 250, 800, and 1150 Hz, respectively, with the exact result on the right and the reconstructed field on the left (plotted with the same scale). The essential

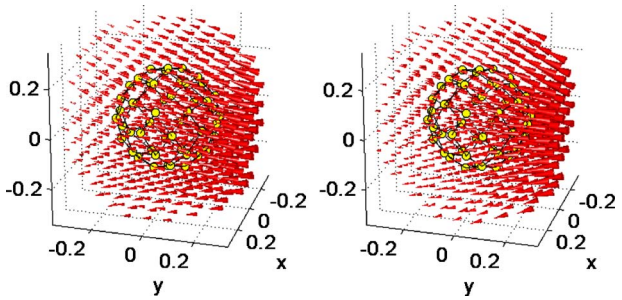


FIG. 5. (Color online) Output from the VAIM at 250 Hz and 25 dB SNR for a point source at 1.0 m is shown on the left vs the exact field shown on the right both plotted on the same scale. The intensity vectors are plotted on a linear scale.

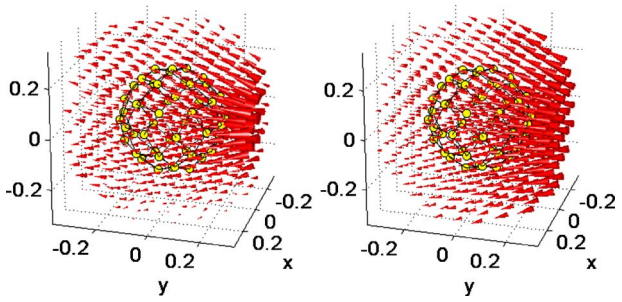


FIG. 6. (Color online) Same as Fig. 5 except at 800 Hz.

conclusion drawn from these figures is that the direction of the point source (located on axis at $y=1.0$) is correctly indicated by the reconstructed field (left-hand plots) and the actual location can be found by the intersection of lines collinear with the intensity vectors near the source. In comparing to the exact result on the right panel of the figure note that accompanying errors in the three main reconstruction volumes $r \leqslant 0.2$, $0.2 < r \leqslant 0.3$ m, and $0.3 < r \leqslant 0.4$ m, were given in Fig. 2. The large errors in the outer volume $0.3 < r \leqslant 0.4$ m at the two higher frequencies are evident in the comparison if one concentrates on the vectors at the outer reaches of the volume in Figs. 6 and 7. Note, however, the mean angle error of the visible vectors in this volume (according to 4) is less than 14° at 800 and 1150 Hz, so that even with the large amplitude errors and field distortion in the outer volume the direction and location of the point source is still determined. Again we want to emphasize that the errors at 800 and 1150 Hz arise from the base system error, and are not related to the SNR. To diminish these errors one must increase the number of microphones in the array so that more spherical harmonics are included as discussed in Sec. III.

The level of noise in these simulations was preset. However, it is generally not known during physical experiments and can be determined by the following procedure.[12,13] We assume that the level of the evanescent waves with $n=6$ associated with the source has decayed beyond the noise level (spatially uncorrelated) at the microphones. This decay is given in the region $kr < n$ by $j_n(kr) \approx (kr)^n/(2n+1)!!$, a power-law *decay toward* the origin. Thus we compute the standard deviation of the noise σ using $\sigma \approx E[\|\hat{P}^{\delta}_{mn}\|]/\sqrt{13}$ for $n=6$ using the norm of the 13 harmonics $m=-6,-5,\ldots,6$. This method works faithfully for frequencies $kr_{\max} < 6$ or $f < 819$ Hz for $r_{\max}=0.4$. Above 800 Hz knowledge of the noise level is unnecessary since the errors are dominated by base system error, as discussed earlier.

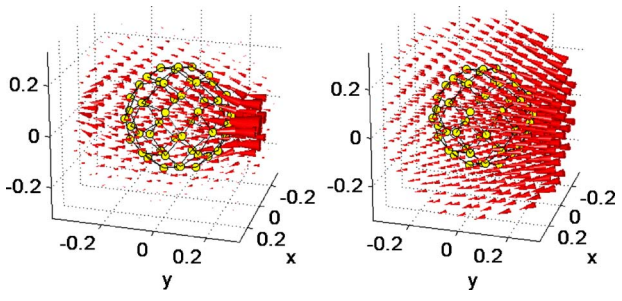


FIG. 7. (Color online) Same as Fig. 5 except at 1150 Hz.

The signals used in the above-noted simulations above were simple deterministic pressure fields measured at the microphone locations. However, in practice stochastic signals must be considered so that the VAIM can be used in field experiments with complex noise sources. We deal with these complex signals in the "front end" signal processing. This front end provides for the construction of partial field holograms using theory that has been developed over the past 15 years.

V. FRONT END SIGNAL PROCESSING

We summarize the front end theory briefly here and the reader is directed to Refs. 14–19 for further information. This theory is used to expand the modes of operation of the VAIM. Although the VAIM can reconstruct intensity fields *without* reference transducers, due to the instantaneous measurement of the pressure data, and identify noise sources by display of the directional intensity vectors radiated from those sources, it is possible to further separate multiple noise sources by using partial field decomposition techniques described in Refs. 14–19. The partial field approach was developed by Hald[14] and has since found a multitude of industrial applications through the use of the spatial transformation of sound fields approach.

Assume M reference transducers recorded simultaneously and Fourier transformed to provide raw spectra represented by $\mathbf{X}(f) \equiv (X_1 X_2 \cdots X_M)^t$ (t is transpose) along with the 50 microphone raw spectra $\mathbf{P}(f) \equiv (P_1 P_2 \cdots P_{50})^t$. The reference transducers are generally attached to candidate (source) machines that are assumed to be random with Gaussian statistics, although they are not necessarily incoherent to one another.

The autospectral density of the ith microphone is given in the usual way by an ensemble average E of the raw spectra (using a long time series broken into shorter segments that are Fourier transformed) of the measured pressure

$$S_{p_i p_i}(f) = E[P_i^*(f) P_i(f)]. \tag{12}$$

Similarly, the cross-spectral density column vector to the ith microphone is given by

$$\mathbf{S}_{\mathbf{x}\mathbf{p}_i} \equiv E[\mathbf{X}^*(f) P_i(f)],$$

and the reference cross-spectral matrix is

$$\mathbf{S}_{\mathbf{x}\mathbf{x}} \equiv E[\mathbf{X}^* \mathbf{X}^{\mathbf{t}}].$$

Partial field decomposition techniques all decompose the autospectral density function of a microphone using an inner product of a partial field column vector $\mathbf{\Psi}_i$ of length M, $\mathbf{\Psi}_i = (\Psi_{1i} \Psi_{2i} \cdots \Psi_{Mi})^t$, that is, $S_{p_i p_i} = \mathbf{\Psi}_i^H \mathbf{\Psi}_i + S_{n_i n_i}$, the latter term being the noise. A partial field matrix Ψ is formed from these using $\mathbf{\Psi} = (\mathbf{\Psi}_1 \mathbf{\Psi}_2 \cdots \mathbf{\Psi}_{50})$. The rows of $\mathbf{\Psi}$ form M partial field holograms, each processed separately for reconstruction of M volumetric intensity fields.

There are two standard procedures for partial field decompositions: (1) the Cholesky method that yields

$$\mathbf{\Psi}_i = (\mathbf{T}^H)^{-1}\mathbf{S}_{\mathbf{xp}_i}, \tag{13}$$

where T (upper triangular matrix) results from the decomposition

$$\mathbf{S}_{\mathbf{xx}} = \mathbf{T}^{\mathbf{H}}\mathbf{T} \tag{14}$$

and (2) the principal component (SVD) method that results in

$$\mathbf{\Psi}_i = \mathbf{\Sigma}^{-1/2}\mathbf{U}^H\mathbf{S}_{\mathbf{xp}_i}, \tag{15}$$

where the singular value decomposition is given by

$$\mathbf{S}_{\mathbf{xx}} = \mathbf{U}\Sigma\mathbf{U}^{\mathbf{H}} \equiv \mathbf{U}\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2}\mathbf{U}^{\mathbf{H}} \tag{16}$$

where $\mathbf{U}$ is unitary and Σ is diagonal. It is important to note that in our research we found that the Cholesky method gave identical results to the signal conditioning approach provided in Bendat and Piersol[20] and in Ref. 21. A great deal of theory is provided in Bendat's book about the signal conditioning approach which can then be directly applied to understanding Cholesky decompositions.

One characteristic of the Cholesky method is that the partial fields are dependent upon the order of the columns and rows of $\mathbf{S}_{\mathbf{xp}}$.[20] Thus it is necessary to carry out some preanalysis to set up the order of the references, choosing the most significant reference at a particular frequency to form the first row. We determine significance by choosing references that have the largest coherence to the microphones, computing the average coherence $\overline{\gamma^2_{x_i p}}$ of the ith reference to the 50 microphones:

$$\overline{\gamma^2_{x_i p}} \equiv \frac{1}{50}\sum_{j=1}^{50} \gamma^2_{x_i p_j}, \tag{17}$$

where $\gamma^2_{x_i p_j} \equiv |S_{x_i p_j}|^2/(S_{x_i x_i} S_{p_j p_j})$ which allows us to rank the references, $x_m, x_n, \ldots, x_k$, with respect to average coherence for each frequency:

$$\overline{\gamma^2_{x_m p}} > \overline{\gamma^2_{x_n p}} > \cdots > \overline{\gamma^2_{x_k p}}. \tag{18}$$

Here x_m and x_n are the references with the first and second largest average coherence, respectively.

Given this ranking of references we reorder $\mathbf{S}_{\mathbf{xx}}$ and $\mathbf{S}_{\mathbf{xp}}$ (separate order for each frequency):

$$\mathbf{S}_{\mathbf{xx}} \equiv \begin{pmatrix} S_{x_m x_m} & S_{x_m x_n} & \cdots & S_{x_m x_k} \\ S_{x_n x_m} & S_{x_n x_n} & \cdots & S_{x_n x_k} \\ \vdots & \vdots & \ddots & \vdots \\ S_{x_k x_m} & S_{x_k x_n} & \cdots & S_{x_k x_k} \end{pmatrix}^{(M\times M)}, \tag{19}$$

$$\mathbf{S}_{\mathbf{xp}} \equiv \begin{pmatrix} S_{x_m p_1} & S_{x_m p_2} & \cdots & S_{x_m p_{50}} \\ S_{x_n p_1} & S_{x_n p_2} & \cdots & S_{x_n p_{50}} \\ \vdots & \vdots & \ddots & \vdots \\ S_{x_k p_1} & S_{x_k p_2} & \cdots & S_{x_k p_{50}} \end{pmatrix}, \tag{20}$$

so that the full reconstruction equation becomes

$$\mathbf{\Psi} = (\mathbf{T}^H)^{-1}\mathbf{S}_{\mathbf{xp}}. \tag{21}$$

The rows of $\mathbf{\Psi}^{(M\times 50)}$ form M separate holograms ranked in order of importance, each of which can be used to reconstruct the volumetric intensity at a given frequency by replacing $p_\infty(a, \theta_j, \phi_j, \omega)$ in Eq. (4) with one of the rows and computing the intensity vector Eq. (3) on a cubic lattice, displaying the results as in Sec. IV.

Figures of merit. Finally we consider three important measures that bracket the significance of the partial coherence method for any particular experiment. The first is called the partial (or virtual) coherence Γ^2_{ij} and represents what fraction of the signal energy $S_{p_j p_j}$ of the jth microphone is taken up by the partial field related to the ith reference:

$$\Gamma^2_{ij} \equiv \frac{|\Psi_{ij}|^2}{S_{p_j p_j}}. \tag{22}$$

This ratio must be less than or equal to one. A value of one indicates that the ith reference accounts for all of the measured signal and that all other references do not participate in the measured signal at the jth microphone (an unlikely occurrence). The average over all the microphones is

$$\Gamma^2_i \equiv \frac{1}{50}\sum_{j=1}^{50} \frac{|\Psi_{ij}|^2}{S_{p_j p_j}}, \tag{23}$$

so that Γ^2_i represents the fraction of the signal energy to all the microphones in the array due to the ith reference. Note that $\Gamma^2_i \leqslant 1$. The final figure of merit is the total fraction of all the signal energy received by the array that is represented by all the references used in the partial field decomposition:

$$\Gamma^2 \equiv \sum_{i=1}^{M} \Gamma^2_i. \tag{24}$$

For example, if this fraction Γ^2 is equal to 0.6 we know that 60% of the signal energy received by the microphones has been represented by the M partial field holograms. When this number is low we might conclude that the reference set was poorly chosen, and that the dominant acoustic sources have been overlooked in the experiment. We will show examples of these measures in the next section.

VI. EXPERIMENTAL APPLICATION

A 50 element spherical array was designed and constructed at the Naval Research Laboratory (NRL) and is shown in Fig. 8 in our in-air laboratory on the left and in flight on the right. The microphones were 1/8 in. diaphram electrets and were amplitude calibrated at 248 Hz. The phase response was not calibrated, but the 50 microphones were preselected for uniformity, resulting in a total phase variation of ±2° at 500 Hz. A light metal frame held the microphones in their positions that were determined by the quadrature algorithm, described earlier. The tips of the microphones formed a spherical array of radius 0.2 m and the signal cables were run down the vertical support rod shown in the figure. One aim of the design was for acoustic transparency.

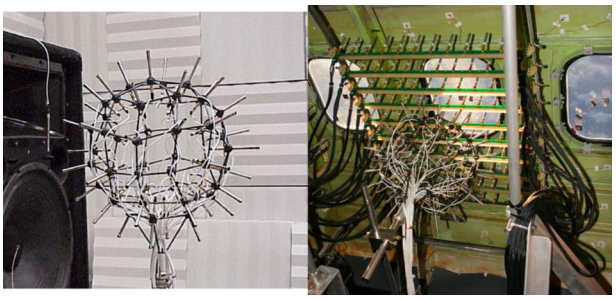


FIG. 8. (Color online) NRL 50 element spherical array in NRL test facility (left) and in flight inside the 757 aircraft (right) shown behind a conformal array (not discussed in this paper) with horizontal beams of elements in front of a fuselage window.

Initial experiments on a loudspeaker source in NRL's in-air facility were successful, but are not reported on here for the sake of brevity.

The picture on the right in Fig. 8 shows the array in the NASA Aries Boeing 757-200 airplane which is dedicated to flight test operations for the NASA Langley Research Center. The data set we report on here is a small subset of a much larger recent experiment using this aircraft.[22] In this experiment most of the seats in the aircraft were removed as well as the trim panels and insulation in four adjacent window sections as shown on the right in the figure. Also shown, but not reported on here, is a conformal nearly planar array of microphones located between the spherical array and the fuselage panel. The aircraft was flown at a tightly controlled altitude of 30 000 ft and speed of 0.8 M. Most of the bare panels and windows were monitored with accelerometers (a total of 31) placed at their centers and extra microphones (a total of 8) were dispersed throughout the aircraft interior. All of these monitors served as possible references for the VAIM as well as monitors of the physics of the vibration/radiation mechanics. All of the transducer channels were sampled simultaneously at 12 000 samples/s and recorded digitally on tape for processing in the laboratory. To provide a known source in the cabin Jacob Klos at NASA[22] created a point source using a long tube coupled to a loudspeaker as shown in Fig. 9. The outlet of the tube was at the center of the window, a few centimeters from its surface. The loudspeaker (shown in the figure on the right) was driven by a pseudorandom signal. This provided a guaranteed incoherent source to compare with the flow noise excited panels also radiating into the interior, and thus is a test of the front-end signal processing that should be able to separate multiple incoherent sources. We will show in the following the first three partial field holograms (rows of $\mathbf{\Psi}$) and volumetric intensity maps produced by our procedure at selected frequencies for this experiment. A second experiment was done with the point source turned off, and again partial field holograms and volumetric intensity maps were determined and compared with the first experiment. The objectives here were (1) to show that VAIM successfully identified the point source and the flow-noise induced panel source and (2) to show that the extraction of the flow-noise induced source was unique (unchanged with or without the point source).

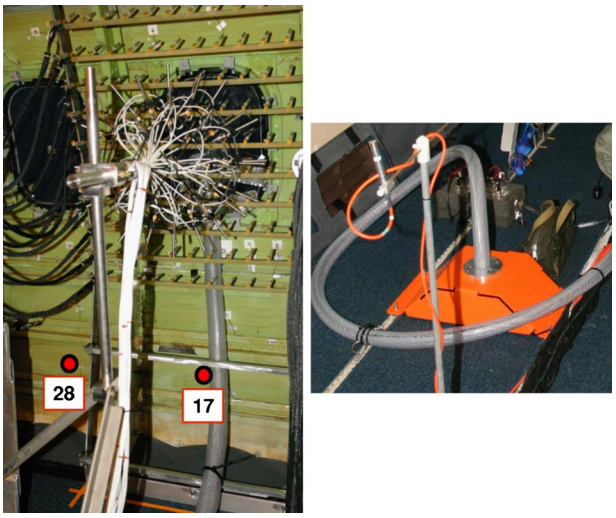


FIG. 9. (Color online) Acoustic point source installed inside the 757 cabin. The end of the vertical tube simulating a point source radiates sound near the center of the window (somewhat obscured by the conformal array). The tube is excited by a loudspeaker resting on the floor of the cabin that is coupled to the other end of the tube (shown on right). Red dots mark the position of two reference accelerometers.

A. Partial field holograms and intensity fields for point and flow noise sources

The 31 accelerometer references and the loudspeaker drive voltage (a pseudorandom bandpass signal between 500 and 1500 Hz) were all used as the reference set ($M=32$) to form the auto and cross-spectral density functions of Eqs. (19) and (20). A set of ensembles [estimates of $X_m(f)$ and $P_i(f)$] was created consisted of 1200 nonoverlapping segments of the time record, each containing 1920 points and Fourier transformed to provide frequency domain data. This was carried out for each microphone and each reference. Since the sample rate was 12 kHz, the total record was 192 s. The average coherence Eq. (17) was then computed from these spectra and the references were ordered as to importance using Eq. (18). This was carried out at each frequency (6.25 Hz bins) in the band of interest (500 – 1400 Hz).

For brevity we present the results at a single frequency bin and note that the results are typical. The autospectral density of the accelerometer on the center of one of the panels below the window (dot marked No. 17 in Fig. 9) showed a maximum in response at 732 Hz indicating a resonance of the panel excited by flow noise. At this frequency the first three partial holograms [the three top rows of the matrix in Eq. (21)] were constructed using the Cholesky method from Eqs (20), (19), (14), and (21). Each of the partial holograms were then processed into volumetric intensity fields which we now describe.

Each partial field, a row of Eq. (21), replacing p_∞ in Eq. (5), is integrated by Lebedev quadrature to determine the Fourier coefficients. These coefficients $\hat{P}^{\delta}_{mn}$ determined up to $n \leqslant 5$ are used to compute the components of the pressure and vector intensity given by Eqs. (1) and (2) leading to Eq. (3). Computations are repeated at a set of field points on an equal spaced cubic reconstruction grid with lattice size of

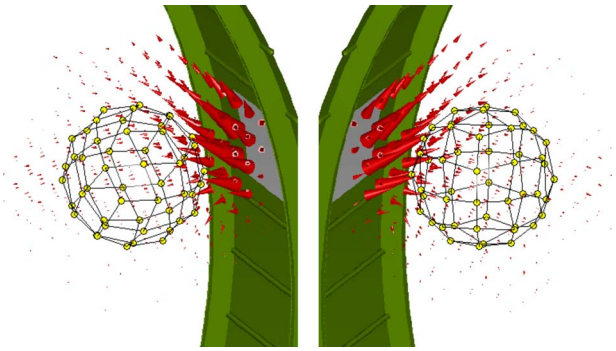


FIG. 10. (Color online) Point source on: 732 Hz first partial field. The first partial field contained 59% of the total signal energy according to Eq. (23). The total acoustic energy in vector field shown was 9.91 μJ.

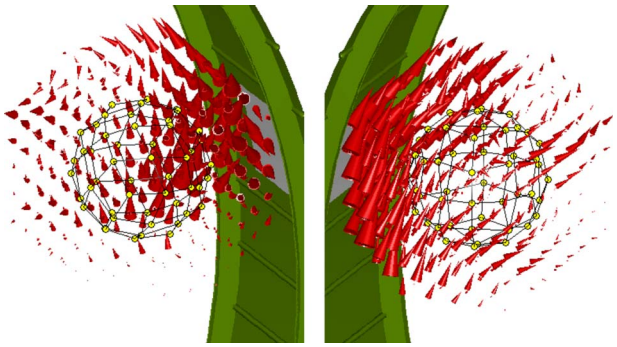


FIG. 12. (Color online) Point source on: 732 Hz second partial field. The second partial field contained 6% of the total signal energy. The total acoustic energy in vector field shown was 0.70 μJ.

8 cm. This Cartesian grid proved to be most effective at display of the intensity vector. The intensity computations are limited to within the sphere of radius 0.4 m, due to the high frequency errors outside this volume (see Fig. 2). The cutoff N for the series was determined by the regularization scheme described earlier, and in this case was $N=5$ (the maximum allowed) for each of the three partial fields.

The result for the reconstruction of the intensity vector at 732 Hz in the described volume for the first partial field is shown in Fig. 10. Two views of the volumetric field are shown and the square grey patch is a cartoon showing the location of the fuselage window (recall that the point source was at the center of this window). The dominant reference x_m [using Eq. (18)] turned out not surprisingly to be the drive voltage for the loudspeaker. Clearly the location of the real source is uncovered by the intensity display, hopefully a convincing demonstration of the success of the approach. The reference x_n related to the second partial field was the accelerometer shown in Fig. 9 (No. 17) located on the resonant panel below the window. This result is shown in Fig. 11. Although the exact location of the resonant panel source (located 0.9 m from the sphere center) is not clearly identifiable, the flow of the intensity clearly indicates a source below the window. Interestingly, the intensity flow diffracts along a path that is tangential to the fuselage surface. (Note we will show in the following that the same diffraction arises when the point source is turned off.) Finally, the third partial field is shown in Fig. 12, which is correlated to accelerometer No. 28 located at the center of the panel directly to the left of the resonant panel described earlier. This panel, identical in shape, was also resonant at 732 Hz, and the results for the volumetric intensity for this partial field are shown in Fig. 12. Note that the intensity appears to now come from the bottom left, that is, from the direction of the resonant panel, although possibly interrupted by the king frame that is located in between. The total amount of signal "energy" represented by the first three partial fields, given by Eq. (24) with $M=3$, was 84% indicating that only 16% of the total signal energy was unaccounted for.

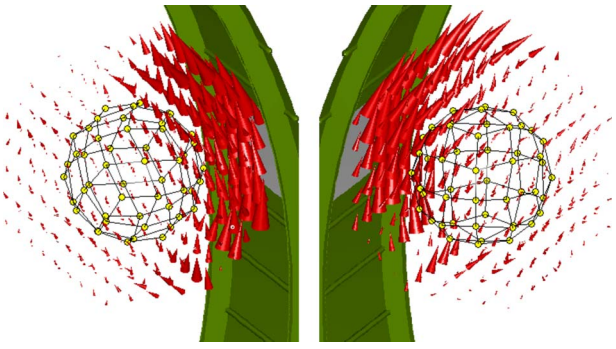


FIG. 11. (Color online) Point source on: 732 Hz second partial field. The second partial field contained 18% of the total signal energy according to Eq. (23). The total acoustic energy in vector field shown was 2.21 μJ.

Discussion. The reference (electrical drive) for the first partial field is incoherent ($\gamma^2 \approx 0.003$) to each of the two accelerometers (Nos. 17 and 26 in Fig. 9) and the Cholesky method works well to provide a unique decomposition and a separation of the incoherent fields. As long as the references are incoherent, the extraction is straightforward and successful as previous research has shown[14–16]. The difficultly here comes with a set of accelerometer references that respond to flow noise that in itself has some spatial coherence, and thus the individual panels on the fuselage section are *somewhat* coherent to one another. After the point source is removed, there are no incoherent sources left, since the various aircraft panels are weakly coupled. However, the Cholesky method forces incoherence (albeit artificial) here. That is, in view of the fact that the Cholesky method is identical to the signal conditioning approach[20] we can understand the Cholesky method as one that extracts the uncorrelated components of the reference signals in successive partial fields. In our case one can view the third partial field as an extraction out of the first two partial fields by creating a signal that is orthogonal or incoherent to them and in the process of doing this creates a volumetric intensity field that appears to come from the panel to the left. The raw accelerometer signals themselves are correlated to each other, in our case with a coherence of 0.33. This orthogonalization is somewhat artificial as it depends on the order of the references chosen before the Cholesky decomposition is carried out. We have chosen to order based on descending average coherences using Eq. (18) but other ordering methods may be used. In other words, when correlated references exist in the reference set, one is not guaranteed a one to one relationship between the partial field vector intensity and the orthogonalized reference. The intensity field may appear to come from one of the

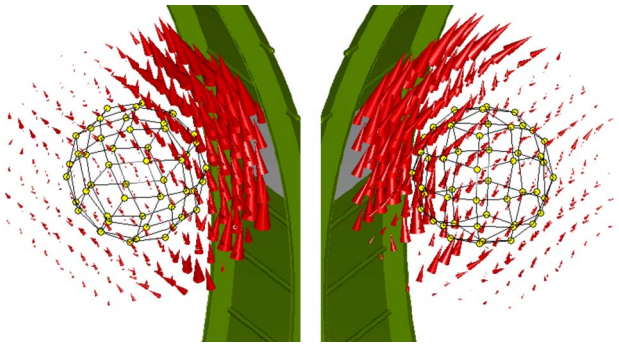


FIG. 13. (Color online) Point source is turned off. Result for the first partial field at 732 Hz. This field contained 45% of the total signal energy. The total energy in vector field shown was 2.53 μJ.

other closely correlated references, instead. However, this is not the case if the references are uncorrelated and the order of the references will not change the results.

B. Repeat experiment with point source turned off

The experiment described in Sec. VI A was repeated about 20 min later with the acoustic point source turned off, and the results were computed in the same frequency bin for comparison. An effort was made to keep the aircraft speed and altitude unchanged. The resulting first partial field is shown in Fig. 13 and the maximum coherence to the microphones determined by Eq. (18) was accelerometer No. 17, the same reference found for the second partial field when the source was on, in the previous experiment. A close comparison to Fig. 11 reveals an important result. The flow fields are identical. This gives creditability to the conclusion that the vector field for the point source was accurately separated from the other incoherent sources in the previous experiment. Furthermore, this agreement shows the robustness of the approach.

Finally we present the results for the second partial field when the source is off, shown in Fig. 14. The reference associated with the field was accelerometer No. 28, and a direct comparison with the previous result Fig. 11, also correlated to the same accelerometer, can be made. Again a remarkable result occurs—the intensity vector fields are nearly identical, again showing an intensity field that appears to emanate from an area near the reference accelerometer (see Fig. 9) on the panel to the left. Again the robustness of the field extraction approach is demonstrated, even in the face of the coherence in the vibration between the two panels involved.

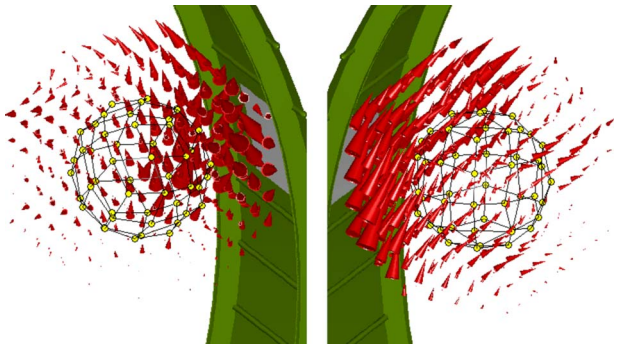


FIG. 14. (Color online) Point source is turned off. Result for the second partial field at 732 Hz. This field contained 18% of the total signal energy. The total energy in vector field shown was 0.81 μJ.

We have shown only 2 of the 32 partial fields computed. The remaining partial fields represent a small fraction of the total signal energy with less than 2.5% each, and appear to be nearly buried in the noise.

VII. CONCLUSIONS

The work presented here is a small fraction of the theory and development of the VAIM, and future papers will discuss more details such as SVD partial field decomposition results, use of self-referencing to a set of microphones in the array without external references and 1/3 octave band results as well as new generation designs of the array. The aim in the design and the algorithm behind the intensity computations is for real time display, so that the system would provide nearly instantaneous results when used with a laptop in the field. With this the VAIM can scan a large space making volumetric reconstructions on the fly. We believe that this will provide an invaluable tool for source identification both in air and underwater.

ACKNOWLEDGMENTS

Work supported by the Office of Naval Research. Experimental work and data analysis was supported by NASA Langley and The Boeing Company provided critical experiment support.

[1]T. J. Schultz, "Acosutic wattmeter," J. Acoust. Soc. Am. **28**, 693–699 (1956).
[2]F. J. Fahy, *Sound Intensity* (Elsevier Applied Science, London, 1989).
[3]S. Nagata, K. Furihata, T. Wada, K. Asano, and T. Yanagisawa, "A three-dimensional sound intensity measurement system for sound source identification and sound power determination by ln models," J. Acoust. Soc. Am. **118**, 3691–3705 (2005).
[4]R. D. Marciniak, "A nearfield, underwater measurement system," J. Acoust. Soc. Am. **66**, 955–964 (1979).
[5]G. Weinreich and E. B. Arnold, "Method for measuring acoustic radiation fields," J. Acoust. Soc. Am. **68**, 404–411 (1980).
[6]M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," J. Acoust. Soc. Am. **118**, 3094–3103 (2005).
[7]*Spherical Near-Field Antenna Measurements*, edited by J. E. Hansen (Peregrinus for IEE, London, 1988).
[8]R. C. Wittmann, "Spherical wave operators and the translation formulas," IEEE Trans. Antennas Propag. **36**, 1078–1087 (1988).
[9]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London, 1999).
[10]V. I. Lebedev, "Values of the nodes and weights of ninth to seventeenth order Gauss-markov quatrature formulai invariant under the octahedron group with inversion," Zh. Vychisl. Mat. Mat. Fiz. **15**(1), 48–54 (1975).
[11]V. I. Lebedev, "Quadratures on a sphere," USSR Comput. Math. Math. Phys. **16**(2), 10–24 (1976).
[12]E. G. Williams, "Regularization methods for near-field acoustical holography," J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[13]P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems* (Siam, Philadelphia, 1998).
[14]J. Hald, "STSF—A unique technique for scan-based nearfield acoustical holography without restriction on coherence," Technical Report, B&K Technical Review, No. 1, 1989.
[15]D. L. Hallman and J. S. Bolton, "Multi-reference nearfield acoustical holography," in Proceedings Inter-noise 1992, Toronto, Canada, pp. 1165–1170.
[16]D. L. Hallman and J. S. Bolton, "A comparison of multi-reference nearfield acoustical holography procedures," in Proceedings Noise-Con

1994, Ft. Lauderdale, FL, pp. 929–934.
[17]H.-S. Kwon and J. S. Bolton, "Partial field decomposition in nearfield acoustical holography by the use of singular value decomposition and partial coherence procedures," in Proceedings Noise-Con 1998, pp. 649–654.
[18]H.-S. Kwon, Y.-J. Kim, and J. S. Bolton, "Compensation for source nonstationarity in multireference, scan-based near-field acoustical holography," J. Acoust. Soc. Am. **113**, 360–368 (2003).
[19]K.-U. Nam and Y.-H. Kim, "A partial field decomposition algorithm and its examples for near-field acoustic holography," J. Acoust. Soc. Am. **116**, 172–185 (2004).
[20]J. S. Bendat and A. G. Piersol, *Random Data Analysis and Measurement Procedures*, 3rd ed. (Wiley, New York, 2000).
[21]M. A. Tomlinson, "Partial source discrimination in near field acoustic holography," Appl. Acoust. **57**, 243–261 (1999).
[22]J. Klos *et al.*, "Comparison of different measurement technologies of the in-flight assessment of radiated acoustic intensity," in Proceedings of Noise-Con 2005, Minneapolis, MN.

An internal streaming instability in regenerators

J. H. So, G. W. Swift, and S. Backhaus[a)]
Condensed Matter and Thermal Physics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 24 April 2006; accepted 7 July 2006)

Various oscillating-wave thermodynamic devices, including orifice and feedback pulse tube refrigerators, thermoacoustic-Stirling hybrid engines, cascaded thermoacoustic engines, and traditional Stirling engines and refrigerators, utilize regenerators to amplify acoustic power (engines) or to pump heat acoustically up a temperature gradient (refrigerators). As such a regenerator is scaled to higher power or operated at lower temperatures, the thermal and hydrodynamic communication transverse to the acoustic axis decreases, allowing for the possibility of an internal acoustic streaming instability with regions of counterflowing streaming that carry significant heat leak down the temperature gradient. The instability is driven by the nonlinear flow resistance of the regenerator, which results in different hydrodynamic flow resistances encountered by the oscillating flow and the streaming flow. The instability is inhibited by several other mechanisms, including acoustically transported enthalpy flux and axial and transverse thermal conduction in the regenerator solid matrix. A calculation of the stability limit caused by these effects reveals that engines are immune to a streaming instability while, under some conditions, refrigerators can exhibit an instability. The calculation is compared to experimental data obtained with a specially built orifice pulse tube refrigerator whose regenerator contains many thermocouples to detect a departure from transverse temperature uniformity.
[DOI: 10.1121/1.2259776]

PACS number(s): 43.25.Nm, 43.35.Ud [RR] Pages: 1898–1909

I. INTRODUCTION

The regenerator[1] is a critical component of many oscillating-wave thermodynamic devices including orifice[2] and feedback[3] pulse tube refrigerators, thermoacoustic-Stirling hybrid engines,[4–6] cascade thermoacoustic engines,[7] and traditional Stirling engines and refrigerators.[8] In combination with the temperature gradient, the regenerator amplifies acoustic power in engines and it allows the acoustic wave to pump heat in refrigerators.

To analyze regenerators, researchers have generally assumed one-dimensional behavior, i.e., that everything is independent of transverse coordinates y and z but depends only on the axial position x. However, awareness is growing that nontrivial multidimensional behavior and an associated penalty in performance can occur in the regenerators of pulse-tube and Stirling cryocoolers. Such multidimensional behavior has been described for transverse variation in the hydraulic radius of regenerators[9] and for three identical regenerators operated in parallel.[10]

Most regenerators are made of stacked screen or other tortuous porous media whose nonlinear flow resistance generates a complicated interaction between the oscillating and steady flows. The result is that the oscillating and steady flows encounter different flow resistances which depend on temperature in different ways. A small spatial perturbation in the mean temperature transverse to the acoustic axis leads to different flow resistances for the oscillating and streaming flows in different regions of the regenerator, resulting in a small circulating streaming flow, as illustrated in Fig. 1. Whether the streaming flow acts to amplify or suppress the original perturbation depends on many variables, including the direction of the streaming flow and the temperatures of the heat exchangers at the two ends of the regenerator. The original perturbation might be amplified by this effect if $T_w < T_a$ and suppressed if $T_w > T_a$ where T_a is ambient temperature and T_w is the working temperature (i.e., the hot temperature for an engine and the cold temperature for a refrigerator). Even if the resulting streaming acts to amplify the original perturbation, other effects, such as thermal conduction or acoustic enthalpy transport, will attempt to suppress it.

As such an engine or refrigerator is scaled to higher power, the cross-sectional area of its regenerator must increase proportionally to keep its design near a thermodynamic optimum, but its length remains approximately constant. At some point, the transverse thermal and hydrodynamic communication within the regenerator becomes so weak that it cannot counteract the streaming caused by a transverse mean-temperature perturbation, and an instability may arise. Whether or not this occurs depends on the balance among all these effects. If an instability does occur, the streaming flow carries additional energy flux down the temperature gradient, resulting in reduced efficiency in the engine or refrigerator.

Using a linear stability analysis, we add to a streaming-free solution a perturbation that grows or decays exponentially in time. By comparing different terms that either drive or inhibit the instability, we show when the combination of these mechanisms results in instability, i.e., an exponentially

[a)]Electronic mail: backhaus@lanl.gov

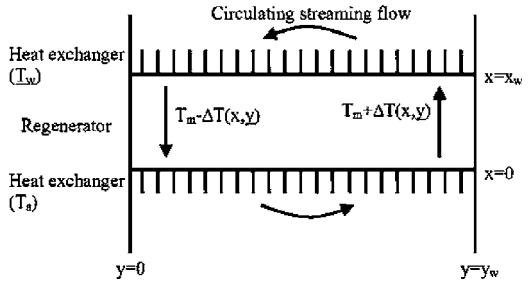


FIG. 1. Schematic drawing of an internal streaming flow in a regenerator. The x axis is the acoustic direction, and the y axis is the long transverse dimension. The extent of the regenerator along the z axis (not shown) is assumed to be small enough so that the acoustic and streaming flows are uniform in that direction. The arrows represent steady streaming flows induced by a small left-right temperature difference.

growing perturbation. The instability is experimentally investigated using a specially designed and heavily instrumented orifice pulse tube refrigerator. Thermocouples placed inside the refrigerator's regenerator detect the presence of streaming through measurements of the mean temperature distribution. When the acoustic power flux is high and the total energy flux and cold-end temperature are low, a mean temperature variation about the axial midplane of the regenerator is detected, indicating the presence of a streaming instability. The values of the acoustic power flux, total energy flux, and cold-end temperature at the threshold of instability are found to be in only qualitative agreement with theory.

II. THEORY

To search for an internal streaming instability analytically, we will combine a thermoacoustic analysis[11] through second order in the acoustic amplitude with a linear perturbation analysis.[12] Carrying the thermoacoustic analysis to second order includes the lowest-order, time-averaged energy and mass-flux effects, which can interact in the perturbation analysis to produce exponential growth of both. A similar approach was taken in an investigation of a related instability.[13] First, a simplified model of the acoustics in the regenerator is used to establish relationships among streaming-free first-order variables. Then the important second-order streaming-free quantities—the time-averaged energy flux and the second-order velocity—are derived from the first-order quantities. Thus far, all variables depend on the axial coordinate in the regenerator but are independent of transverse coordinates. Then, a perturbation is added to the solution that depends on the transverse coordinates, and the analysis shows that it can either grow or decay exponentially with time, depending on the competition among a number of stabilizing and destabilizing effects. To make the mathematics as simple as possible, we make a number of restrictive and potentially unrealistic assumptions. We neglect any time-phase difference between first-order pressure and first-order velocity throughout the regenerator. We assume that the velocity perturbation is nonzero only parallel to the acoustic axis. We neglect oscillations at any frequency but the fundamental. We arbitrarily choose a mathematically simple functional form for the axial spatial dependence of the perturbation. We hope that some or all of these restrictive assumptions can be avoided in future work.

A. Streaming-free solution

Using a variation on the usual acoustic expansion and time-harmonic notation,[11] the streaming-free solution can be written

$$p(x,t) = p_m + p_{1R}(x)\cos\omega t - p_{1I}(x)\sin\omega t + p_{20}(x) + p_{22R}(x)\cos 2\omega t - p_{22I}(x)\sin 2\omega t + \cdots, \quad (1)$$

$$u(x,t) = u_{1R}(x)\cos\omega t - u_{1I}(x)\sin\omega t + u_{20}(x) + u_{22R}(x)\cos 2\omega t - u_{22I}(x)\sin 2\omega t + \cdots, \quad (2)$$

$$T(x,t) = T_m(x) + T_{1R}(x)\cos\omega t - T_{1I}(x)\sin\omega t + T_{20}(x) + T_{22R}(x)\cos 2\omega t - T_{22I}(x)\sin 2\omega t + \cdots, \quad (3)$$

where p, u, and T are the working-gas pressure, x component of velocity, and temperature, respectively. Here and throughout this paper, variables with the subscript $m, 1R, 1I, 22R, 22I$, or 20 are real. The fundamental angular frequency of the oscillation is ω, t is time, and x is the coordinate along the axis of the regenerator, with $x=0$ at the ambient face and $x=x_w$ at the working-temperature face. Variables describing the gas, such as $u_{1R}(x)$ and $T_{1R}(x)$, represent local spatial averages over small volumes of gas including many pores, not microscopic values needed to describe spatial dependencies within the tortuous geometry of a single pore. However, the time-averaged energy-flux density $\dot{h}$, which describes energy flux through both the gas and solid, is a local spatial average over a small volume of both gas and solid including many pores. Thus, if A is the regenerator's cross-sectional area and ϕ is its volume porosity, the total volume flow rate is ϕAu and the total energy flux is $A\dot{h}_x$.

The representation in Eqs. (1) through (3) is equivalent to the usual complex notation

$$p(x,t) = p_m + \mathrm{Re}[p_1(x)e^{i\omega t}] + p_{20}(x) + \mathrm{Re}[p_{22}(x)e^{i2\omega t}] + \cdots, \quad (4)$$

$$u(x,t) = \mathrm{Re}[u_1(x)e^{i\omega t}] + u_{20}(x) + \mathrm{Re}[u_{22}(x)e^{i2\omega t}] + \cdots, \quad (5)$$

$$T(x,t) = T_m(x) + \mathrm{Re}[T_1(x)e^{i\omega t}] + T_{20}(x) + \mathrm{Re}[T_{22}(x)e^{i2\omega t}] + \cdots, \quad (6)$$

with, e.g., $p_1(x)=p_{1R}(x)+ip_{1I}(x)$ and with variables having subscript 1 or 22 being complex. At each step in the calculation, the choice of notation will be determined by convenience.

Using the notation in Eqs. (4)–(6), the continuity equation expanded to first order and averaged over the small scale is[14]

$$\frac{du_1}{dx} - \frac{1}{T_m}\frac{dT_m}{dx}u_1 = -\frac{i\omega\rho_1}{\rho_m}. \tag{7}$$

In effective regenerators, the excellent thermal contact between the working gas and the regenerator solid renders the density oscillations nearly isothermal, so that $\rho_1 = (\rho_m/p_m)p_1$. If the gas volume in the regenerator is small enough that its compressibility can be ignored, the ρ_1 term on the right-hand side (RHS) of Eq. (7) can be neglected, so that

$$u_1(x) = u_1(0)\frac{T_m(x)}{T_a}. \tag{8}$$

The subscripts a and w refer to mean variables evaluated at the ambient end $x=0$ and the working end $x=x_w$ of the regenerator, respectively. Without loss of generality, the phase of $u_1(0)$ can be chosen so that

$$u_{1R}(x) = u_{1R}(0)\frac{T_m(x)}{T_a} > 0, \tag{9}$$

$$u_{1I}(x) = 0. \tag{10}$$

Under typical operating conditions, p_1 and u_1 at $x=0$ are nearly in phase so that

$$p_{1R}(x) > 0, \tag{11}$$

$$p_{1I}(x) = 0. \tag{12}$$

In general, the second-order mass-flux density is given by

$$\dot{m}_{20} = \mathrm{Re}[\rho_1 u_1^*]/2 + \rho_m u_{20} = \frac{\rho_m}{2_{pm}} p_{1R}u_{1R} + \rho_m u_{20}, \tag{13}$$

where the superscript * stands for complex conjugation. We have again assumed that the density oscillations are nearly isothermal. In the streaming-free state $\dot{m}_{20}=0$, and if the typically small variation in p_{1R} through the regenerator is ignored, u_{20} can be expressed

$$u_{20}(x) = -\frac{p_{1R}u_{1R}(x)}{2p_m} = u_{20}(0)\frac{T_m(x)}{T_a}, \tag{14}$$

showing that u_{1R} and u_{20} have approximately the same spatial dependence.

If the perimeter of the regenerator is well insulated, the time-averaged second-order energy-flux density $\dot{h}_{2,x}$ is independent of both x and t in the streaming-free steady state, indicating there is no buildup of energy inside the regenerator.[11] A perfect regenerator would have $\dot{h}_{2,x}=0$, and in realistic regenerators $\dot{h}_{2,x}$ is small—much smaller than the acoustic intensity. The small T_1 that is neglected above in the analysis of $u_1(x)$ contributes significantly to this small energy-flux density and cannot be neglected in $\dot{h}_{2,x}$.

B. Perturbed solution

An exponentially growing or decaying perturbation is added to the streaming-free solution derived in the previous section, so that the complete solution is of the form

$$p(x,y,t) = p_m + p_{1R}\cos\omega t + p_{20}(x) + [\delta p_{1R}(x,y)\cos\omega t - \delta p_{1I}(x,y)\sin\omega t + \delta p_{20}(x,y)]e^{\epsilon t}, \tag{15}$$

$$u(x,y,t) = u_{1R}(x)\cos\omega t + u_{20}(x) + u_{22R}(x)\cos(2\omega t) - u_{22I}(x)\sin(2\omega t) + [\delta u_{1R}(x,y)\cos\omega t - \delta u_{1I}(x,y)\sin\omega t + \delta u_{20}(x,y) + \delta u_{22R}(x)\cos(2\omega t) - \delta u_{22I}(x)\sin(2\omega t)]e^{\epsilon t}, \tag{16}$$

$$T(x,y,t) = T_m(x) + \delta T_m(x,y)e^{\epsilon t} + T_{1R}(x,y)\cos\omega t - T_{1I}(x,y)\sin\omega t + [\delta T_{1R}(x,y)\cos\omega t - \delta T_{1I}(x,y)\sin\omega t]e^{\epsilon t}, \tag{17}$$

$$\rho(x,y,t) = \text{similar to } T, \tag{18}$$

$$\dot{h}_x(x,y,t) = \dot{h}_{2,x} + \delta\dot{h}_x(x,y)e^{\epsilon t}, \tag{19}$$

$$\dot{h}_y(x,y,t) = \delta\dot{h}_y(x,y)e^{\epsilon t}, \tag{20}$$

where we only allow variation in one of the coordinates transverse to the acoustic axis, i.e., y and we have carried the expansions only as far as the terms that we will discuss later. In this calculation, we have in mind a regenerator with one long (y) and one short (z) transverse dimension such that variations in temperature due to a perturbation are more likely to take hold in the long dimension. The perturbation includes both oscillating and nonoscillating terms. The oscillating perturbations are assumed to have the same frequency as the corresponding terms in the streaming-free solution and amplitudes that change slowly, but exponentially, in time compared to the acoustic period, i.e., $|\epsilon| \ll \omega$. This two-time-scale approach allows the explicit separation of the slow change of the instability from the rapid acoustic oscillations. Nonoscillating terms also change exponentially in time with the same time constant as the amplitudes of the oscillating terms. We assume $\delta\omega$ is zero. For a refrigerator this can be regarded as a simple consequence of how the system is driven, e.g., by a linear motor at fixed frequency. For a thermoacoustic-Stirling hybrid engine, cascaded thermoacoustic engine, or Stirling engine, one can similarly assume that the complex load impedance is deliberately varied to keep ω fixed.

Substituting the full solution of Eqs. (15)–(20) into the continuity equation, expanding to first order in the perturbation, taking $\delta\rho_1=0$ for the reasons described below Eq. (7), utilizing the results of Eq. (8), and recalling that we are neglecting transverse flow, we find

$$\frac{d\delta u_1}{dx} - \frac{1}{T_m}\frac{dT_m}{dx}\delta u_1 = \frac{u_1(0)}{T_a}\left[\frac{d\delta T_m}{dx} - \frac{\delta T_m}{T_m}\frac{dT_m}{dx}\right]. \tag{21}$$

The solution to this differential equation for δu_1 is given by

$$\delta u_1(x,y) = \delta u_1(0,y)\frac{T_m(x)}{T_a} + \frac{u_1(0)}{T_a}\delta T_m(x,y), \tag{22}$$

where the boundary condition $\delta T_m(0,y)=0$ has been used. This boundary condition implies that the heat transfer provided by the heat exchanger at $x=0$ is sufficient to hold the temperature perturbation at zero at the heat exchanger despite the streaming. (Later, we will use the same condition at $x=x_w$.)

Equation (22) can be broken up into real and imaginary parts

$$\delta u_{1R}(x,y) = \delta u_{1R}(0,y)\frac{T_m(x)}{T_a} + u_{1R}(0)\frac{\delta T_m(x,y)}{T_a}, \tag{23}$$

$$\delta u_{1I}(x,y) = \delta u_{1I}(0,y)\frac{T_m(x)}{T_a}. \tag{24}$$

To determine $\delta u_{1R}(0,y)$ and $\delta u_{1I}(0,y)$ in Eqs. (23) and (24), we must consider the first-order Navier-Stokes equation. The full solution given in Eqs. (15) through (17) is substituted into the spatially averaged Navier-Stokes equation for screens,[14] and the result is expanded to first order in the perturbation for all terms (dropping terms that are obviously of zero order in the perturbation). Additionally, the linear term is expanded to acoustic first order, and the nonlinear term is expanded to acoustic second order. This intermediate result is

$$\begin{aligned}
&-\frac{\partial}{\partial x}[\delta p_{1R}\cos\omega t - \delta p_{1I}\sin\omega t] \\
&= \frac{c_1}{8r_h^2}[\mu_m\delta u_{1R}\cos\omega t - \mu_m\delta u_{1I}\sin\omega t + \delta\mu_m u_{1R}\cos\omega t] \\
&\quad + \frac{c_2}{2r_h}\rho_m[u_{1R}\cos\omega t + \delta u_{1R}\cos\omega t - \delta u_{1I}\sin\omega t]|u_{1R}\cos\omega t \\
&\quad + \delta u_{1R}\cos\omega t - \delta u_{1I}\sin\omega t| + \frac{c_2}{2r_h}\delta\rho_m u_{1R}^2\cos\omega t|\cos\omega t|,
\end{aligned} \tag{25}$$

where c_1 and c_2 parametrize the regenerator flow resistance,[14] r_h is the hydraulic radius,[14] and μ is the gas viscosity. The presence of the absolute value sign in the nonlinear term complicates the expansion by forcing the temporary retention of certain terms that may appear to be of higher order than we later keep. To isolate δp_{1R} or δp_{1I}, Eq. (25) is multiplied by $(\cos\omega t)/\pi$ or $(\sin\omega t)/\pi$ and integrated with respect to ωt from 0 to 2π, being careful to split the integration into two at the zero crossings of the terms inside the absolute value signs. To lowest order in δu_{1I}, the result for δp_{1I} is

$$-\frac{\partial\delta p_{1I}}{\partial x} = \left[\frac{c_1\mu_a}{8r_h^2}\left(\frac{T_m}{T_a}\right)^b + \frac{4c_2}{3\pi r_h}\rho_a u_{1R}(0)\right] \times \frac{T_m(x)}{T_a}\delta u_{1I}(0,y), \tag{26}$$

where b accounts for the temperature dependence of viscosity via $\mu(T)=\mu_a(T/T_a)^b$. Assuming that the spaces at the ends of the regenerator are open enough so that δp_{1I} and δp_{1R} cannot have any y dependence at $x=0$ or x_w, i.e., $\delta p_{1I}(0,y)=\delta p_{1I}(x_w,y)=0$, the integral of Eq. (26) from 0 to x_w must be zero. The terms inside of the square brackets are always positive, and therefore $\delta u_{1I}(0,y)=0$ and $\delta u_{1I}(x,y)=0$. This result, combined with the assumption that $\delta p_{1I}(0,y)=0$ and Eq. (26), shows that $\delta p_{1I}(x,y)=0$.

With the perturbation greatly simplified by the conclusion that $\delta u_{1I}(x,y)=0$, δp_{1R} is isolated from Eq. (25) yielding

$$\begin{aligned}
-\frac{8r_h^2}{c_1\mu_a}\frac{\partial\delta p_{1R}}{\partial x} &= \delta u_{1R}(0)\left[\left(\frac{T_m}{T_a}\right)^{1+b} + 2\Gamma\left(\frac{T_m}{T_a}\right)\right] \\
&\quad + \frac{\delta T_m}{T_a}u_{1R}(0)\left[(1+b)\left(\frac{T_m}{T_a}\right)^b + \Gamma\right]
\end{aligned} \tag{27}$$

to lowest order in δu_{1R}, where $\Gamma=8c_2N_{R,a}/3\pi c_1$ and $N_{R,a}$ is the Reynolds number $N_R=4\rho_m u_{1R}r_h/\mu_m$ evaluated at the ambient end of the regenerator. Next, Eq. (27) is integrated with respect to x from 0 to x_w under the following assumptions:

$$\delta p_{1R}(0,y) = \delta p_{1R}(x_w,y) = 0, \tag{28}$$

$$T_m(x) = T_a + (T_w - T_a)x/x_w, \tag{29}$$

$$\delta T_m(x,y)/T_a = f(y)\sin(\pi x/x_w), \tag{30}$$

where $f(y)$ is a yet unknown function of y. Equation (28) is again a statement that the space at the ends of the regenerator is open enough so that it cannot support a transverse pressure gradient. Equation (29) approximates the streaming-free mean temperature profile as linear. The selection of the sinusoidal behavior of ΔT_m in Eq. (30) is not based on a rigorous solution of the governing equations. Instead, this particular functional form simplifies the computation, provides a close representation of the experimentally measured mean temperature deviation when nonzero acoustic streaming is present,[15] and satisfies the boundary conditions $\delta T_m(0,y)=\delta T_m(x_w,y)=0$. The result of integrating Eq. (27) from $x=0$ to $x=x_w$ is

$$\delta u_{1R}(0,y) = -\frac{f(y)u_{1R}(0)}{\pi}\frac{2\Gamma+\beta(\tau,b)}{(\tau^{b+2}-1)/[(\tau-1)(b+2)]+(\tau+1)\Gamma}, \tag{31}$$

where $\tau=T_w/T_a$ and

$$\beta(\tau,b) = (1+b)\int_0^\pi [1+(\tau-1)z/\pi]^b\sin z\,dz. \tag{32}$$

This result combined with Eq. (23) and the assumed form of $\delta T_m(x,y)$ given in Eq. (30) gives a complete solution for $\delta u_{1R}(x,y)$ in terms of the streaming-free variables and the unknown function $f(y)$.

Next, we consider the streaming-velocity perturbation $\delta u_{20}(x,y)$ by investigating the second-order, time-average mass-flux density perturbation $\delta\dot{m}_{20}$. In contrast to $\dot{m}_{20}$, it is not obvious from the outset that $\delta\dot{m}_{20}$ is a constant throughout the regenerator. Considering only $\delta\dot{m}_{20}$, conservation of energy to linear order in the perturbation yields

$$(1-\phi)\rho_s c_s \epsilon \delta T_m = -\phi c_p \frac{\partial (T_m \delta \dot{m}_{20})}{\partial x}, \tag{33}$$

where the heat capacity of the gas $\phi\rho_m c_p$ has been ignored relative to the heat capacity of the regenerator screen $(1-\phi)\rho_s c_s$. As the local mean temperature of the regenerator changes in time, the local gas density will also change, i.e., $\epsilon\delta T_m \approx -\epsilon T_m \delta\rho_m/\rho_m$. Conservation of mass at linear order in the perturbation requires that the change in local density is fulfilled by a gradient in $\delta\dot{m}_{20}$

$$-\epsilon\delta\rho_m = \frac{\partial \delta\dot{m}_{20}}{\partial x}. \tag{34}$$

Combining Eqs. (33) and (34) and the discussion in between yields

$$\frac{1}{\delta\dot{m}_{20}}\frac{\partial\delta\dot{m}_{20}}{\partial x} = -\varepsilon\frac{1}{T_m}\frac{\partial T_m}{\partial x}\Bigg/(1+\varepsilon), \tag{35}$$

where $\varepsilon = \phi\rho_m c_p/(1-\phi)\rho_s c_s$. In efficient regenerators, $\varepsilon \ll 1$, and therefore, the change in $\delta\dot{m}_{20}$ with x from one side of the regenerator to the other is small.

Using Eq. (13) as the general definition of $\dot{m}_{20}$, and with $\dot{m}_{20}=0$ in the steady state, one expression for $\delta\dot{m}_{20}$ is

$$\delta\dot{m}_{20} = \frac{\rho_m}{2p_m} p_{1R} u_{1R}\left[\frac{\delta u_{1R}}{u_{1R}} - \frac{\delta u_{20}}{u_{20}}\right], \tag{36}$$

where the $\delta p_{1R}/p_{1R}$ term has been ignored because Eqs. (27), (30), and (31) show that it is smaller than the other two terms by a factor $[p_{1R}(0)-p_{1R}(x_w)]/p_{1R}(0)$. Recognizing that $\delta\dot{m}_{20}$ and the prefactor on the RHS of Eq. (36) are approximately independent of x, and using Eqs. (8) and (22), it can be shown that

$$\frac{\delta u_{20}(x,y)}{u_{20}(x)} = \frac{\delta u_{20}(0,y)}{u_{20}(0)} + \frac{\delta T_m(x,y)}{T_m(x)}. \tag{37}$$

To complete the solution for $\delta u_{20}(x,y)$, its value at $x=0$ still needs to be determined from the second-order, time-averaged Navier-Stokes equation. We begin by substituting the full solution from Eqs. (15)–(17), but with $\delta u_{1I}=0$ and $\delta p_{1I}=0$ as established above, into the time-dependent, spatially averaged Navier-Stokes equation for screens,[14] expanding to first order in the perturbation, expanding the linear term to second order in the acoustic amplitude and the nonlinear term to third order, and time averaging. An intermediate result is

$$-\frac{\partial\delta p_{20}}{\partial x} = \frac{c_1\mu_m u_{20}}{8r_h^2}\left[\frac{\delta u_{20}(0)}{u_{20}(0)} + (1+b)\frac{\delta T_m}{T_m} + \frac{4c_2 N_R}{3\pi c_1}\left(\frac{3\delta u_{20}(0)}{u_{20}(0)} - \frac{\delta u_{1R}(0)}{u_{1R}(0)} + \frac{\delta T_m}{T_m} - \frac{2\delta p_{1R}}{p_{1R}} \text{ plus } u_{22} \text{ and } \delta u_{22} \text{ terms}\right)\right] \tag{38}$$

For the reason given below Eq. (36), the $\delta p_{1R}/p_{1R}$ can be dropped. However, the terms proportional to u_{22} and δu_{22} are more problematic. By assuming the density oscillations at 2ω are still isothermal, it can be, shown that $u_{22}(x)=u_{22}(0)[T_m/T_a]$, but $u_{22}(0)$ is still undetermined. In an orifice pulse tube refrigerator (OPTR), adjustments to the driving piston's motion could be made to force $p_{22}(0)=p_{22}(x_w)$, but, this may not ensure that $u_{22}(0)$ is actually zero because the nonlinearities in the regenerator, at interfaces between components, and due to finite motion of the piston itself are continually generating 2ω terms throughout the system. Without knowledge of u_{22}, there is no way to calculate δu_{22}. The only way to close the equations without introducing a whole host of additional equations that address these nonlinearities is to simply drop the u_{22} and δu_{22}.

To extract $\delta u_{20}(0)$, Eq. (38) is integrated from $x=0$ to x_w. Assuming that the spaces at the two ends of the regenerator are open enough to disallow a transverse pressure gradient, the integral of the LHS of Eq. (38) is zero. Using Eqs. (14) and (29)–(31), and the temperature dependence of μ_m, the result of integrating Eq. (38) is

$$\delta u_{20}(0) = -\frac{f(y)u_{20}(0)/\pi}{(\tau^{b+2}-1)/[(b+2)(\tau-1)] + 3(\tau+1)\Gamma/4} \times \left[\frac{\Gamma(\tau+1)[2\Gamma+\beta(\tau,b)]/4}{(\tau^{b+2}-1)/(b+2)(\tau-1)+(\tau+1)\Gamma} + \beta(\tau,b) + \Gamma\right]. \tag{39}$$

Equations (14), (29), (30), (37), and (39) provide a full solution for $\delta u_{20}(x,y)$.

With solutions for δu_{1R} and δu_{20} in hand, the calculation can proceed to the energy equation, which reveals the origin of the instability. Simplified for the two-dimensional geometry considered here, the energy equation is given by

$$(1-\phi)\rho_s c_s \frac{\partial}{\partial t}\left[T_s + \frac{\phi\rho c_v}{(1-\phi)\rho_s c_s}T\right] = -\frac{\partial\dot{h}_x}{\partial x} - \frac{\partial\dot{h}_y}{\partial y}, \tag{40}$$

where T_s, c_s, and ρ_s are the temperature, heat capacity, and density of the regenerator solid matrix, respectively. In good regenerators, $\phi\rho c_v/(1-\phi)\rho_s c_s \ll 1$ so the second term in the square brackets can be safely ignored. On the slow time scale of the initial growth of the perturbation, the heat transfer between the regenerator solid and the gas is adequate to ensure $\partial T_s/\partial t \approx \epsilon\delta T_m$. Using these approximations, substituting the full solution into Eq. (40), and expanding to first order in the perturbation yields

$$(1-\phi)\rho_s c_s \epsilon\delta T_m = -\frac{\partial\delta\dot{h}_x}{\partial x} - \frac{\partial\delta\dot{h}_y}{\partial y}. \tag{41}$$

Since we have assumed there are no acoustic or streaming flows in the y direction, the energy flux perturbation along y can be written

$$\frac{\partial\delta\dot{h}_y}{\partial y} = -k_y\frac{\partial^2\delta T_m}{\partial y^2} = -k_y T_a \sin(\pi x/x_w)\frac{d^2 f(y)}{dy^2}. \tag{42}$$

Here, k_y is the effective thermal conductivity of the regenerator solid in the y direction, which is much lower than the thermal conductivity of the regenerator solid k_s for three reasons. First, the screen wires do not fill up the entire cross-

sectional area of the screen. Second, the length of a piece of wire between two points is longer than the distance between the points due to the sine-wavelike bending of the wire as it passes over and under the wires running in the perpendicular direction. Third, the wires are not all aligned with the direction of heat flow. Taking all of these effects into account, k_y is given by

$$k_y = \frac{1-\phi}{2K(\phi)}k_s, \tag{43}$$

where

$$K(\phi) = \frac{2}{\pi}\int_0^{\pi/2} \sqrt{1+4(1-\phi)^2\sin^2 z}\, dz. \tag{44}$$

Applying the simplifications mentioned above, substituting Eqs. (30) and (42) into Eq. (41) and integrating from $x=0$ to x_w yields an expression for the growth rate ϵ

$$(1-\phi)\rho_s c_s f(y)\epsilon = \frac{\pi}{2x_w}\frac{[\delta\dot{h}_x(0) - \delta\dot{h}_x(x_w)]}{T_a} + k_y\frac{d^2f(y)}{dy^2}. \tag{45}$$

To complete the calculation, we only need to express $\delta\dot{h}_x(0)$ and $\delta\dot{h}_x(x_w)$ in terms of δT_m, i.e., in terms of $T_a f(y)$.

The streaming-free solution for the energy flux along x is given by $\dot{h}_x = \langle \tilde{m}_x c_p \tilde{T}\rangle - k_x \partial T/\partial x$, where k_x is the effective conductivity of the screen in the x direction.[14,16] Here, we have been forced to deviate from our original variable definitions by the form of the expression for $\dot{h}$. Variables with a tilde are microscopically varying[14] and $\langle\ldots\rangle$ indicates a local spatial average including a reasonable number of pores. Time averaging, expanding to linear order in the perturbation, and realizing that we only need the energy flux at $x=0$ or x_w yields

$$\dot{h}_{2,x} = \dot{m}_{1R}c_p\langle\tilde{T}_{1R}\rangle_{\tilde{u}_{1R}}/2 - k_x\,\partial T_m/\partial x \equiv \dot{h}_c + \dot{h}_k \tag{46}$$

$$\delta\dot{h}_x = \dot{h}_c\left[\frac{\delta u_{1R}}{u_{1R}} + \frac{\langle\delta\tilde{T}_{1R}\rangle_{\tilde{u}_{1R}}}{\langle\tilde{T}_{1R}\rangle_{\tilde{u}_{1R}}}\right] + \dot{h}_k\left[\frac{\partial\delta T_m/\partial x}{\partial T_m/\partial x}\right] + \delta\dot{m}_{20}c_p T_m. \tag{47}$$

We have again deviated somewhat from our original definitions of T_1 and δT_1 by working with velocity-weighted averages,[14] such as $\langle\tilde{T}_{1R}\rangle_{\tilde{u}_{1R}} = \langle\tilde{T}_{1R}\tilde{u}_{1R}\rangle/\langle\tilde{u}_{1R}\rangle$, instead of simple spatial averages. In the derivation of Eq. (47), we have assumed $\langle\tilde{T}_{1R}\rangle_{\tilde{\delta u}_{1R}} = \langle\tilde{T}_{1R}\rangle_{\tilde{u}_{1R}}$; this seems plausible since both $\tilde{u}_{1R}$ and $\tilde{\delta u}_{1R}$ flow through the same regenerator matrix and should experience similar small-scale spatial fluctuations.

The temperature oscillations are computed from Eq. (27) of Swift and Ward,[14] where they assume that the heat capacity of the regenerator solid is much larger than that of the gas and the gas-to-regenerator heat-transfer coefficient h does not depend on Reynolds number. We make the additional simplification that $\omega\rho_m c_p/(h/r_h) \approx (r_h/\delta_\kappa)^2 \ll 1$. The result is

$$\langle\tilde{T}_{1R}\rangle_{\tilde{u}_{1R}} = -\frac{\rho_m c_p u_{1R}(\partial T_m/\partial x)}{h/r_h}. \tag{48}$$

Substituting the full solution into Eq. (27) of Swift and Ward[14] and expanding to first order in the acoustic variables and linear order in the perturbation, we find

$$\langle\tilde{\delta T}_{1R}\rangle_{\tilde{u}_{1R}} = \langle\tilde{T}_{1R}\rangle_{\tilde{u}_{1R}}\left[\frac{\delta u_{1R}}{u_{1R}} + \frac{\partial\delta T_m/\partial x}{\partial T_m/\partial x}\right], \tag{49}$$

where we have again assumed $\langle\tilde{T}_{1R}\rangle_{\tilde{\delta u}_{1R}} = \langle\tilde{T}_{1R}\rangle_{\tilde{u}_{1R}}$. In Eq. (49), we have dropped a term proportional to $\delta\rho_m$ because $\langle\tilde{\delta T}_{1R}\rangle_{u_{1R}}$ is only needed at $x=0$ and x_w where $\delta\rho_m$ is zero. Substituting Eqs. (23), (30), (47), and (49) into Eq. (45) we find

$$\frac{2x_w}{\pi}(1-\phi)\rho_s c_s T_a\epsilon f(y) = \frac{2\pi\dot{h}_{2,x}f(y)}{\tau-1} + \delta\dot{m}_{20}c_p T_a(1-\tau) + \frac{2x_w k_y T_a}{\pi}\frac{d^2f(y)}{dy^2}. \tag{50}$$

To arrive at this result, the two terms of $\dot{h}_{2x}$, $\dot{h}_c$, and $\dot{h}_k$, are each taken to be independent of x. Equations (22), (30), (31), (36), (37), and (39) can be combined to show that $\delta\dot{m}_{20}$ is proportional to $f(y)$, so Eq. (50) shows that $f(y)$ is proportional to $\sin(n\pi y/y_w)$ or $\cos(n\pi y/y_w)$. The value of n is constrained by the temperature boundary conditions imposed by the regenerator geometry. For example, n is even for annular regenerators of circumferential extent y_w, to force continuity of temperature. (The requirement that $\int\delta\dot{m}_{20}dy=0$ rules out $n=0$.) Substituting Eq. (36) evaluated at $x=0$ into Eq. (50) and using the ideal gas equation of state, we find

$$\frac{x_w y_w z_w(1-\phi)\rho_s c_s T_a}{\pi^2 E_a}\epsilon = -\frac{(H/E_a)}{1-\tau} - n^2\frac{k_y T_a z_w x_w/y_w}{E_a} + \frac{\gamma}{\gamma-1}\frac{1-\tau}{2\pi}\left(\frac{\delta u_{1R}(0)}{u_{1R}(0)} - \frac{\delta u_{20}(0)}{u_{20}(0)}\right)\Bigg/ f(y). \tag{51}$$

Here, H and E_a are the total energy flux and the total acoustic power flux at the ambient end of the regenerator, z_w is the short dimension of the regenerator transverse to x_w, and γ is the ratio of specific heats of the working gas.

Equation (51) begins to shed some light on when an instability may arise. All factors on the LHS of Eq. (51), other than perhaps ϵ itself, are positive. Therefore, the sign of ϵ is determined only by the RHS of the equation. The transverse conductivity term, the term proportional to k_y, is negative for both engines and refrigerators and, therefore, always contributes to the stability of the streaming and temperature distribution. It is proportional to n^2, so it selects the broadest possible mode for the instability: $n=2$ for our annular regenerator, $n=1$ for a wide rectangular regenerator. [For circular regenerators, $f(y)$ would be something like $J_0(kr)\cos(n\theta)$ with k such that $(d/dr)J_0(kr)=0$ at the circumference of the regenerator, and $n=1$ would be selected.] For

engines, $\tau>1, E_a>0$, and $H<0$, making the first term on the RHS negative. In refrigerators, $\tau<1, E_a>0$, and $H>0$, so the first term is still negative. Therefore, this term always contributes to the stability of the streaming distribution. The third term on the RHS is more difficult to analyze. For values of τ, b, and Γ over the wide range we have explored, the difference inside the parentheses divided by $f(y)$ is always positive. Therefore, for engines, the third term on the RHS is always negative and this calculation predicts that engines are inherently immune from streaming instabilities. However, for refrigerators, the third term on the RHS is positive, leading to the possibility of a streaming instability. In the case of refrigerators, a streaming instability will arise (i.e., $\epsilon>0$) when

$$\frac{H}{E_a} < \frac{\gamma}{\gamma-1}\frac{(1-\tau)^2}{2\pi}\left(\frac{\delta u_{1R}(0)}{u_{1R}(0)} - \frac{\delta u_{20}(0)}{u_{20}(0)}\right) \Bigg/ f(y) - n^2(1-\tau)\frac{k_y T_a z_w x_w}{y_w} \Bigg/ E_a, \tag{52}$$

with $\delta u_{1R}(0)/f(y)u_{1R}(0)$ given by Eq. (31) and $\delta u_{20}(0)/f(y)u_{20}(0)$ given by Eq. (39). One interesting result is that for a regenerator with a linear flow resistance, i.e., $c_2=0$ in Eq. (26), $\Gamma=0$, $\delta u_{1R}(0)/u_{1R}(0) = \delta u_{20}(0)/u_{20}(0)$, and there is no instability of the type described here. The conclusion is that a parallel plate regenerator with uniform plate spacing and a Reynolds number less than about 2000 does not suffer from this type of instability.

III. APPARATUS

Since the acoustic streaming instability is sometimes expected in refrigerators but never in engines, we have built an orifice pulse tube refrigerator (OPTR) to specifically search for it. Figure 2 shows a scale drawing of the OPTR used in these measurements. It consists of three heat exchangers, a regenerator, and a pulse tube. The rest of the hardware includes a piston driven by a linear motor,[17] an experimental offset, and a variable acoustic network. The system is filled with 30-bar helium gas and driven at a fixed frequency of 45 Hz. The 0.10-m diameter piston (not shown) is located directly beneath the experimental offset.

The offset is simply an open cylinder that allows for impedance matching between the OPTR and the piston/linear motor. It also allows access to the inner cooling-water channel of the aftercooler. The offset also has a port for measuring acoustic pressure using a piezoresistive transducer.[18]

Above the offset is the aftercooler, which is the principal ambient heat exchanger. It is made from a 5.1-cm-thick brass block drilled with a total of 90 3.2-mm-diameter holes in two circular rows. Annular cooling-water channels are located just inboard and outboard of the holes.

Above the aftercooler is the regenerator. It is made from plain square-weave stainless-steel screen with an inner diameter of 5.5 cm and outer diameter of 7.4 cm. The screen is cut by wire electrical-discharge machining, cleaned, and then packed into the regenerator housings with an annular die to a height of $x_w=5.08$ cm. The inner and outer housings around

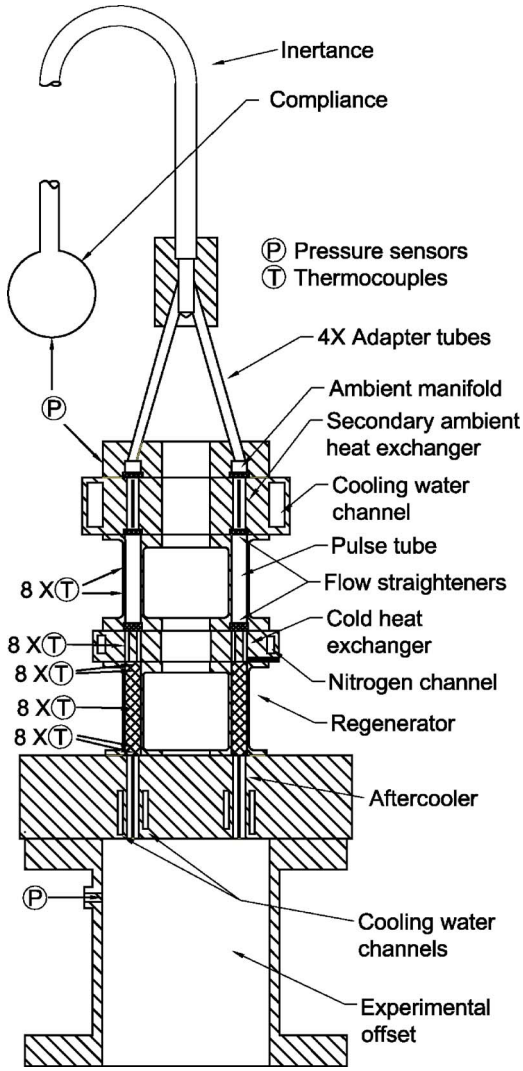


FIG. 2. Scale drawing of the orifice pulse tube refrigerator used in this study. All components from the aftercooler to the ambient manifold are annular. The rest of the components are circular. The four adapter tubes transition the annular geometry back to circular.

the regenerator have wall thicknesses of 1.65 and 1.32 mm, respectively. On each end, four layers of 20-mesh copper screen with a wire diameter of 0.4 mm are inserted as simple spacers to allow room for the flow to spread out after exiting the heat exchangers. Temperatures are measured with 40 1.0-mm-diameter sheathed type-K thermocouples located every 45° around the azimuth at the five axial positions as shown in Fig. 2. The thermocouples are inserted into tight-fitting pockets drilled halfway through the regenerator's annular thickness, and the stainless-steel sheaths are soft soldered to slightly larger, short tubes that are brazed to the outer housing, thereby making a leak-tight seal. The top and bottom sets of eight are centered in the copper spacers. The other three sets are in the regenerator itself, just inside the copper spacers and at the axial midpoint.

Two different regenerators are tested. The first has a porosity of 0.686 and a hydraulic radius[8] of 22.2 μm. The second has a porosity of 0.686 and a hydraulic radius of 13.9 μm. In both regenerators, the hydraulic radius is much smaller than the thermal penetration depth, about 204 μm at

300 K, ensuring good thermal contact between the helium gas and the screen. The first regenerator never showed an instability. Data from it are only discussed briefly, and the rest of this article focuses on the second regenerator.

Above the regenerator is the cold heat exchanger. It is made from a 1.9-cm-thick block of oxygen-free high-conductivity copper drilled with a total of 180 2.4-mm-diameter holes in three circular rows. A channel around the outside of the heat exchanger allows for cooling or heating of the heat exchanger with liquid or gaseous nitrogen or a water-antifreeze mixture as necessary. Eight type-K thermocouples are inserted into drilled pockets around the perimeter near the lower face of the cold heat exchanger at the same angular locations as in the regenerator.

The annular pulse tube is located above the cold heat exchanger. Its stainless-steel inner and outer shells are nearly identical to the regenerator housings. The pulse tube provides thermal insulation between the cold and ambient heat exchangers while transmitting acoustic power out of the cold zone. The surfaces facing the helium gas are polished to ensure that the surface roughness is much less than the viscous and thermal penetration depths. Several layers of copper screen at either end of the pulse tube serve as flow straighteners. Temperature in the pulse tube is measured with 16 type-K thermocouples spot welded to the outer wall every 45° around the azimuth at two axial locations.

Above the pulse tube is the secondary ambient heat exchanger, which is similar in construction to the aftercooler. It consists of a drilled brass block with a cooling-water channel around its outside. Next, an annular manifold and four 6.4-mm-diameter tubes adapt the annular geometry back to a tubular geometry. A flow straightener in the manifold keeps the flow from the four tubes from jetting through the secondary ambient heat exchanger. A second acoustic pressure sensor is located in the manifold.

Above the four adapter tubes, 3.4 m of 12.7-mm-diameter inertance tubing extends to a 2.3-liter compliance tank forming an acoustic network[19] used to set the volumetric flow rate at the aftercooler. A third pressure sensor is located in compliance. All three pressure sensors are calibrated using a steady pressure measured with a National Institute of Standards and Technology traceable Bourdon-tube pressure gauge.

IV. PRELIMINARY MEASUREMENTS

Several quantities must be measured accurately to make a comparison with the theory: τ, Γ, E_a, H, and k_y. For the first three of these, we require accurate measurements of the temperature, complex pressure p_{1a}, and volumetric velocity U_{1a} at the ends of the regenerator.

Thermocouples in the copper spacers, regenerator, cold heat exchanger, and pulse tube are calibrated *in situ* by submerging the entire assembly in baths of known temperature: liquid nitrogen (75 K at Los Alamos atmospheric pressure) and a dry ice-acetone bath (195 K). A third calibration point is obtained by letting the insulated system sit undisturbed overnight (with the cooling water shut off) and reading all of the thermocouples in the morning. Quadratic fits to the temperature deviations are used to interpolate the corrections between the calibration points. The range of the temperature corrections is ±3 K and ±8 K at dry ice-acetone and liquid-nitrogen temperatures, respectively.

The measurement of the complex pressure amplitude is straightforward. All pressure sensors in the OPTR are read out with the same lock-in amplifier, ensuring accurate magnitude and phase information. The magnitude and phase of the pressure amplitude in the experimental offset changes very little along its length, so the pressure measured at this sensor provides the complex pressure amplitude at both the piston face $p_{1,p}$ and the ambient end of the regenerator p_{1a}.

The complex volumetric-velocity amplitude is more difficult to determine. A mutual-inductance-based linear variable-displacement transducer (LVDT) (Ref. 20), on the linear motor provides a measure of the complex displacement amplitude of the piston. The flow rate at the piston face $U_{1,p}$ is then determined from knowledge of the piston area and angular frequency of the oscillation, ω. However, U_1 changes significantly between the piston face and the ambient face of the regenerator. Most of the change is in the imaginary part and is due to the compliance of the experimental offset. This change is easily accounted for by measuring the total volume between the piston face and the aftercooler, and knowing $\omega, p_{1,a}$, and the mean pressure. However, there is also a change in the real part of U_1 due to thermal and viscous dissipation on the experimental offset wall E_{offset}, as well as dissipation due to leakage through the clearance seal between the compressor's piston and cylinder E_{seal}. Instead of trying to compute each source of dissipation, a set of measurements is carried out to determine the combined effect of all of them. At the top of the experimental offset, the OPTR is replaced by a variable acoustic load consisting of a 2.2-liter tank connected to the offset by a water-cooled globe valve.[21] By adjusting the valve, varying amounts of acoustic power E_{load} can be dissipated in the load. Pressure sensors in the tank and offset are used to obtain E_{load}.[21] The acoustic power leaving the piston face E_p is determined from measurements of $U_{1,p}$ using the LVDT and $p_{1,a}$ using the pressure sensor in the experimental offset. Both measurements are made using the lock-in amplifier, which allows for accurate determination of the phase between $p_{1,a}$ and $U_{1,p}$. With the load valve closed, $E_p = E_{\text{offset}} + E_{\text{seal}}$. As the load valve is opened, the additional power drawn by the load must originate at the piston face. Therefore, $E_p = E_{\text{offset}} + E_{\text{seal}} + E_{\text{load}}$. If $p_{1,a}$ is held constant for each valve setting, E_{seal} and E_{offset} should be nearly constant. Therefore, if E_p is plotted versus E_{load} for various valve settings, the result should be a straight line of slope 1 and an E_p-axis intercept of $E_{\text{offset}} + E_{\text{seal}}$. Figure 3 shows the results of these measurements. Each set of data is fitted with a straight line, and the slopes range from 1.00 to 1.05 demonstrating the accuracy of the E_p and E_{load} measurements. The fitted intercepts giving $E_{\text{offset}} + E_{\text{seal}}$ are displayed in Fig. 4.

The entire system is modeled with DeltaE.[22] To account for the dissipation in the piston seal and experimental offset, an extra side-branch acoustic resistance that dissipates $E_{\text{seal}} + E_{\text{offset}}$ from Fig. 4 is added to the DeltaE model at the piston end of the experimental offset. In combination with a

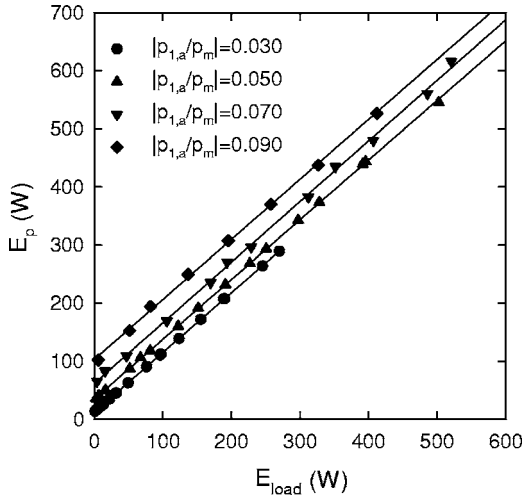


FIG. 3. Acoustic power leaving the piston face E_p versus the power dissipated in the variable acoustic load E_{load} for several different $|p_{1,a}/p_m|$. The lines are least squares fits to the data. The slopes of the lines range from 1.0 to 1.05, demonstrating the accuracy of determining E_p from $p_{1,a}$ and LVDT measurements of the piston location. The E_p intercept at each value of $|p_{1,a}/p_m|$ determines the acoustic dissipation due to boundary-layer processes in the experimental offset and piston seal leakage.

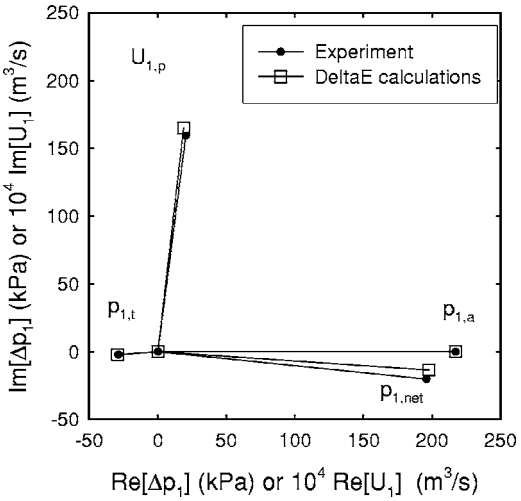


FIG. 5. Measured and calculated acoustic pressure and volumetric velocity phasors throughout the experimental system at $|p_{1,a}/p_m|=0.07$ and $T_c=77$ K.

segment that models the compliance of the offset, the extra resistance makes appropriate changes to $U_{1,p}$ to give $U_{1,a}$, which also modifies E_p to give E_a. For each data point, measurements of $U_{1,p}, p_{1,a}, p_{1,\text{net}}$, and $p_{1,t}$ are compared with calculations using the DeltaE model. Here, $p_{1,\text{net}}$ and $p_{1,t}$ are the complex pressure amplitudes in the ambient manifold just below the acoustic network and in the compliance tank, respectively. Typical results are shown in Fig. 5. The phase of $p_{1,a}$ has been arbitrarily set to zero, and $|p_{1,a}|$ is forced to be the same in the model and experiment. The measured $U_{1,p}$ phasor is in good agreement with the model. The only other measure of U_1 in the system is the pressure oscillation in the compliance tank. The good agreement between the measurement of $p_{1,t}$ and the model indicates U_1 into the tank is close to the model predictions. With agreement between measurement and model predictions for U_1 at the piston and in the compliance tank, the model predictions of $U_{1,a}$ (which determines Γ) and U_1 throughout the system can be used with some confidence.

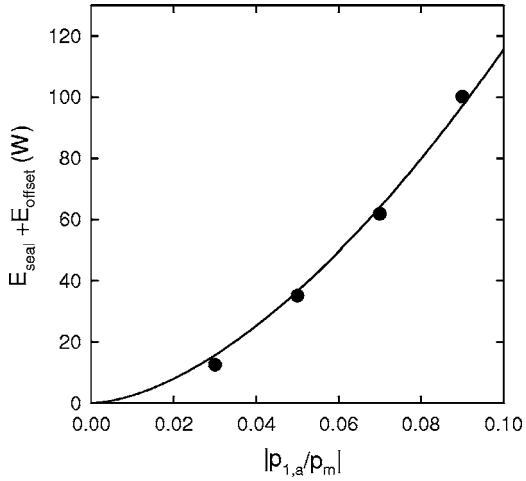


FIG. 4. Acoustic power dissipated by boundary-layer processes in the experimental offset and by piston seal leakage vs $|p_{1,a}/p_m|$. The fit to the data is used to interpolate between the data points.

In principle, the total energy flux H through the regenerator could be determined from the difference between E_a and the measured heat rejected at the aftercooler. However, this would not provide a very accurate measure because H is relatively small compared to both of these quantities. Therefore, we resort to our knowledge of U_1 and p_1 throughout the system to numerically compute[22] H. However, for this computation to be accurate, we must know the flow impedance and heat-transfer properties of the regenerator, which depend on r_h and ϕ. By measuring the total mass of the screen in the regenerator and its volume, ϕ is accurately determined. The hydraulic radius is determined from ϕ and the diameter of the screen wire d_w as quoted by the manufacturer, via $r_h=(d_w/4)\phi/(1-\phi)$, yielding 13.9 μm for the second regenerator. The acoustic pressure drop across the OPTR $p_{1,a}-p_{1,\text{net}}$ is used as a check of this value of r_h. Figure 5 shows the typically 80% agreement between the measurements of $p_{1,a}-p_{1,\text{net}}$ and the numerical computation.[22] A 20% uncertainty in $p_{1,a}-p_{1,\text{net}}$ implies a 10% uncertainty in r_h, which in turn implies a 20% uncertainty in our computed values of H.

V. EXPERIMENTS

To search for the instability, we vary the ratio H/E_a. We interpret any azimuthal dependence of the steady-state temperature in the midplane of the regenerator as a sign that an instability has arisen, has grown in time, and has stopped growing when it reaches a balance with some other nonlinear effect that is beyond the scope of this paper. Data are taken by keeping both T_a and T_w (i.e., τ) nearly fixed while varying $|p_{1,a}|$. Under these conditions, E_a goes approximately as $|p_{1,a}|^2$. The hydrodynamically transported energy flux $\dot{h}_c$ in Eq. (46) also grows as $|p_{1,a}|^2$. However, $\dot{h}_k$, the thermal-conduction component of the energy flux in Eq. (46), stays fixed. Therefore, by varying $|p_{1,a}|$, we can vary H/E_a in Eq. (52) while keeping τ fixed. At each acoustic amplitude, the

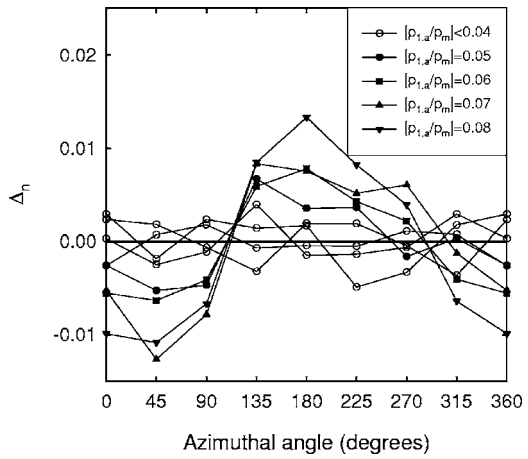


FIG. 6. Normalized measured temperatures around the azimuth at the axial midpoint of the regenerator for various $|p_{1,a}/p_m|$ and T_c=77 K. See the text for how the temperatures have been normalized. The appearance of a sine-wavelike distribution at higher $|p_{1,a}/p_m|$ signals the onset of an acoustic streaming instability.

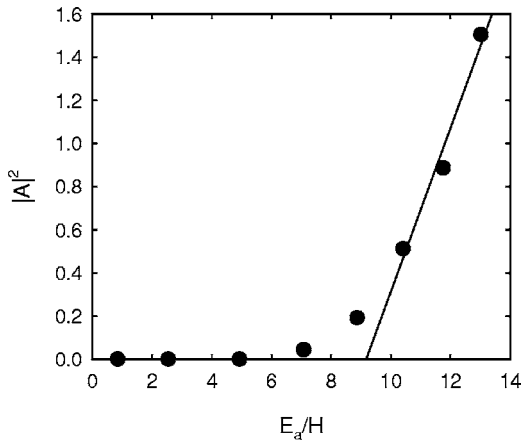


FIG. 7. The amplitude squared of the Fourier transform of the temperature data in Fig. 6. The line is a least squares fit to the last three data points. The $|A|^2$=0 intercept is defined to be the onset of the instability, i.e., $(H/E_a)_{\text{crit}}$.

system is allowed to reach a steady state indicated by negligible changes in temperature of any of the thermocouples in the regenerator. All the thermocouple temperatures, the acoustic pressures, and the piston amplitude are then recorded. The temperatures at the ambient and cold end of the regenerator and $|p_{1,a}|$ are used as inputs to the numerical model that calculates H. The acoustic power E_a is determined as described in the previous section.

At low amplitude where H is dominated by $\dot{h}_k$ and H/E_a is large, the temperature at the axial midpoint of the regenerator is found to be roughly independent of the azimuthal angle. As $|p_{1,a}|$ is increased, the angular temperature distribution changes little until a critical $|p_{1,a}|$ is reached, where a sine-wavelike angular temperature distribution appears. If $|p_{1,a}|$ is increased beyond the critical value, the amplitude of the temperature variation increases. A typical set of data for $T_a \approx 300$ K and $T_c \approx 77$ K is shown in Fig. 6. Here, the temperature at the axial midpoint at each angle, $T_{\text{mid},n}$ is normalized by the temperatures at the ambient and cold ends $T_{a,n}$ and $T_{c,n}$, respectively, by

$$\Delta_n = \frac{T_{\text{mid},n} - (T_{a,n} + T_{c,n})/2}{T_{a,n} - T_{c,n}} - Z_n. \tag{53}$$

Any residual bias from systematic errors in the thermocouple calibration is removed by subtracting off from each Δ_n a zero Z_n which is obtained by averaging data sets whose $|p_{1,a}|$ are clearly below the critical value. The resulting Δ_n values are plotted in Fig. 6 for several $|p_{1,a}|$. A numerical calculation indicates that a sinusoidally y-dependent streaming mass flux with a peak of only $0.003\,\text{Re}[\rho_1 u_1^*]/2$ would cause the changes in temperature observed at $|p_1/p_m|=0.08$.

To accurately determine the location of the transition to streaming, the magnitude of the temperature variation in Fig. 6 is determined from a Fourier transform via

$$|A|^2 = \left| \frac{\pi}{4} \sum_{n=1}^{8} \Delta_n \cos(n\pi/4) \right|^2 + \left| \frac{\pi}{4} \sum_{n=1}^{8} \Delta_n \sin(n\pi/4) \right|^2, \tag{54}$$

and $|A|^2$ is plotted versus E_a/H in Fig. 7. Consistent with Eq. (52), as E_a/H becomes larger a critical value is reached where $|A|^2$ begins to rapidly grow. There is some rounding near the transition, and the critical value $(E_a/H)_{crit}$ is defined as the $|A|^2=0$ intercept of a line fitted to the data points beyond the visible rounding.

The same data taking and analysis procedure is applied to data sets with $T_a \approx 300$ K and T_c spanning 77 to 285 K. The results for $(H/E_a)_{\text{crit}}$ versus τ are plotted as the circles in Fig. 8. The detailed acoustic and thermal conditions are shown in Table I. Operating conditions above and to the right of the circles in Fig. 8 do not show azimuthal temperature variations, whereas operating conditions below and to the

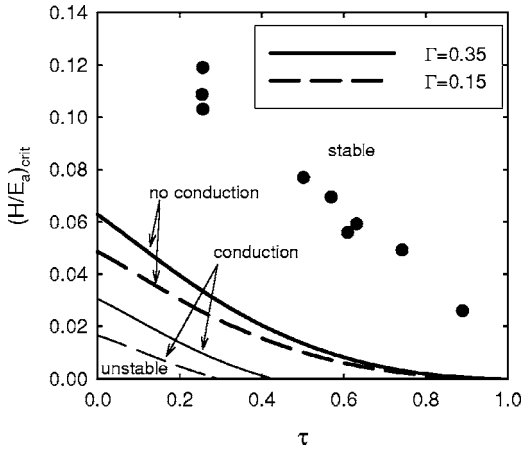


FIG. 8. $(H/E_a)_{\text{crit}}$ versus τ. Each point is an experimental value obtained from a data set like that of Fig. 7. The thin and bold lines are Eq. (52) with and without the transverse conduction term, respectively. To check the repeatability of the measurements, the group of three points near τ=0.25 were taken under nearly identical acoustic conditions, as was the pair of points near τ=0.60. The primary sources of uncertainty in $(H/E_a)_{\text{crit}}$ are the scatter in the individual temperature measurements used to determine $|A|^2$ and the choice of how many points in the plots of $|A|^2$ vs E_a/H to include in the linear fit (see Fig. 7).

TABLE I. Acoustic and thermal conditions for the data in Fig. 8.

T_c (K)	τ	$(H/E_a)_{\text{crit}}$	H (W)	$p_{1,a}$ (kPa)	$10^3 U_{1,a}$ (m^3/s)	Phase $U_{1,a}$ (deg)	Γ
76	0.25	0.119	15.7	145	2.45	42.1	0.29
75	0.25	0.109	17.4	155	2.65	38.9	0.32
75	0.25	0.103	20.3	171	3.10	41.8	0.36
151	0.50	0.077	9.6	174	2.00	44.3	0.24
171	0.57	0.069	7.6	171	1.78	43.6	0.21
188	0.62	0.056	7.7	192	1.84	39.0	0.22
189	0.63	0.059	7.2	186	1.76	42.4	0.21
224	0.74	0.049	4.2	168	1.31	38.8	0.15
276	0.90	0.026	4.1	233	1.71	37.5	0.19

left of the circles show azimuthal variations presumably due to the acoustic streaming instability described earlier. The first regenerator, with a 22.2 μm hydraulic radius, always yielded a higher (H/E_a) for a particular τ, i.e., above and to the right of the data points in Fig. 8, and never showed an instability.

Comparing the experimental results in Fig. 8 with Eq. (52) is not straightforward because τ is not the only independent parameter on the right-hand side of Eq. (52). The other independent parameter is Γ, which is a measure of the Reynolds number at the ambient end of the regenerator. Experimentally, Γ can be varied by using a valve (not shown in Fig. 2) in the acoustic network to vary the impedance of the network. However, increasing Γ by a factor is two is not feasible because the pressure drop across the regenerator would become comparable to $|p_{1,a}|$, violating one of the assumptions in the calculation. Decreasing Γ by a factor of two is also not possible for several reasons. First, lower values of Γ imply smaller $|U_{1,a}|$ at the same $|p_{1,a}|$. The consequence would then be large phase shifts in U_1 from the ambient to the cold end of the regenerator, violating our assumption that p_1 and U_1 are in phase throughout the regenerator. Second, to maintain the phase of $p_{1,a}$ relative to $U_{1,a}$ near zero would require an acoustic network that had nearly the same inertial component (i.e., reactive component) while reducing the dissipative component by a factor of two. This is not possible at the current power level. Therefore, we are restricted to the narrow range of Γ allowed by the current apparatus.

In Fig. 8, experimental values of Γ range from about 0.35 at the lowest τ to 0.15 at the highest. The variation in Γ is somewhat subtle. At a particular $|p_{1,a}|$, the volumetric flow rate at the cold end of the regenerator $U_{1,c}$ is fixed by the dimensions of the inertance tube and the compliance tank. Equation (8) shows that $U_{1,a}=U_{1,c}/\tau$. Therefore, $U_{1,a}$ and Γ are smaller for large τ and larger for small τ. Instead of calculating an instability threshold for each data point, we present two curves in Fig. 8; one for the threshold with $\Gamma=0.35$ and another for $\Gamma=0.15$. To help distinguish the magnitude of the two contributions on the right-hand side of Eq. (52), the bold curves only take into account the first term while the thin curves also take into account the transverse conduction term. Just as with the experimental data, operating points below and to the left of the curves are unstable while those above and to the right are stable.

The qualitative agreement between measurements and calculations in Fig. 8 is encouraging. Both experiment and calculation show an instability for operation below a critical value of H/E_a, and both show that $(H/E_a)_{\text{crit}}$ decreases as τ increases. However, the quantitative agreement is not very good. As the bold curves show, the disagreement is not simply due to an overestimate of the stabilizing effect of transverse conduction. Instead, we may be underestimating the destabilizing effects of streaming or simply leaving out one or more important effects. The exact solutions for oscillating flow and heat transfer in a screen regenerator are unknown, forcing us to use steady-flow correlations in a oscillating flow.[14] Inaccuracies certainly result, and the effects on this subtle calculation are difficult to estimate. In addition, since we cannot solve the equations for the perturbation exactly, we have assumed the functional form of the temperature perturbation profile based on observations of a different system when streaming was present. Although we believe this form to be close to reality, perhaps the instability threshold would be significantly changed by picking a different functional form. The calculation we have presented assumes that p_1 and U_1 are in phase throughout the regenerator. The data in Table I show that $U_{1,a}$ leads $p_{1,a}$ by approximately 40°. The effect of this phase difference is difficult to estimate without including it at all stages of the calculation. Finally, we assumed the *transverse* acoustic and streaming velocity perturbations are both zero. In reality this is certainly not the case because, although not presented in this paper, the solutions arrived at in this calculation show a y-dependent p_1 and $p_{2,0}$ inside the regenerator, i.e., everywhere but $x=0$ or $x=x_w$. Quick initial estimates show that the energy flows driven by the resulting transverse acoustic and streaming flows are either stabilizing or have no effect. However, more detailed calculations are required to verify these estimates.

VI. CONCLUSIONS

A calculation has been presented showing that an acoustic streaming instability can arise in the regenerators of oscillating-wave refrigerators, while engines are immune. The calculation begins by assuming that a region "A" near the midplane of the regenerator becomes a little hotter than it should while region "B" (also near the midplane but some transverse distance away) becomes a little cooler. Some of the consequences of such a mean temperature perturbation will always remove heat from region A and deposit heat in region B, leading to a suppression of the original perturbation. Referring to Eq. (51), these include transverse thermal

conduction through the regenerator screen (proportional to k_y) and the axial total energy flux H (proportional to dT_m/dx). The other term in Eq. (51), $\delta u_{1R}(0)/u_{1R}(0) - \delta u_{20}(0)/u_{20}(0)$, is proportional to the second-order streaming mass-flux perturbation. Under all conditions we have explored, the resulting steady mass flux is always positive (i.e., directed from T_a to T_w in regions where the midplane temperature is higher). In engines, heat is extracted from region A and deposited in region B by this term, which suppresses the mean temperature perturbation and leads to the stability of engines to this type of perturbation. However, the opposite is true in refrigerators where additional heat is deposited in the region A and removed from region B. The details of the state of the refrigerator's regenerator dictate whether the heat flux perturbation due to this streaming can overcome the stabilizing effects of transverse thermal conduction and axial total energy flux. If it does, the perturbation grows, resulting in an instability in refrigerators.

When a perturbation is present, the changes in the heat flux due to the transverse thermal conduction and the axial energy flux are easily understood. However, the second-order streaming mass-flux perturbation is more complicated. It is nonzero because the nonlinear flow resistance of the regenerator results in different flow resistances for the first-order acoustic flow and the second-order streaming flow and because these resistances change with mean temperature in different ways.

The calculation of the instability threshold is difficult because the three effects mentioned above are all about the same size. A small inaccuracy in the calculation of any one of three can lead to a significant inaccuracy in the result for the instability threshold. An accurate numerical calculation could presumably predict the instability threshold. However we chose to attempt an analytical calculation to gain understanding by identifying the specific stabilizing and destabilizing mechanisms.

An orifice pulse tube refrigerator with a well-instrumented regenerator was built to determine the instability threshold. The calculation described above and the experimentally determined threshold show qualitative agreement, but quantitative agreement is still lacking. Possible routes of additional theoretical investigation include better models for the flow and heat transfer in regenerators, determining the exact solution for the x dependence of the mean-temperature perturbation, and including acoustic and streaming flow transverse to the main acoustic axis.

ACKNOWLEDGMENTS

The authors would like to acknowledge D. Gardner for work on an early version of this experiment, C. Espinoza for his expert assistance in the construction of the experiment, and the Office of Basic Energy Sciences within the U.S. DOE Office of Science for financial support.

[1] A. J. Organ. *The Regenerator and the Stirling Engine* (Mechanical Engineering Publications, Ltd., London 1997).
[2] R. Radebaugh, "A review of pulse tube refrigeration," Adv. Cryog. Eng. **35**, 1191–1205 (1990).
[3] G. W. Swift, D. L. Gardner, and S. Backhaus, "Acoustic recovery of lost power in pulse tube refrigerators," J. Acoust. Soc. Am. **105**, 711–724 (1999).
[4] S. Backhaus and G. W. Swift, "A thermoacoustic-Stirling heat engine," Nature (London) **399**, 335–338 (1999).
[5] T. Yazaki, A. Iwata, T. Maekawa, and A. Tominaga, "Traveling wave thermoacoustic engine in a looped tube," Phys. Rev. Lett. **81**, 3128–3131 (1998).
[6] C. M. de Blok and N. A. H. J. van Rijt, "Thermo-acoustic system," U. S. Patent No. 6,314,740, Nov. 13, 2001.
[7] D. L. Gardner and G. W. Swift, "A cascade thermoacoustic engine," J. Acoust. Soc. Am. **114**, 1905–1919 (2003).
[8] A. J. Organ, *Thermodynamics and Gas Dynamics of the Stirling Cycle Machine* (Cambridge University Press, Cambridge, England, 1992).
[9] David Gedeon, "Flow circulation in foil-type regenerators produced by non-uniform layer spacing," in *Cryocoolers 13*, edited by R. G. Ross (Springer, New York, 1999), pp. 421–430.
[10] C. S. Kirkconnell, "Experimental investigation of a unique pulse tube expander design," in *Cryocoolers 10*, edited by R. G. Ross (Plenum, New York, 1999), pp. 239–247.
[11] G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America, Sewickley, PA 2002).
[12] S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability* (Dover, New York, 1961).
[13] S. Backhaus and G. W. Swift, "An acoustic streaming instability in thermoacoustic devices utilizing jet pumps," J. Acoust. Soc. Am. **113**, 1317–1324 (2003).
[14] G. W. Swift and W. C. Ward, "Simple harmonic analysis of regenerators," J. Thermophys. Heat Transfer **10**, 652–662 (1996).
[15] Unpublished data measured in a different apparatus [S. Backhaus, E. Tward, and M. Petach, "Traveling-wave thermoacoustic generator," Appl. Phys. Lett. **85**, 1085–1087 (2004)] using five thermocouples attached to the outside of a thin-wall can containing the regenerator show that this profile is closely represented by a sine wave.
[16] M. A. Lewis, T. Kuriyama, F. Kuriyama, and R. Radebaugh, "Measurement of heat conduction through stacked screens," Adv. Cryog. Eng. **43**, 1611–1618 (1998).
[17] Model C2, Qdrive, Inc., Troy NY, www.qdrive.com
[18] Model 8510B-500. Endevco Corporation, 30700 Rancho Viejo Road, San Juan Capistrano, CA.
[19] D. L. Gardner and G. W. Swift, "Use of inertance in orifice pulse tube refrigerators," Cryogenics **37**, 117–121 (1997).
[20] Schaevitz Model 500 HR, Measurement Specialties, Inc. Fairfield NJ, www.schaevitz.com
[21] A. M. Fusco, W. C. Ward, and G. W. Swift, "Two-sensor power measurements in lossy ducts," J. Acoust. Soc. Am. **91**, 2229–2235 (1992).
[22] W. C. Ward and G. W. Swift, "Design environment for low amplitude thermoacoustic engines (DeltaE)," J. Acoust. Soc. Am. **95**, 3671–3672 (1994). Software and user's guide available either from the Los Alamos thermoacoustics website at www.lanl.gov/thermoacoustics/or from the Energy Science and Technology Software Center, US Department of Energy, Oak Ridge, Tennessee.

Experimental investigation of the effects of water saturation on the acoustic admittance of sandy soils

Kirill V. Horoshenkov[a] and Mostafa H. A. Mohamed
School of Engineering, Design and Technology, University of Bradford, Bradford, West Yorkshire, BD7 1DP, UK

(Received 31 December 2005; revised 26 July 2006; accepted 26 July 2006)

A novel technique for the laboratory characterization of the frequency-dependent acoustic surface admittance of partly saturated samples of sands is presented. The technique is based on a standard laboratory de-watering apparatus coupled with a standard acoustic impedance tube. The dependence of the surface admittance on the degree of water saturation is investigated for two samples of sand with widely different flow resistivities. It is shown that a relatively small change (e.g., from 0% to 11% by volume) in the degree of water saturation can result in a much larger change (e.g., twofold) in the acoustic surface admittance. An empirical relationship is found between the peaks observed in the real part of admittance spectra for the low flow resistivity sand and the degree of water saturation. The data are compared with predictions of two widely used ground impedance models: a semiempirical single parameter model and a two parameter model. A modified two-parameter version of a single-parameter model is found to give comparable fit to the two-parameter model. However, neither model provides an accurate fit.
 [DOI: 10.1121/1.2338288]

PACS number(s): 43.28.En, 43.55.Ev, 43.20.Gp, 43.58.Bh [KA] Pages: 1910–1921

I. INTRODUCTION

Outdoor sound propagation at short and medium ranges is strongly affected by the presence of the porous ground. This effect has been studied in detail over the last 20–30 years and can be calculated for a given frequency of sound provided that the corresponding value of the complex acoustic surface admittance (β_s) of the ground is known. The ground admittance is significantly affected by the type of material in the porous ground,[1] its roughness,[2] degree of compaction[3] and the moisture content.[4,5] Thus the relationship between the above parameters of the ground and the corresponding value of the admittance becomes rather complex. As a result, theoretically founded methods are typically abandoned in favor of purely empirical models which predict the ground admittance. For example, the HARMONOISE method for the prediction of noise propagation outdoors,[6] uses the empirical expression for the characteristic impedance ($z_b = 1/\beta_s$) proposed in 1970s by Delany and Bazley[7] (Eq. (35) in Ref. 7). The model by Delany and Bazley is based solely on the *effective* flow resistivity which is treated as an adjustable parameter to achieve the best fit between prediction and measurement (e.g., Ref. 8). The application of the model proposed by Delany and Bazley for use in acoustic ground characterization is questionable since: (i) the model was originally developed to predict the acoustic properties of fiberglass rather than loose granular media and (ii) it violates the Kramers-Kronig relations for causal behavior.[9] The only way the model can account for any changes in soil conditions due to the amount of precipitation, sun and wind effects, is by an adjustment of the effective flow resistivity.

A change in the degree of water saturation with time is a common problem when considering outdoor sound propagation. Precipitation, sun activity and morning dew all result in a change in the degree of water saturation of porous soil which is likely to result in a related change in the behavior of the frequency-dependent acoustic surface admittance. Generally, there are two major effects on the acoustic admittance caused by changes in the partial saturation in a porous soil. These may result from: (i) layer thickness changes with mean water level (i.e., position of the hard-backed layer); and (ii) changes in the size distribution of open, interconnected pores. If the soil consists of parallel capillary tubes, the cross sections of which are independent of depth, then effect (ii) will be predominant. Effect (i) will be predominant if the soil consists of large pores in which the capillary forces are relatively small compared to the gravity force. The microstructure of sandy soils, is much more complex than a stack of parallel capillary tubes. Smaller pores are connected to larger pores in a peculiar pattern which is hard to establish without resorting to x-ray tomography. Therefore, the overall effect of moisture on the acoustic performance of soil is hard to quantify in a relatively simple acoustic experiment. As a result and despite several attempts (e.g., Refs. 4, 5, 10, and 11) these effects have not yet been properly characterized under controlled laboratory conditions over a range of water saturation. A critical requirement in this type of experiment is the accurate, noninvasive measurement of the proportion of open pores and the distribution of the moisture across the area of the sample. This requirement has not been sufficiently met either in the pioneering work by Dickinson and Doak[4] or in the more recent work by Cramond and Don.[5] One can also argue that the inversion method based upon sound propagation data and empirical, noncausal expressions

[a]Author to whom correspondence should be addressed; electronic mail: k.horoshenkov@bradford.ac.uk

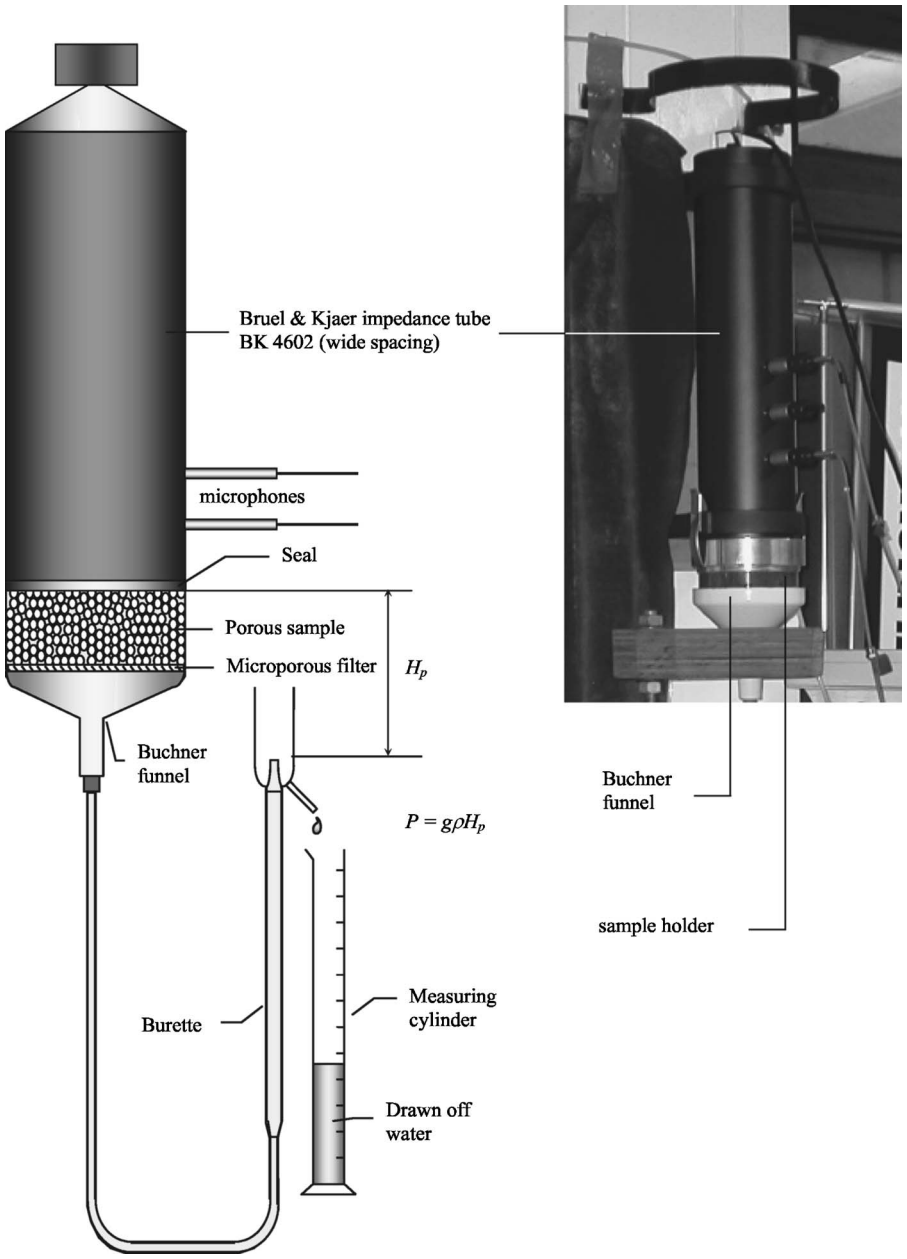


FIG. 1. Apparatus for determining simultaneously the pore size distribution and the acoustic surface impedance of water saturated porous samples.

for the ground impedance (e.g. Ref. 5) may not provide sufficiently representative information on the realistic acoustic properties of partly saturated soils. Such a technique assumes that the Delany and Bazley model is able to predict accurately the acoustic impedance (z_b) and propagation constant (k_b) in such soils.

The paper is organized as follows. Section II describes controlled laboratory measurements of the effect of varying moisture content on the acoustic properties of two samples of fine and coarse sand. Section III discusses the resulting data. Section IV presents two ground impedance models and suggests a simple modification to one of them to enable tolerable agreement with data. In Sec. V predictions of these models are compared with the results of a controlled laboratory experiment in which the surface acoustic admittance is measured directly and linked to the degree of water saturation. Section V offers some concluding remarks.

II. EXPERIMENTAL PROCEDURE

A controlled water suction experiment[12] was carried out on a saturated porous sample using a standard Buchner funnel setup,[13] coupled with the simultaneous measurement of the surface acoustic impedance using the impedance tube method.[14] Figure 1 shows the arrangement used. A vertical Bruel and Kjaer 100 mm impedance tube (Type BK 4206), operated in the wide-spacing mode (100 mm microphone spacing) in the frequency range between 50 and 1600 Hz, is connected to a standard Bruel and Kjaer PULSE™ data acquisition system. The sample holder is attached to the Buchner funnel and subsequently to a glass burette and measuring tube. The lower part of the sample holder is 50 mm deep and its internal diameter is machined precisely to match the internal diameter of the impedance tube. The upper part of the sample holder is 25 mm deep and its internal diameter was machined to match the external diameter of the impedance

TABLE I. A summary of the material properties of porous samples used in the experiments.

Material	Chemical composition	Grain density, kg/m^3	Mean grain size, mm	Mean pore size, μm	Porosity %	Dry flow resistivity kPa s m^{-2}
Coarse sand	SiO_2	2650	0.65	98	39	85
Fine sand	SiO_2	2665	0.37	65	44	314

tube. The sample holder slides onto the impedance tube and is clamped using the two standard brackets. The circumference where the impedance tube and the sample holder are joined is sealed with a rubber gasket. The bottom of the sample holder is perforated and lined with a single layer of a filter paper having 5 μm pores. Filter paper is used to enable the continuous flow of water out of the sample and to prevent the leaching of fine particles. The edge of the filter paper is sealed using silicon sealant to ensure that there is a continuous water phase between the pore water and that in the Buchner funnel. The ambient temperature and atmospheric pressure of the air in the impedance tube are measured to an accuracy of 0.1 °C and 100 Pa, respectively. These data are used to compensate for the related change in the sound speed and air density as suggested in the BS EN 10534-2.[14]

Although several types of sandy soil have been investigated, in this paper we report the data only for coarse and fine silica sand samples. These materials have flow resistivity and porosity values that are sufficiently representative of many outdoor porous surfaces[2,5,7,8] (see Table I). The sands used in the experiments are natural silica sands, prewashed, graded and supplied by WBB Minerals, UK. The basic physical and chemical characteristics for these materials are presented in Table I. The dry flow resistivity data presented in Table I were determined using the standard procedure.[15] Figure 2 shows scanning electron micrographs of the particles of coarse and fine sand. These illustrate that the particle size and shape distributions in the investigated soil samples are relatively uniform. The median grain size of the coarse sand is 0.65 mm and is approximately double of that of the fine sand. Figure 3 presents the probability density functions for the pore size distribution data for the two samples of sand obtained by using a standard water suction experiment.[12] The pore size data shown in Fig. 3 suggest that approximately 90% of pores in the sample of fine sand have radius s such that 51 μm$<s<$70 μm. Approximately 86% of the pores in the sample of coarse sand are such that 80 μm$<s<$105 μm. This choice of material samples provides the opportunity of studying the acoustic and related effects of two different bands of pore size distribution. It can be expected that the majority of pores in these materials will release water once the matric suction pressure exerted by the water in the burette ($P=\rho_w g H_p$) reaches the capillary pressure ($P_c=2\mu\cos\theta/\langle s\rangle$).[16] In the above expressions $\rho_w=998.23$ kg/m^3 is the density of water, $g=9.807$ m/s^2 is the acceleration of gravity, H_p is the difference between the effective level of water in the porous sample and the water level in the burette (see Fig. 1), $\mu=0.0728$ N/m is the surface tension of the water/air interface, $0\leqslant\theta\leqslant\pi$ is the contact angle which depends mainly on the rheological properties of the pore wall material and the chemistry of pore water and whether the water-air interface is advancing or receding, and $\langle s\rangle$ is the mean pore radius in the sample.

Fully saturated samples of these materials were prepared by pouring slowly dry sand into the Buchner funnel which was partially filled with water. Since the dry sand was always poured into the water, it is unlikely that air would be trapped in the pore voids, thus ensuring that the samples were fully saturated. Upon filling the Buchner funnel with saturated sand, the sand surface was leveled using a specially designed tool and any excess water and sand particles were carefully removed. Stratification of sand layers was unlikely to occur provided that sand particles were dropped from the same height and a constant water head was maintained above the sand surface. In other words, the deposition time of the sand particles was constant. Poorly graded sand samples having no fine particles (less than 75 μm) were used. Such a particle size range ensured that movement of particles in the water was minimal. This was confirmed by the visual observation of the color of the retreated water and by inspection of the sand samples after completing the experiment.

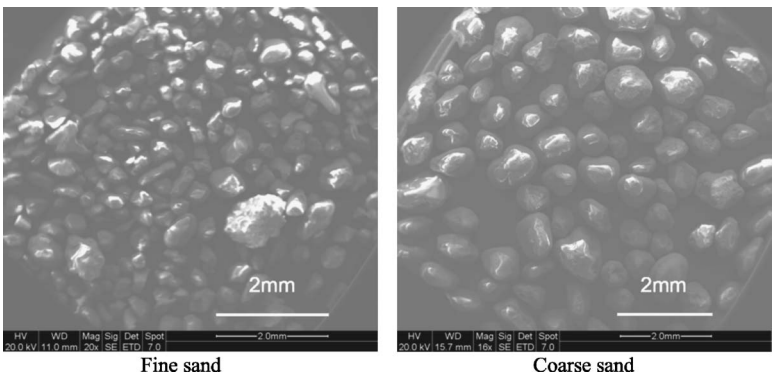


FIG. 2. Scanning electron microscope photographs of the investigated samples of fine and coarse sands.

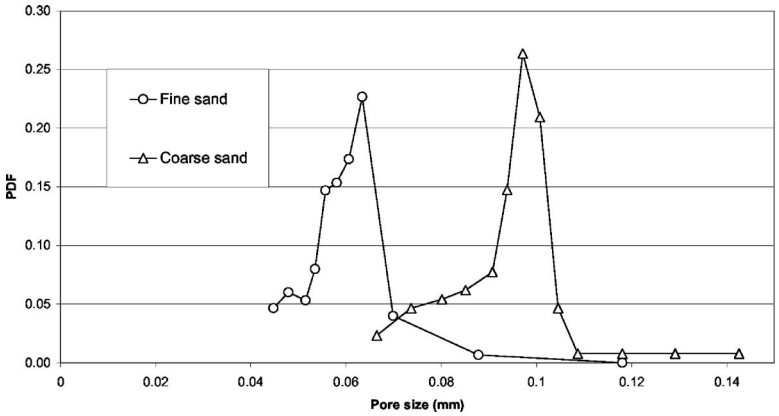


FIG. 3. Measured pore size distribution in fine and coarse sands.

The Buchner funnel was then attached to the impedance tube before starting the acoustic measurements. The acoustic impedance was initially measured at full saturation. An incremental increase in the suction head was applied to the sample and the outflow of water was measured at each stage. A stable distribution of water within the pore voids was reached when there was no further outflow of water. At each stage of equilibrium distribution of water within the sand sample (i.e., when $P_c = P$), the acoustic impedance of the sample was measured. The experiments were continued until the residual degree of saturation of 11%–15% was achieved, which was marked by no further draining of water even with further increase in the value of matric suction head. The average degree of saturation within the sand sample and the applied matric suction head were calculated at each stage (for further details see Ref. 17). In these experiments no change in the sample volume or shape was observed at the residual degrees of water saturation. In separate experiments the standard sample holder was loaded with samples of dry sand of equivalent mass and the acoustic impedance was measured.

III. RESULTS

The measured real and imaginary parts of the normalized surface acoustic admittance (β_s) for 50-mm-thick samples of coarse and fine silica sand as a function of frequency and for various values of the degree of water saturation, S (%), are presented in Figs. 4 and 6. The degree of saturation was determined from the following expression, $S = V_w / V_p \times 100\%$, where V_w is the volume of water in the sample and V_p is the total volume of pores. The small notches and peaks in the measured admittances (see Figs. 4 and 6) were caused by the limited signal-to-noise ratio in the impedance tube. The signal level was controlled to avoid any nonlinear behavior of the sample, i.e., to prohibit the generation of the higher-order harmonics which can be observed in loose particles of sand.[18]

A. Coarse sand

First, we consider the results for the sample of coarse sand. The data for this material show that the greatest effect of moisture is observed during the transition from the fully

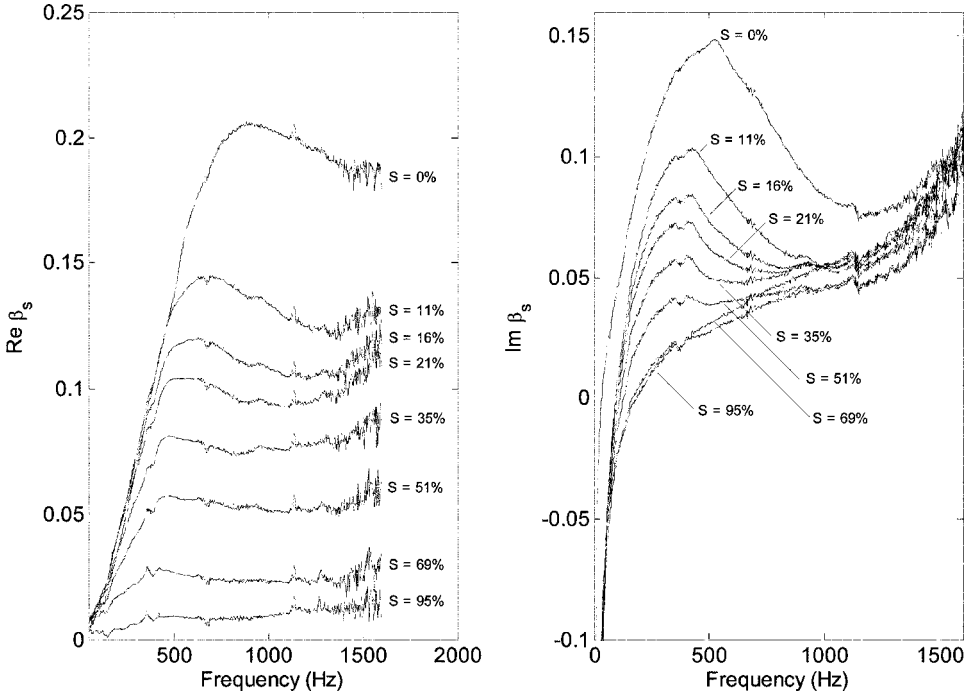


FIG. 4. Surface admittance of a 50 mm layer of coarse sand.

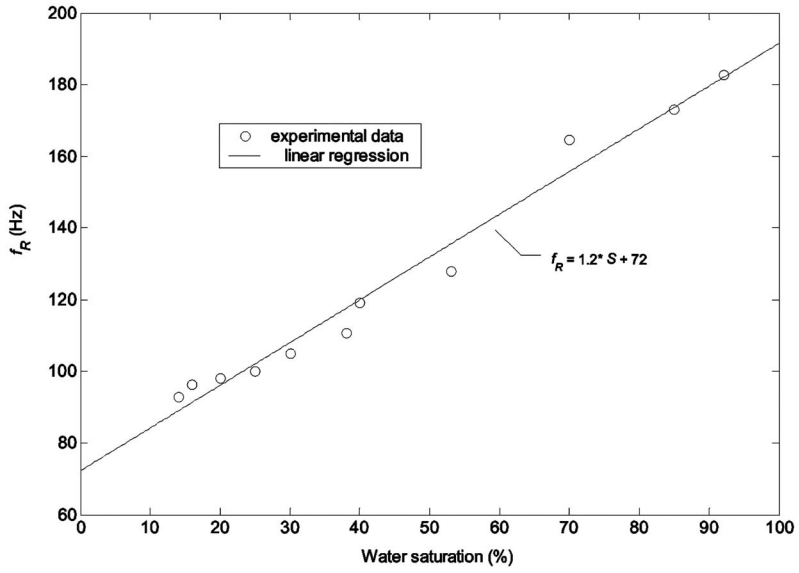


FIG. 5. The dependence of the frequency of the frame resonance in the sample of coarse sand on the degree of water saturation.

dry ($S=0\%$) to a low degree of saturation (e.g., $S\leqslant 11\%$). This transition results in up to 100% reduction in the value of the acoustic surface admittance (see Fig. 4). There is an almost 100% decrease in the real part of the admittance around 1000 Hz when the degree of saturation changes from 0% to 11%. The results suggest that for low degrees of saturation the effect of the moisture on the real and imaginary parts of the admittance is different. The maximum effect on the real part is observed at frequencies above 300–500 Hz. As the frequency decreases, the data for the real part of the surface admittance asymptotically approach the theoretical low frequency limit Re $\beta_s \to 0$. The effect of moisture on the imaginary part is most pronounced around 500 Hz and is less in the lower and higher frequency ranges. In the range of $11\% < S < 95\%$ the reduction in the real and imaginary parts of the admittance appears to depend almost linearly on the increase in the degree of saturation in the frequency range between 100 and 1500 Hz (see Fig. 4). The peak in the 500–600 Hz imaginary part of the admittance relates directly to the interference between the direct wave and the wave reflected from the rigid backing (effective water level). This behavior can be accurately predicted for $S\leqslant 11\%$ using the four-parameter Pade approximation model[19] or a similar model which is able to account for the finite layer thickness effect in a low porosity medium.

The real part of the admittance of coarse sand is almost independent of frequency for $S\geqslant 11\%$ above 500 Hz. At higher degrees of saturation the imaginary part of the admittance consistently increases with the frequency and approaches the dry sample value at $f\geqslant 1500$ Hz. In the low frequency range, below 100 Hz, and for $S\geqslant 11\%$ the effect of moisture on the imaginary part of the admittance is insignificant (see Fig. 4).

An interesting phenomenon is the presence of the frame resonance in the sample of partly saturated coarse sand. The frame of porous soil is formed by individual sand particles. These particles are loose and their density is considerably greater than that of air and comparable to that of water. At the medium and high frequencies the inertia force of an individual particle in a dry (low water saturated) sample is much greater than the viscous drag due to the oscillatory flow of air in the material pores. As a result, the particle (frame) vibration velocity is limited and the admittance value is governed largely by the oscillatory flow in the material pores. At the low frequencies the resisting inertia force becomes comparable to the viscous drag and a resonance occurs at the low frequency when Im $\beta_s=0$. The behavior of the frequency at which this resonance is observed in the sample of coarse sand is presented in Fig. 5 as a function of the degree of water saturation. There is a linear relationship between the frequency of the frame resonance and the degree of water saturation given by the following simple expression $f_R=1.2S+72$ (Hz).

Because of the problems in obtaining reliable information about the microscopic distribution of water between the individual sand grains it is difficult to determine whether the frequency of vibration mainly relates to: (i) the effective thickness of the top unsaturated layer of sand; (ii) the moisture-dependent stiffness of contact points between the grains[11]; or (iii) surface tension effects in the closed pores.[20] It is possible that a combination of the three effects controls the frequency of the frame resonance.

B. Fine silica sand

Second, we consider the behavior of the acoustic surface admittance of the sample of fine sand (Fig. 6). The results show that the real part of the admittance of the fine sand has an approximately linear dependence on the degree of water saturation. There is a proportional reduction in the value of the real part of the admittance for a given increase in the degree of saturation across the considered frequency range. Generally, there is a 5 to 15-fold reduction in the real part of the admittance as the degree of saturation increases from 0% to 95%. The imaginary part of the admittance is hardly affected during the transition from dry to a low degree of saturation (i.e., from $S=0\%$ to $S=15\%$) and any further increase in the degree of saturation results only in a marginal decrease in the imaginary part. The imaginary part of the admittance

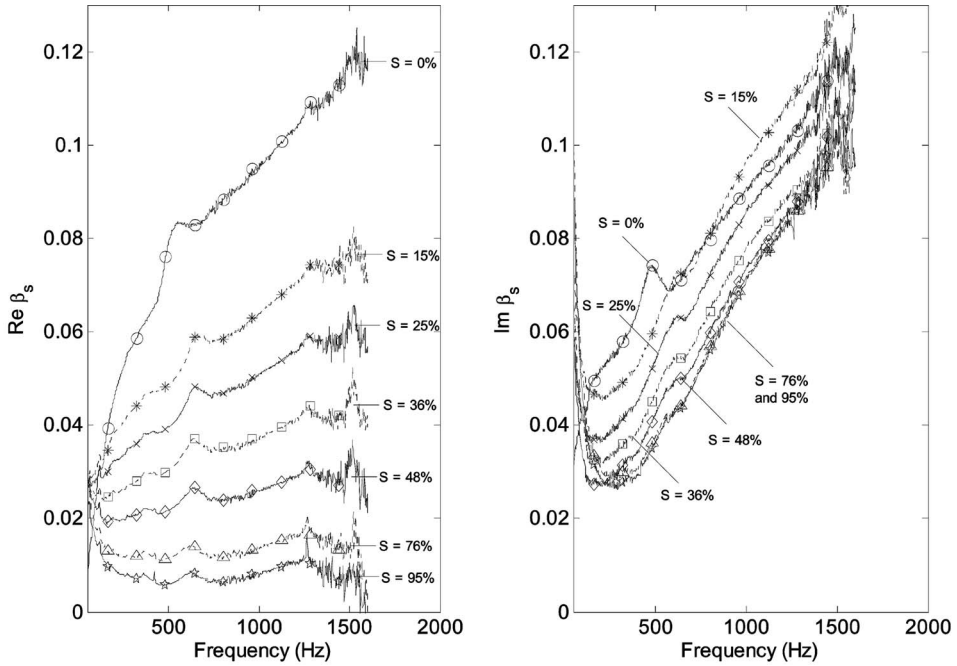


FIG. 6. Surface admittance of a 50 mm layer of fine silica sand.

appears unaffected by moisture for $S \geqslant 76\%$. Indeed the imaginary part is approximately constant for all the degrees of saturation investigated in this work.

Frame resonance can be also observed in the case of fine silica sand, resulting in the nonrigid frame behavior visible in the data for both the real and imaginary parts of the surface admittance in the low frequency range (below 100 Hz) for $S \neq 0\%$. Unlike the case of coarse sand, the imaginary part of the surface admittance of fine silica sand stays positive across the considered frequency range. The imaginary part reaches a minimum between 200 and 300 Hz and then increases as the frequency decreases (see Fig. 6). It is difficult to obtain the exact frequency of this minimum for a given degree of saturation and there does not seem to be any clear dependence between the position of this minimum and the degree of water saturation. In general, the data suggest that the value of the minimum decreases with increasing values of S.

IV. APPLICABILITY OF GROUND IMPEDANCE MODELS

In this section we consider the performance of two models for the acoustic properties of rigid-frame porous media when applied to the characterization of porous sands with an arbitrary degree of saturation. The models selected in this study have been used widely to characterize the acoustic behavior of porous soils and are based on effective flow resistivity. These models predict directly the acoustic impedance which can then be translated into the admittance data used in a majority of formulas for the acoustic excess attenuation due to the presence of porous ground (e.g., Refs. 7 and 8). Since there is no standard methodology for selecting the nonacoustic parameters from the measured impedance data, we apply regression analysis directly to the measured impedance data in order to illustrate the ability of these models to predict the frequency/moisture dependent behavior of porous soils.

A. Delany and Bazley model

A common method of modeling the acoustic admittance of porous ground, $\beta_s = 1/z_b$, is to use the Delany and Bazley model,[6] i.e., Eq. 35 in Ref. 7,

$$z_b = 1 + 9.08\left(\frac{1000f}{\sigma}\right)^{-0.75} - i11.9\left(\frac{1000f}{\sigma}\right)^{-0.73}, \qquad (1)$$

where σ is the effective flow resistivity (Pa s m^{-2}), $i = \sqrt{-1}$ and time dependence $e^{+i\omega t}$ is understood. Cramond and Don used a two-microphone technique for their outdoor sound propagation experiments and the full Delany and Bazley model for a hard-backed layer of a partly saturated soil to deduce the effective flow resistivity.[5] The authors suggested the following empirical formula for the flow resistivity as a function of the water saturation[5]

$$\sigma = 150 \times 10^3 - 360 \times 10^3 \ln(1 - S/S_{\max}), \quad S < S_{\max}, \qquad (2)$$

where $S_{\max}$ is the saturation value which corresponds to the dry porosity of porous soil. Figure 7 presents the flow resistivity predicted by formula (2) for a soil with $S_{\max} = 40\%$ and the deduced flow resistivity for coarse and fine sands as a function of the degree of saturation. The deduced values of the flow resistivity were obtained using the direct search optimization method[21]

$$F(\sigma) = \int_{\omega_{\min}}^{\omega_{\max}} |z_{\exp}(\omega,\sigma) - z_b(\omega,\sigma)| d\omega \to \min, \qquad (3)$$

where $z_{\exp}(\omega,\sigma)$ is the measured data for the normalized surface impedance, $z_b(\omega,\sigma)$ is the impedance predicted by expression (1), $\omega_{\min} = 2\pi \times 200$ and $\omega_{\max} = 2\pi \times 1200$ rad/s.

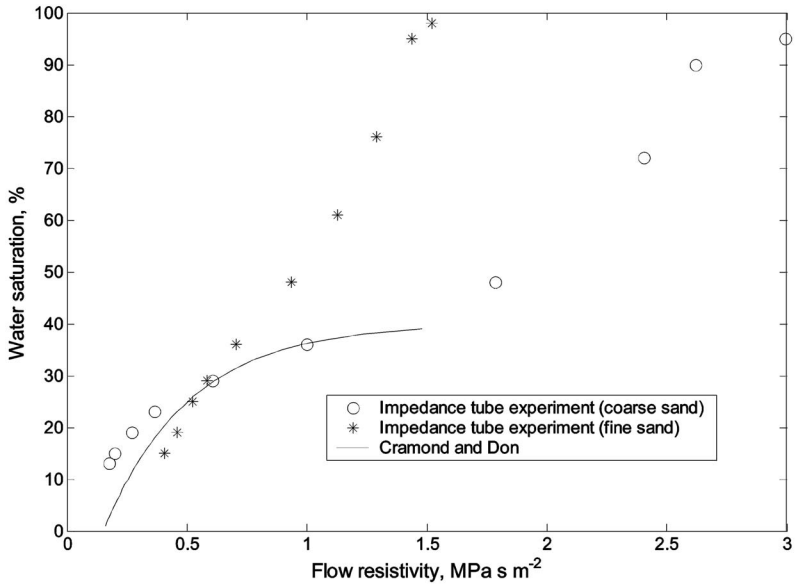


FIG. 7. Comparison of the effective flow resistivities deduced from the impedance tube data for fine sand and those predicted by the Cramond and Don formula for grassland (see Ref. 5).

The results of the optimization (see Fig. 7) show that the Cramond-Don formula captures satisfactorily the behavior of the deduced flow resistivity of the sample of coarse sand for $S<40\%$. The prediction of the behavior of the deduced flow resistivity of fine sand is only satisfactory for $S<30\%$.

B. Modified Delany and Bazley model

Close examination of the behavior of the measured data and the results predicted for the deduced values of σ [expression (3)] reveal that the measured real and imaginary parts of the admittance cannot be accurately modeled by adjusting the effective flow resistivity in expression (1).[22] A particular value of the flow resistivity in the original Delany and Bazley formula [expression (1)] can only provide a satisfactory fit to the experimentally determined values of either the real or imaginary part of the admittance. An alternative[22] is to modify expression (1) by adopting two different values of the effective flow resistivity (σ_R) and (σ_I) which allows a more accurate and simple method of predicting the real and imaginary parts of the acoustic impedance of fine sand at an arbitrary degree of the water saturation. It is straightforward to apply a least-mean-squares (lms) regression analysis and expression (1) to find the optimal flow resistivity values that would provide the best fits to the measured real and imaginary parts of the acoustic impedance. Applying the lms regression to (1) we find that the real part of the impedance is best predicted with the following value of the effective flow resistivity:

$$\log \sigma_R = \frac{1}{0.75} \log\left\{ \frac{\sum_{m=1}^{N} f_m^{-0.75}(\mathrm{Re}\, z_{\mathrm{exp},m} - 1)}{a \sum_{m=1}^{N} f_m^{-1.5}} \right\}. \quad (4)$$

Similarly, the best fit value of the flow resistivity for the measured imaginary part of the impedance is

$$\log \sigma_I = \frac{1}{0.73} \log\left\{ \frac{-\sum_{m=1}^{N} f_m^{-0.73}\, \mathrm{Im}\, z_{\mathrm{exp},m}}{b \sum_{m=1}^{N} f_m^{-1.46}} \right\}. \quad (5)$$

Here $a=5.11\times10^{-2}$, $b=7.683\times10^{-2}$, N is the number of data points in the impedance spectrum and $z_{\mathrm{exp},m}$ is the measured surface impedance at the frequency f_m. Expression (1) can now be modified to include the two separate values of the flow resistivity, i.e.

$$z_b = 1 + 9.08\left(\frac{1000f}{\sigma_R}\right)^{-0.75} - i11.9\left(\frac{1000f}{\sigma_I}\right)^{-0.73}. \quad (6)$$

We will call expression (6) the *modified* Delany and Bazley model.

C. Two-parameter Attenborough model

An alternative practical method to predict the acoustic properties of porous ground is to use the two-parameter model proposed by Attenborough.[23] This model requires a knowledge of the effective flow resistivity (σ_e) and the effective rate of change in the porosity with the layer depth (α_e)

$$z_b = 0.436\sqrt{\frac{\sigma_e}{f}} - i\left(0.436\sqrt{\frac{\sigma_e}{f}} + 9.75\alpha_e/f\right). \quad (7)$$

The effective flow resistivity in the above model can be adjusted to match the measured data for the real part of the acoustic impedance. In order to match the measured data for the imaginary part of the acoustic impedance the effective rate of change in porosity can be adjusted. We can use expression (7) and the lms method to determine the values of σ_e and α_e so that

TABLE II. The deduced values of the effective flow resistivities in the modified (two-parameter) Delany and Bazley model [expression (6)].

Material	Degree of saturation %	Effective flow resistivity (σ_R), kPa s m^{-2}	Effective flow resistivity (σ_I), kPa s m^{-2}	Error in the real part (E_R),%	Error in the imaginary part (E_I),%
Coarse sand	0	112.5	106.4	47.2	18.6
	11	224.1	134.8	47.6	26.2
	51	835.5	537.0	18.5	34.5
	95	244.9	1707	37.5	5.9
Fine sand	0	347.8	276.9	12.1	16.5
	15	332.6	448.4	5.4	15.6
	48	367.3	977.4	23.3	4.7
	95	123.8	1392	48.7	7.0

$$\log \sigma_e = -2 \log\left\{ \frac{0.436 \sum_{m=1}^{N} f_m^{-1}}{\sum_{m=1}^{N} \mathrm{Re}(z_{\mathrm{exp},m}) f_m^{-0.5}} \right\} \tag{8}$$

and

$$\alpha_e = -\frac{\sum_{m=1}^{N} \mathrm{Im}\, z_{\mathrm{exp},m} f_m^{-1} + 0.436 \sigma_e^{0.5} \sum_{m=1}^{N} f_m^{-1.5}}{9.48 \sum_{m=1}^{N} f_m^{-2}}. \tag{9}$$

V. DISCUSSION

The models presented in Sec. IV have been used to deduce the relevant nonacoustic parameters of 50 mm-thick samples of fine and coarse sand in the frequency range of 200–1600 Hz for three different degrees of saturation, $S < 95\%$ and in the frequency range of 400–1600 Hz for $S = 95\%$. In the case of the modified Delany and Bazley model [expression (6)] the deduced parameters were the two values of the effective flow resistivity [expressions (4) and (5)]. In the case of the two-parameter Attenborough model [expression (7)] the deduced parameters were the effective flow resistivity [expression (8)] and the effective rate of change in the porosity with the layer depth [expression (9)].

The results of this analysis are compiled in Tables II and III. Tables II and III also present the relative errors in predicting the real and imaginary parts of the impedance

$$E_R = \frac{\sqrt{\langle (\mathrm{Re}\, z_{\mathrm{exp}} - \mathrm{Re}\, z_{\mathrm{th}})^2 \rangle}}{\langle \mathrm{Re}\, z_{\mathrm{th}} \rangle} \times 100\,\% \quad \text{and}$$

$$E_I = \frac{\sqrt{\langle (\mathrm{Im}\, z_{\mathrm{exp}} - \mathrm{Im}\, z_{\mathrm{th}})^2 \rangle}}{\langle \mathrm{Im}\, z_{\mathrm{th}} \rangle} \times 100\,\%, \tag{10}$$

where z_{th} is the predicted acoustic impedance and $\langle \ldots \rangle$ stands for arithmetical averaging in the frequency domain. Figures 9 and 10 present the comparison between the measured and predicted frequency dependent behavior of the acoustic surface impedance for the four selected degrees of water saturation. The predictions were obtained using expressions (6) and (7) and the corresponding values of the microstructural parameters listed in Tables II and III.

The results show that changes in the degree of saturation strongly affect the values of the microstructural parameters in the considered impedance models. In the case of the modified Delany and Bazley model there is a progressive discrepancy between the values of the flow resistivity required to provide the best fit to the measured real and imaginary parts of the acoustic impedance. Specifically, in the case of coarse sand at $S = 95\%$, the ratio of $\sigma_I / \sigma_R \approx 7$. In the case of fine sand at the same degree of saturation this ratio is $\sigma_I / \sigma_R \approx 12$. This effect cannot be accounted for by the original Delany and Bazley formula [expression (1)]. In the case of

TABLE III. The deduced values of the effective flow resistivity and rate of change in the porosity in the two-parameter Attenborough model [expression (7)].

Material	Degree of saturation %	Effective flow resistivity (σ_e), kPa s m^{-2}	Effective rate of change in the porosity (α_e), m^{-1}	Error in the real part (E_R),%	Error in the imaginary part (E_I), %
Coarse sand	0	48.2	−13.5	41.1	30.0
	11	106.6	−82.3	37.5	30.1
	51	472.6	−140.9	12.9	21.5
	95	89.9	1285	41.9	5.9
Fine sand	0	145.5	−27.7	5.9	8.4
	15	144.2	132	10.5	9.8
	48	150.5	623	28.5	6.7
	95	40.9	1126	52.7	7.7

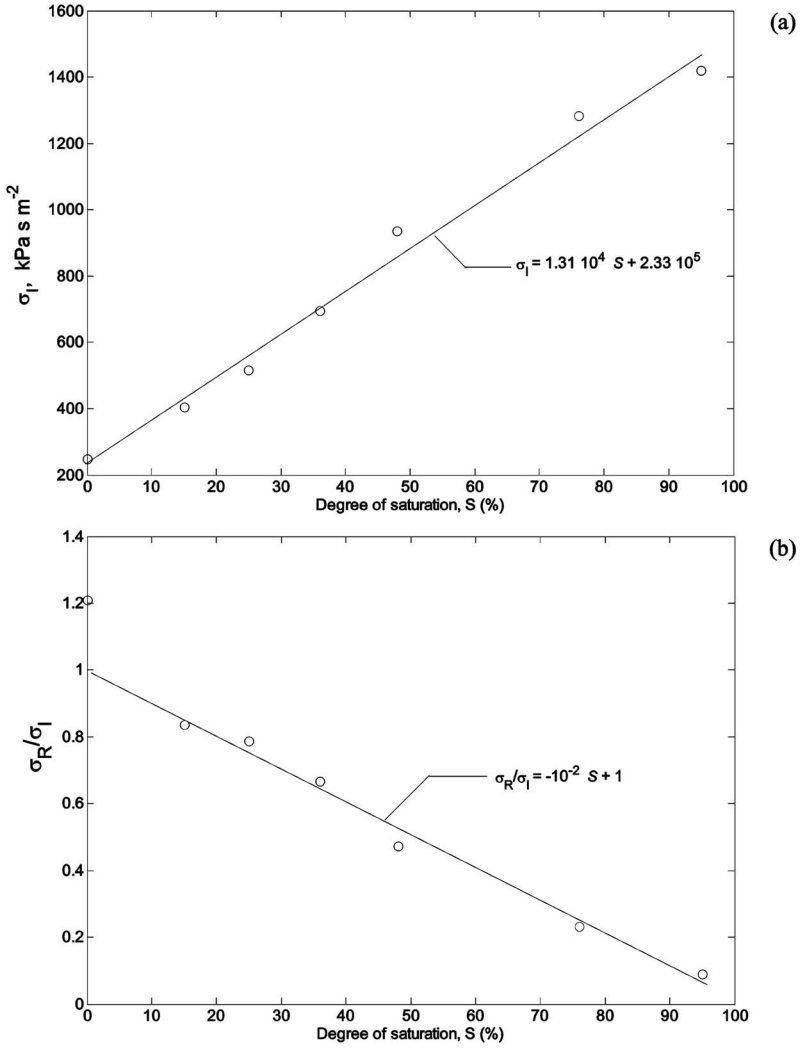


FIG. 8. The dependence of the deduced values of the effective flow resistivity, σ_I, (a) and the ratio of σ_R/σ_I (b) on the degree of water saturation in fine silica sand.

fine silica sand there is a simple relationship between the deduced effective flow resistivities (σ_I, σ_R) and the degree of saturation (S) as presented in Fig. 8. As a result two empirical relations have been proposed[22]

$$\sigma_I = 1.31 \times 10^4 S + 2.33 \times 10^5 \text{ Pa s m}^{-2} \quad \text{and}$$
$$\sigma_R/\sigma_I = -10^{-2} S + 1. \tag{11}$$

Modified formulae (6) and (11) seem to be accurate to better than 50% when used to predict the acoustic impedance of fine sand within the frequency range considered in this work. In general, the model appears more accurate in predicting the behavior of the imaginary part of the acoustic impedance at frequencies above 300–400 Hz. Specifically, there is a remarkably good fit with $E_R < 10\%$ between the measured and predicted imaginary part for $S=95\%$ in the case of coarse sand and for $S=48\%$ and 95% in the case of fine silica sand (see Table II and Figs. 9 and 10).

A similar accuracy of prediction can be achieved by using the two-parameter Attenborough model [expression (7)]. The results shown in Table III suggest that the effective flow resistivity and rate of change in the porosity used by the Attenborough model can be adjusted to capture the basic behavior of both the real and imaginary parts of the acoustic impedance of partly saturated sands investigated in this work in the adopted frequency range of 200–1600 Hz. The relationship between the degree of saturation and the two microstructural parameters used in the Attenborough model is complex. There seems to be a gradual increase in the deduced value of the effective flow resistivity for $S<50\%$. At the high degree of saturation of 95% a sudden reduction in the flow resistivity can be observed. It is interesting to note the progressive increase in the modulus of the deduced value of α_e with the increasing degree of water saturation (see Table III). We note also that the effective rate of change in the porosity is able to take either positive or negative sign which is equivalent to the positive and negative porosity gradients across the sample depth, respectively. In the case of coarse sand this change is from the negative to positive and it occurs between $S=51\%$ and $S=95\%$. In the case of fine silica sand the negative sign of α_e changes to positive between $S=0\%$ and $S=15\%$. In the case of coarse sand the accuracy of the Attenborough model estimated according to formulae (10) in the frequency range of 200–1600 Hz is better than 42% ($S=95\%$). In the case of fine silica sand this model is able to predict the imaginary part of the impedance to within better than 10%, but it seems less accurate in predicting the real part for which the discrepancy between the

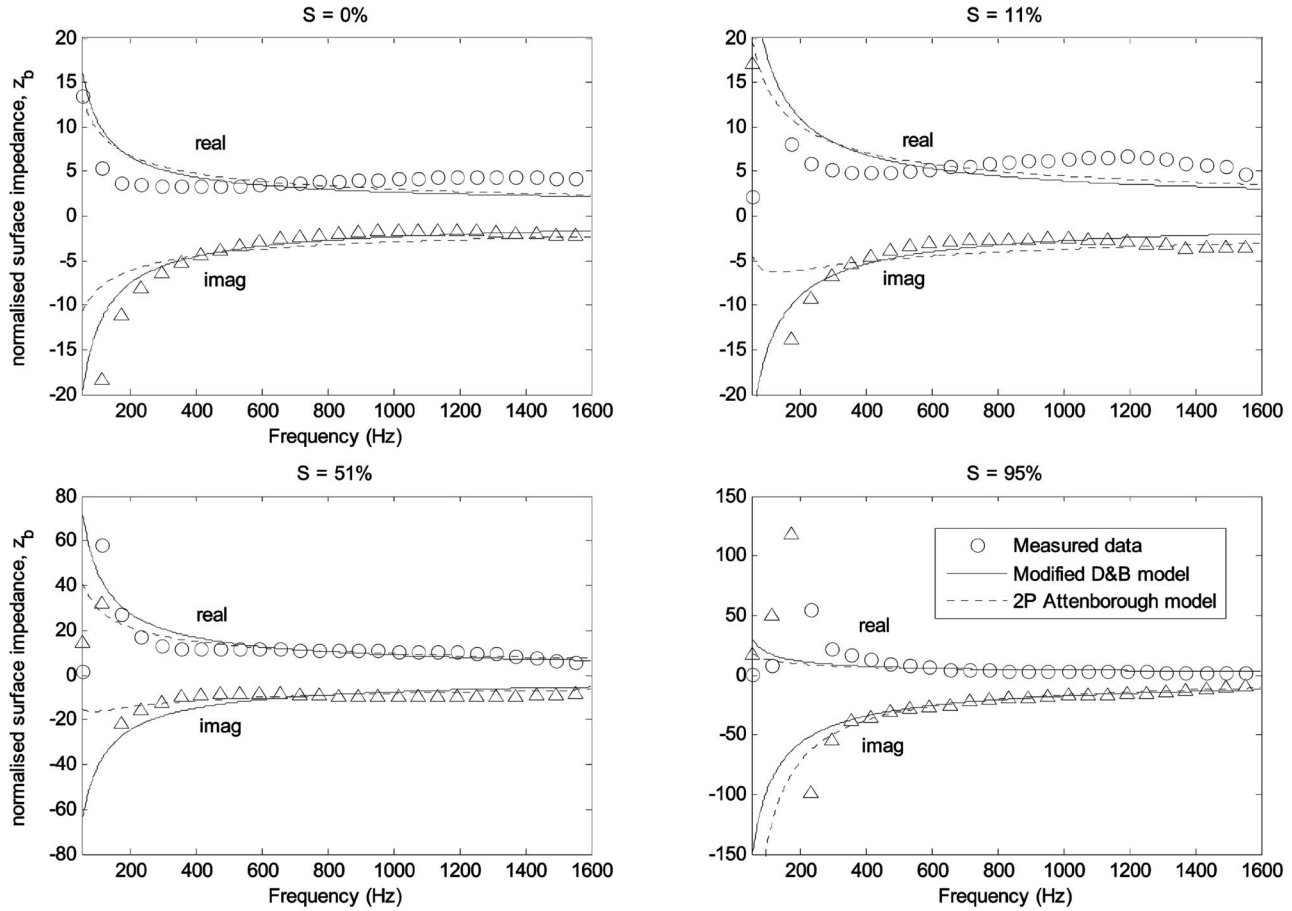


FIG. 9. The measured and predicted frequency dependence of the normalized surface impedance of a 50 mm layer of coarse sand.

measured data and the model increases progressively with the increasing degree of saturation (E_R=5.9% for S=0% and E_R=52.7% for S=95%).

We note that the real part of the effective flow resistivity for coarse sand has a peak value at 51% saturation which is predicted by the modified Delany and Bazley (Table II) and two-parameter Attenborough (Table III) models. The penetration depth $[1/k(\omega,S)]$ in coarse sand is greater than the thickness of investigated sample. As a result, the measured acoustic surface admittance of coarse sand is controlled by the expression $\beta_s(\omega,S)=\beta(\omega,s)\tanh(ik(\omega,S)d)$ rather than $\beta_s(\omega,S)\cong\beta(\omega,s)$, where d is the mean water level below the sample surface and $\beta(\omega,s)$ and $k(\omega,S)$ are frequency/moisture dependent characteristic admittance and wave number in sand, respectively. The frequency-dependent behavior of $\beta_s(\omega,S)$ is likely to be the combination of the change in the mean water level and size distribution of open, interconnected pores. Such a behavior of the surface impedance, particularly at the low frequencies (see Fig. 9 and 10), is difficult to predict theoretically because of the lack of microstructural data. Although not reported in detail here, attempts have been made to apply other multiple parameter models for the acoustic properties of rigid frame porous media, (e.g. Refs. 19, 24, and 25). It has been found that inclusion of more detailed microstructural parameters does not offer any significant improvement in the fit to the experimental data for materials with $S>10\%-15\%$. This can be attributed to two main reasons: (i) in the low frequency range the rigid frame approximation is no longer valid because the effect of the frame vibration on the acoustic admittance is comparable with that of the viscous friction in the oscillatory two-phase flow in the partly saturated material pores (see the low frequency behavior of the impedance shown in Figs. 9 and 10) and (ii) wet sandy soil is no longer a homogeneous porous material but a patchily saturated medium with a complex combination of water-blocked and open interconnected pores resulting in the complicated variation in the deduced values of σ_e, α_e, σ_R and σ_I (see Tables II and III).

Moreover, it is uncertain whether models that allow for an elastic frame are able to offer any advantages. Recent attempts (e.g., Johnson,[10] Tserkovnyak[26] and Umnova[27]) have illustrated the problems in applying the full Biot formulation to predict low frequency sound propagation in a partly saturated porous medium. Difficulties arise due to the lack of reliable data on the poro-elastic structure and on the nature of the mechanical bonds between individual, wetted particles. The distribution of water in the sample and the connectivity of air-filled pores cannot be easily characterized. This leads to problems in determining the sign and value of the porosity and permeability gradients, which are likely to change as the water is progressively extracted from the sample. There is also a considerable uncertainty about the mean water level in the sample when the water is removed from the larger and

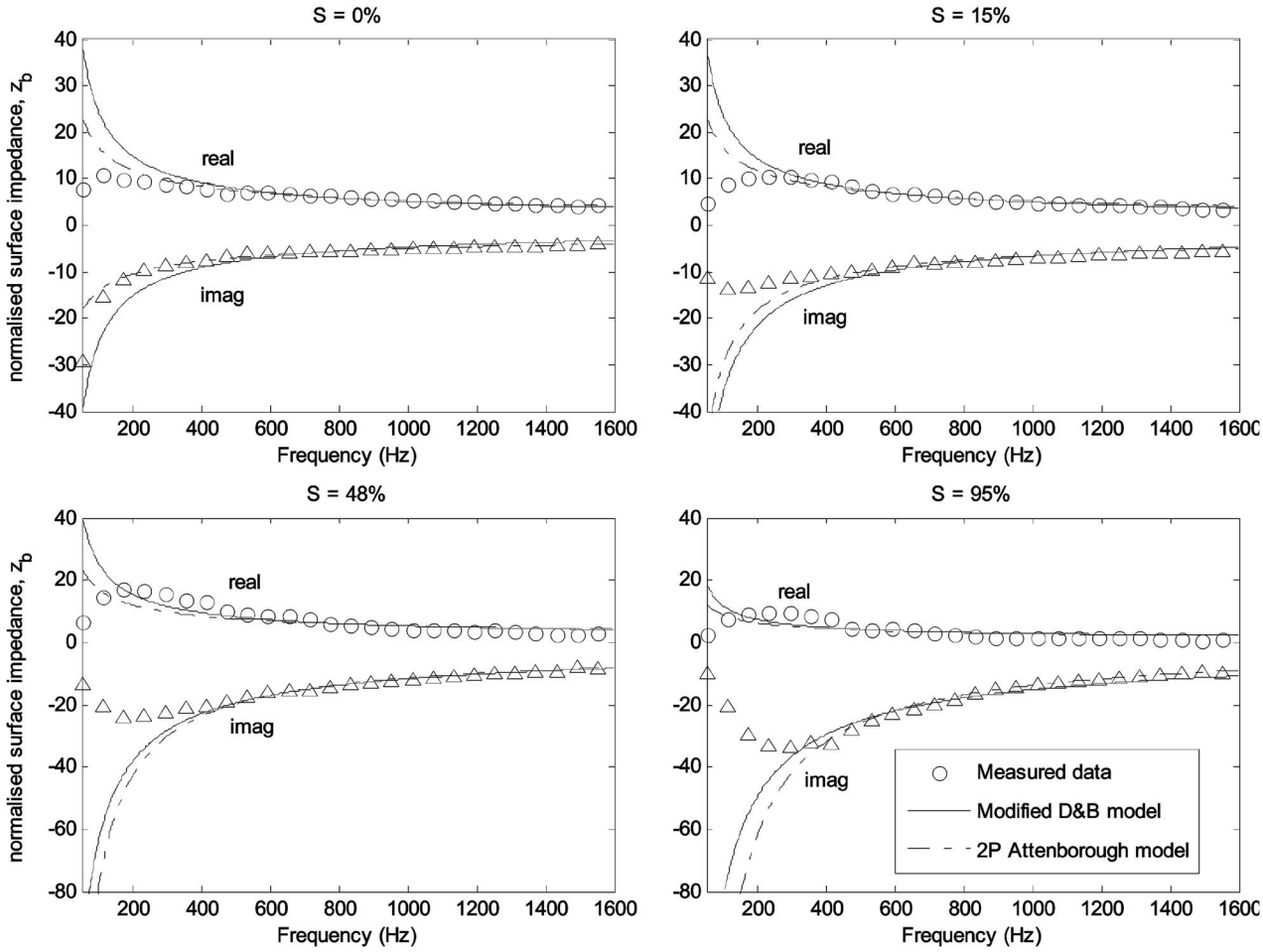


FIG. 10. The measured and predicted frequency dependence of the normalized surface impedance of a 50 mm layer of fine silica sand.

smaller pores which are often interconnected. From the acoustical point of view, a layer of partly saturated sand in the impedance tube appears to be neither hard backed nor semi-infinite.

VI. CONCLUSIONS

An experimental facility has been developed and used for the first time to study the effect of moisture on the acoustic admittance of porous sands and sandy soils with the precisely controlled degree of water saturation. It has been shown that the acoustic behavior of partly saturated sandy soils is a very complex phenomenon which cannot be simply predicted.

Specifically, it has been shown that small (10%–20%) variations in the pore moisture content can result in considerable variations in the measured surface acoustic admittance of the porous soil. In the case of coarse sand a change from the dry state to the 11% saturated state results in 100% decrease in the real part of the surface admittance at frequencies close to 1000 Hz. In the case of fine silica sand a similar reduction in the admittance is observed when the degree of saturation increases from 0% to 29%.

It has been shown that the range of values of the effective flow resistivity required to model the acoustic behavior of coarse and fine sands is considerable. As the degree of water saturation of fine sand increases from 15% to 95% the value of the optimal effective flow resistivity used in the original Delany and Bazley model increases eightfold. In the case of coarse sand the same change in the degree of water saturation results in a 17-fold increase in the effective flow resistivity. According to the new EU HARMONOISE prediction method[7] this increase is equivalent to the change from the normal uncompacted ground (80 kPa s m^{-2}) to compacted dense ground (2000 kPa s m^{-2}). This range implies a significant uncertainty when modeling ground impedance with a single effective flow resistivity. It is proposed that the effect of moisture should be accounted for in the Standard.

It has been shown that the Delany and Bazley model in its original form is not able to predict the frequency dependence of both the real and imaginary parts of the admittance of water-saturated samples of sand. In the case of fine sand the prediction error can exceed 700%. Moreover, below 300 Hz the measured admittance data for coarse sand shows distinctive frame resonance behavior. In the case of coarse sand the resonance frequency of frame vibration is in the range of 90–190 Hz. In this regime the rigid frame approximation is no longer valid and a new model is needed to account for the following effects: (i) viscous absorption in partly saturated pores; (ii) the vibration of the top unsaturated layer of sand; (iii) the moisture-dependent stiffness of contact points between the grains; (iv) surface tension of water in the closed pores.

A modified two-parameter form of the Delaney and Bazley expression for the characteristic impedance [see expression (6)] has been suggested introducing two different values of the effective flow resistivity, σ_R and σ_I, for the real and imaginary parts, respectively. It has been shown that at particular high saturation states, e.g. for $S=95\%$, there is the requirement for $\sigma_I \cong 12\sigma_R$. Linear regression formulas have been presented to deduce the best fit values of these two flow resistivities from the measured acoustic impedance data [expressions (4) and (5)]. In the case of fine silica sand it is possible to relate the two flow resistivities directly to the degree of water saturation [expressions (11)]. The new empirical expressions offer a considerable improvement in terms of the accuracy of the impedance prediction of better than 50%.

The performance of the two-parameter Attenborough model in the frequency range of 200–1600 Hz has also been studied. Two regression formulas [expressions (8) and (9)] have been proposed to deduce the best fit effective flow resistivity and the rate of porosity change from the measured real and imaginary parts of the acoustic admittance. It has been shown that the model is capable of predicting the impedance of partly saturated coarse and fine silica sand with accuracy of better than 53%. On the basis of the data reported here, the two-parameter Attenborough model should be regarded as a better alternative to the original Delany and Bazley single-parameter model advocated in the HARMONOISE prediction method as it is able to provide useful indications of the value and sign of the porosity gradient in partly saturated porous soils.

ACKNOWLEDGMENTS

The authors are grateful to the School of Engineering, Design and Technology for the financial support of this work. The authors express particular gratitude to Clive Leeming for his advice on the design of the instrumentation and for its accurate implementation. The authors are grateful to Siow N. Ting and Pinkie Letang for their help with the experiments which often run into the night. The authors are grateful to Professor David C. Hothersall for his advice on the format of this manuscript. The authors would like to thank Professor Keith Attenborough for his technical comments and suggested corrections to the style and grammar.

[1]M. J. M. Martens, L. A. M. Vanderheijden, H. H. J. Walthaus, and W. J. J. M. Vanrens, "Classification of soils based on acoustic-impedance, air-flow resistivity, and other physical soil parameters," J. Acoust. Soc. Am. **78**, 970–980 (1985).

[2]K. Attenborough and S. Taherzadeh, "Propagation from a point source over a rough finite impedance boundary," J. Acoust. Soc. Am. **98**, 1717–1722 (1995).

[3]N. N. Voronina and K. V. Horoshenkov, "A new empirical model for the acoustic properties of loose granular media," Appl. Acoust. **64**, 415–432 (2003).

[4]P. J. Dickinson and P. E. Doak, "Measurement of normal acoustic impedance of ground surfaces," J. Sound Vib. **13**, 309–322 (1970).

[5]A. J. Cramond and C. G. Don, "Effects of moisture-content on soil impedance," J. Acoust. Soc. Am. **82**, 293–301 (1987).

[6]DGRM—Technical Report No. HAR32TR-030715-DGMR03—Harmonoise WP 3 D17 Engineering method for road traffic and railway noise (Draft) (2003).

[7]M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).

[8]T. F. W. Embleton, J. E. Piercy, and G. A. Daigle, "Effective flow resistivity of ground surfaces determined by acoustical measurements," J. Acoust. Soc. Am. **74**, 1239–1244 (1983).

[9]K. V. Horoshenkov and S. N. Chandler-Wilde, "On the behavior of some impedance models for the acoustic properties of rigid frame porous media," *Proceedings of the 6th International Congress on Sound and Vibration*, Copenhagen, 653–660 (1999).

[10]D. L. Johnson, "Theory of frequency dependent acoustics in patchy-saturated porous media," J. Acoust. Soc. Am. **110**, 682–694 (2001).

[11]F. D. Shields, J. M. Sabatier, and M. Wang, "The effect of moisture on compressional and shear wave speeds in unconsolidated granular material," J. Acoust. Soc. Am. **108**, 1998–2004 (2000).

[12]P. Leclaire, M. Swift, and K. V. Horoshenkov, "Specific area from water-suction porosimetry in application to porous acoustic materials," J. Appl. Phys. **84**, 6886–6890 (1998).

[13]J. Bear, *Hydraulics of Groundwater* (McGraw–Hill, New York, 1979), pp. 193–197.

[14]British Standard BS EN 10534-2: Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes—Transfer function method (2001).

[15]BS EN 29053:1993, ISO 9053:1991. Acoustics—Materials for acoustical applications—Determination of airflow resistance (1999).

[16]J. Bear, *Dynamics of Fluids in Porous Media* (American Elsevier, 1972), pp. 439–449.

[17]R. S. Sharma and M. H. A. Mohamed, "An experimental investigation of LNAPL migration in an unsaturated/saturated sand," Eng. Geol. (Amsterdam) **70**, 305–313 (2003).

[18]D. Donskoy, A. Ekimov, N. Sedunov, and M. Tsionskiy, "Nonlinear seismo-acoustic land mine detection and discrimination," J. Acoust. Soc. Am. **111**, 2705–2714 (2002).

[19]K. V. Horoshenkov and M. J. Swift, "The acoustic properties of granular materials with pore size distribution close to log-normal," J. Acoust. Soc. Am. **110**, 2371–2378 (2001).

[20]P. B. Nagy and A. H. Nayfeh, "Generalized formula for the surface stiffness of fluid-saturated porous-media containing parallel pore channels," Appl. Phys. Lett. **67**, 1827–1829 (1995).

[21]G. R. Reklaitis, A. Ravindran, and K. M. Ragsdell, *Engineering Optimization Methods and Applications* (Wiley, New York, 1983), pp. 150–165.

[22]K. V. Horoshenkov, M. H. A. Mohamed, and S. Adamidou, "Acoustic Properties of Partly Saturated Porous Soils," *CD-ROM Proceedings of the International Symposium on Acoustics of Poro-elastic Medium (SAPEM)*, Lyon, France, 7–9 December 2005.

[23]K. Attenborough, "Ground parameter information for propagation modelling," J. Acoust. Soc. Am. **92**, 418–427 (1992).

[24]D. K. Wilson, "Relaxation-matched modeling of propagation through porous media, including fractal pore structure," J. Acoust. Soc. Am. **94**, 1136–1145 (1993).

[25]Y. Champoux and J.-F. Allard, "Dynamic tortuosity and bulk modulus in air-saturated porous media," J. Appl. Phys. **70**, 1975–1979 (1991).

[26]Y. Tserkovnyak and D. L. Johnson, "Capillary forces in the acoustics of patchy-saturated porous media," J. Acoust. Soc. Am. **114**, 2596–2606 (2003).

[27]O. Umnova, *Private correspondence*, University of Bradford and University of Salford, UK (2006).

Constrained comparison of ocean waveguide reverberation theory and observations

Charles W. Holland[a)]
The Pennsylvania State University, Applied Research Laboratory, State College, Pennsylvania 16804-0030

(Received 5 May 2006; revised 20 July 2006; accepted 21 July 2006)

Measurements of long-range (order 10^4 m) shallow-water reverberation in the Straits of Sicily at 900 and 1800 Hz are compared with theoretical predictions. All of the required environmental inputs for the theory are obtained independently, that is to say there are no free parameters. The reflection coefficient and the scattering strength are measured by direct path methods; both quantities show strong frequency dependence. The theoretical reverberation predictions using these measurements are in good agreement with directional reverberation data, i.e., within the expected uncertainty bounds. The good agreement suggests that the supporting environmental measurement techniques are robust and that the physics associated with reverberation in a waveguide is reasonably well understood, at least in simple environments. The ability to independently measure the seabed scattering strength and reflection coefficient is a crucial step for the advancement of inverse methods using reverberation (e.g., rapid environmental assessment) inasmuch as it provides the means for quantitatively measuring the robustness of those methods. [DOI: 10.1121/1.2338811]

PACS number(s): 43.30.Gv, 43.30.Ma, 43.30.Zk [AIT] Pages: 1922–1931

I. INTRODUCTION

The ability to predict acoustic reverberation in shallow water ocean waveguides is important for the design and employment of active sonar. Hence, reverberation models and measurements have been in development for many decades. Eyring *et al.*[1] conducted early measurements of reverberation at 24 kHz and developed sonar equation models to explain their observations. They noted that the seabed characteristics play a significant, often dominant, role in shallow water reverberation and analyzed data from seabeds characterized by rock, mud, and a sand-mud mixture. For the rocky bottom, they suggested that the bottom backscattering strength is proportional to $\sin\theta$, where θ is the grazing angle. Urick[2] conducted shallow water measurements at nearly the same frequency, making the tacit assumption that scattering strength is independent of grazing angle and showed that the bottom scattering over a sandy site is roughly isotropic in azimuth. MacKenzie concluded[3] that the scattering strength was independent of frequency over seven octaves (later observations, e.g., Refs. 4 and 5, show a much richer frequency and grazing angle dependence than was concluded from the early measurements) and developed an energy flux type model[6] in integral form using a scattering kernel ($\mu \sin\theta_i \sin^n\theta_o$) where θ_i and θ_o indicate incident and scattered angles, respectively.

Over the years, modeling advancements were made using a variety of approaches, including: energy flux approaches (e.g., Refs. 4 and 7), normal mode methods (e.g., Refs. 8 and 9), ray theory (e.g., Ref. 10), and the parabolic equation (e.g., Ref. 11). Over time, many of these approaches have advanced to the point where they can treat a variety of system and environmental complexities including various scattering mechanisms, coherence, environmental range dependence, and bi-static geometries. In parallel with the development of the models has been an abiding interest in using the models to extract environmental information; most frequently scattering strength (e.g., Refs. 3 and 12–14) but also the reflection coefficient (or equivalently sediment geoacoustic properties), e.g., Ref. 15.

An important goal for both the modeling and inversion communities is a constrained comparison of the models with measurements, "constrained" meaning that the environmental variables are obtained independently. Such a comparison has not yet been made to the author's knowledge, and is the objective of this work.

The importance of such a comparison is that it is a fundamental step for both the modeling and environmental measurement communities to help demonstrate the validity of the theory and the experimental techniques. It is clearly also a foundational step for the development of inverse methods (such as rapid environmental assessment and through-the-sensor environmental measurement strategies) inasmuch as there must be some way to judge the "success" or "failure" of such methods. As an example of potential failures in scattering strength inversion is that large biases may occur when the incorrect scattering kernel is assumed.[16]

If one were attempting solely a model validation, a scaled tank experiment would be an attractive approach. However, in this case, since we wish to draw on environmental measurement techniques, an at-sea experiment is desirable. From the experimental side, our approach was to conduct the experiment in relatively simple conditions, i.e., few biologics, low sea state, and nearly range independent bathymetry and geoacoustic properties. This permits the testing of one aspect of the theory, bottom reverberation, which

a)Electronic mail: holland-cw@psu.edu

is generally viewed as the most important and least understood component of reverberation. From the modeling side, the energy flux approach was selected for the theoretical predictions. While there are no generally agreed upon "benchmark" solutions as yet for waveguide reverberation problems, the energy flux and normal mode approaches have been shown to be in agreement.[17]

The paper is organized in the following way. In Sec. II the theory of reverberation from a waveguide is reviewed. In conjunction with discussion of the theory, the approach for the constrained model-to-data comparison is explained in more detail. Following a review of the experiment site and its characteristics in Sec. III, the long-range reverberation measurements are described (Sec. IV). Sections V and VI describe the required environmental measurements, seabed scattering and reflection, respectively. In Sec. VII, a comparison is made between the measured data and theoretical predictions, where it is shown that the theory and data agrees well within the expected uncertainty bounds.

II. THEORY AND APPROACH

A. Reverberation in a waveguide

The reverberation in an isovelocity waveguide, assuming many propagating modes, can be written

$$I = I_o e^{-4\beta r} \frac{N}{H^2 r} \frac{c\tau}{2} \int_0^{\theta_c} \int_0^{\theta_c} \int_{\psi-\gamma/2}^{\psi+\gamma/2} M(\theta_i, \theta_o, \phi) \times R(\theta_i)^{r \tan\theta_i/H} R(\theta_o)^{r \tan\theta_o/H} d\theta_i d\theta_o d\phi, \quad (1)$$

where I_o is the source intensity; β is the attenuation in the seawater (see Appendix); r is range; N is the number of insonified patches (i.e., $N=2$ for an array of linear elements in a homogeneous environment except at end fire where $N=1$); ψ is the steering direction; γ is the horizontal beamwidth; c is the sound speed, τ is pulse length; H is water depth; θ_c is the critical angle; and M is the scattering kernel which depends upon incoming and outgoing vertical angles, θ_i, θ_o and azimuthal angle ϕ, and R is the plane wave pressure reflection coefficient.

When the intensity reflection coefficient is approximated as, $R^2 \sim \exp(-\alpha\theta)$ (sometimes called the Weston α, e.g., Ref. 18 which he employed for propagation modeling), then the reverberation[19] can be written (e.g., Refs. 4 and 7)

$$I = I_o e^{-4\beta r} \frac{N}{H^2 r} \frac{c\tau}{2} \int_0^{\theta_c} \int_0^{\theta_c} \int_{\psi-\gamma/2}^{\psi+\gamma/2} M(\theta_i, \theta_o, \phi) \times e^{-\alpha r \theta_i^2/2H} e^{-\alpha r \theta_o^2/2H} d\theta_i d\theta_o d\phi. \quad (2)$$

where the small angle approximation $\theta \sim \tan\theta$ has been made. From Eqs. (1) and (2) note that in order to predict seabed reverberation, both the seabed reflection coefficient and the scattering kernel must be known. In the following section we develop a parametrization for the scattering kernel M.

B. Parametrization of the scattering kernel

In this section we consider the dependence of the scattering kernel on parameters that can be extracted from the measurements. One could either use parameters based on scattering theories, or upon empirical relations. Ideally, it would be preferable to use a theoretical scattering kernel, however, it is more convenient here to use empirical scattering laws. This is because the empirical laws provide flexibility in treating arbitrary angular dependencies in the scattering kernel, whereas most of the current theories (e.g., perturbation theory) have a prescribed angular dependence below the critical angle (e.g., $\sin^4\theta$).

1. Generalized scattering kernel

Closed form expressions for the reverberation exist for several forms of empirical scattering kernels, and in particular forms that are separable in terms of the incident and scattered angle. For example, Zhou and Zhang[14] assume an isotropic scattering kernel that has symmetry in the incident and scattered angle dependence

$$M = \mu \sin^n\theta_i \sin^n\theta_o \quad (3)$$

and obtain a far-field, high frequency approximation for the reverberation, valid when

$$\xi = \frac{\alpha r \theta_c^2}{2H} > 1. \quad (4)$$

Harrison[7] finds closed form solutions valid at all ranges and frequencies for Lambert's law [$n=1$ in Eq. (3)] and for angle-independent scattering $M=\mu$. He also gives a far-field, high frequency approximation for the Lommel-Seeliger law

$$M = \mu_{\mathrm{LS}} \sin\theta_i \sin\theta_o / (\sin\theta_i + \sin\theta_o). \quad (5)$$

In practice, we would like to have a scattering law as general as possible. To this end, we can generalize the scattering laws to allow the form

$$M = \mu \sin^m\theta_i \sin^n\theta_o, \quad (6)$$

where m and n can vary independently[20] between 0 and 1. The reverberation can be written as a generalization of solutions in Ref. 7 as approximately (within a few tenths of a dB to the exact solution at all ranges):

$$I \simeq I_o e^{-4\beta r} \left(\frac{\pi}{2H}\right)^{2-p} \varepsilon_n \varepsilon_m \frac{N\mu}{\alpha^p r^{p+1}} \frac{\gamma c\tau}{2} \times (1 - e^{-\xi})^{2p-2} [\mathrm{erf}(\sqrt{\xi})]^{4-2p}, \quad (7)$$

where $p = 1 + n/2 + m/2$ and

$$\varepsilon_j^2 = \exp[(-1/2 + |j - 1/2|)/3] \quad (8)$$

for $j = n, m$.

2. Vertically bistatic versus monostatic angles

One interesting characteristic of the reverberation is that it turns out not to be sensitive to the details of the scattering law in terms of the dependence on the incident and scattered angles. This can be seen in Eq. (7) where all the terms depend upon the sum of the powers, except for $\varepsilon_{m,n}$ which is nearly unity for all m, n. So, for example, the scattering function $M = \mu \sin^{1/2}\theta_i \sin^{1/2}\theta_o$ gives almost exactly the same reverberation as $M = \mu \sin\theta_o$ except for a small (0.7 dB) dif-

ference due to ε. Another way to say this is that incoherent reverberation in a waveguide is dominated by backscattering; the vertically bistatic paths do not contribute significantly.

This is an important point for purposes of this paper, because in practice it is very difficult to directly measure the dependence of seabed scattering upon vertically bistatic angles. While measuring the vertically bistatic scattering strength is possible at relatively high angles in deep[21] and shallow water,[5] it is very difficult to measure the vertical bistatic angle dependence at low angles because of hybrid multipaths (see Figs. 2,3 in Ref. 5). Thus, Eq. (7) indicates that measuring the backscattering strength is sufficient for modeling reverberation and the errors resulting from not knowing the precise dependence on incident and scattered angle are small (less than 1 dB).

Although this conclusion is based on the assumption of a separable scattering kernel [in the form of Eq. (6)], it appears that at least for some cases, this conclusion is also valid for nonseparable scattering kernels. For example, the nonseparable scattering kernel in Eq. (5) yields reverberation nearly identical (within a dB or so) to other laws that have the same backscattering behavior, e.g., $M=\mu \sin^{1/2}\theta_i \sin^{1/2}\theta_o$.

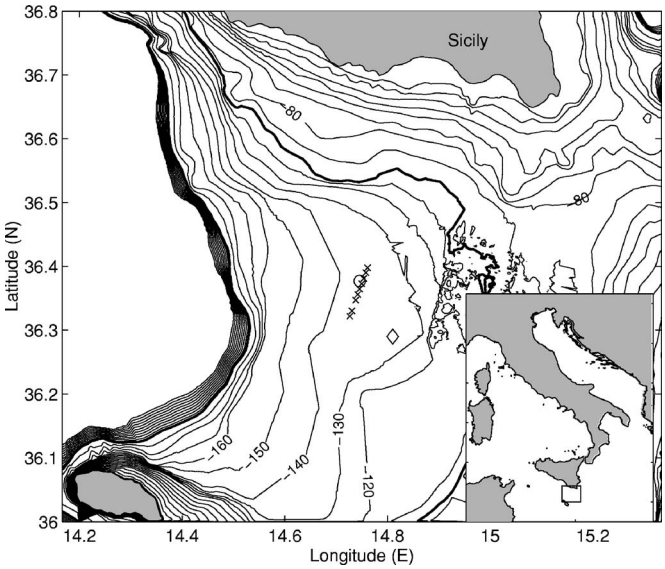


FIG. 1. Map of experimental area in the Straits of Sicily (Malta Plateau). Depths are in meters and were taken from historical soundings, except for a small section along the Ragusa Ridge between about 36.3 and 36.5° N. Shot locations are shown by x. The reflection (◦) and scattering (◇) experiment sites are also indicated.

C. Approach

In order to constrain the reverberation model-to-data comparisons, the seabed reflection coefficient and the scattering kernel must be obtained. Our strategy is based on local, or single seabed interaction, measurements. The general idea is to measure the seabed scattering strength within a given patch and measure the reflection coefficient along the source-to-patch-to-receiver path. Measured directional reverberation at a time and azimuth corresponding to the location of the scattering patch can then be compared with the model predictions using the measured scattering strength and reflection coefficient. In practice, we make the comparison for nine different measurements of reverberation, where the scattering patch is at various ranges and azimuths, in part to examine possible dependence of scattering strength on azimuth. Though the approximations leading to Eq. (7) are useful for understanding the reverberation, the model-to-data comparisons make no assumptions about the form of R. That is, Eq. (1) is employed in the model-to-data comparisons, where the integral is easily evaluated using standard numerical techniques (in this case Simpson's rule).

The seabed measurement techniques are briefly outlined here and then explained in detail in the following sections. The scattering strength is measured using a direct path technique[5] which yields backscattering strength as a function of angle and frequency. This technique averages over an area of the seabed roughly 600 m $\times$ 600 m. The location of the scattering strength measurement is 36.2893N 14.8151E (see Fig. 1). The measured scattering patch, with respect to the source-receiver location for the nine reverberation pings, varies from approximately 9–13 km and 110–160° in azimuth.

The seabed reflection coefficient is also measured using a local, direct path technique, described in Ref. 22 which effectively averages over an area of roughly 500 $\times$ 100 m (the dimensions depending on frequency, i.e., Fresnel zone width). The location of the measurement is shown in Fig. 1 and was selected to lie along the reverberation track. We make the assumption that the seabed reflection coefficient is constant along the track and from the track to the scattering patch. This assumption is based on a study of seabed variability,[23] that showed that this area is very nearly homogeneous with respect to the seabed properties ratio μ/α^2. While it is possible that this ratio is constant because changes in μ are always offset by changes in α, a simpler and more reasonable explanation is that the two quantities are nearly constant. Thus, we assume that the reflection coefficient varies slowly across that region (later we will show that this is a good assumption).

The theory [Eq. (1)] requires R and M over the angular range from 0° up to θ_c. In practice it is difficult to measure very low grazing angles from direct path measurements. Thus, the data are extrapolated in angle using models. For the reflection coefficient, a geoacoustic model derived from the reflection measurements is employed to extrapolate the measurements from the lowest angle of observation, ~8.5°, down to 0°. For the scattering strength, an empirical model is used to extrapolate the observations from the lowest angle of observation, 9° at 900 Hz and 5° at 1800 Hz, down to 0°. An important aspect of the extrapolation is that only the observations of R and M are used for the extrapolation (and not the reverberation).

III. DESCRIPTION OF ENVIRONMENT

The Malta Plateau (see Fig. 1) in the Straits of Sicily occupies the northern edge of the North African passive continental margin and is a submerged section of the Hyblean Plateau of mainland Sicily. A discussion of the geology of this area, along with seismic reflection data and core data can be found in Refs. 24 and 25. The region is divided by the Ragusa Ridge, which forms a spine ~15 km wide between

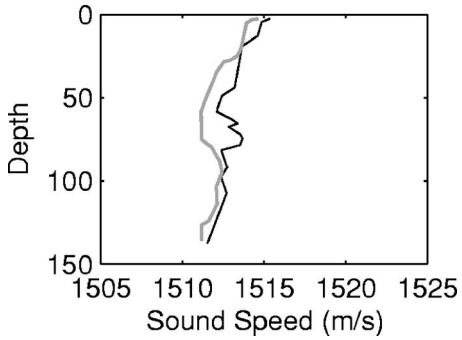


FIG. 2. XBT measurements at the beginning (black) and end (gray) of the reverberation track (see Fig. 1). Source depth is nominally 91 m.

Sicily and Malta. The ridge has exposed rock (presumably limestone) on its eastern and western edges. West of the ridge the seabed is blanketed with silty-clay sediment layer that is roughly 0.25 m thick at water depths of 130–150 m. Below the silty clay, there is a thick (several hundred meters) sequence of unconsolidated sediments.

Sound speed profiles were collected using expendable bathythermographs (XBTs, see Fig. 2) at the beginning and end of the track. The profiles are slightly downward refracting with a gradient of $\sim 0.02\ s^{-1}$. Model predictions show that the effect of this gradient on reverberation is negligible. Wind speed was generally less than 6 knots and the mean rms wave height, measured with a waverider buoy, was 0.16 m with a standard deviation of 0.007 m. We assume that under these calm conditions, the sea surface can be treated as a smooth pressure release interface. Measurements with a normal incidence Simrad EY-500 sonar indicated that there were very few fish schools in the area during the measurement evolution.

IV. REVERBERATION MEASUREMENTS

Broadband directional reverberation data were collected in April 14 1998 during the SCARAB98 experiment (see "x" in Fig. 1). Impulsive sources were employed (Mk61 Sound Undersea Signal, SUS) at 91 m depth approximately every 5 min along each track (see Fig. 1). The SUS were launched from the vessel and initiated within a few hundred meters of the receiver. In the modeling, the geometry is assumed to be monostatic, which is a reasonable approximation given that the ranges of interest are greater than 8 km. Source levels were obtained from empirical equations.[26]

The receiver was a three-aperture nested horizontal array consisting of 256 elements with element spacing of 0.5, 1 and 2 m towed at a depth of $\sim$50 m. Data from the 0.5 m aperture are used in this analysis. The data were digitized at 6000 Hz sampling rate and the data were time-domain beamformed with Hanning shading. The data were incoherently averaged in frequency in 100 Hz bands from 100 to 1800 Hz.[27] In the processing, an integration time of 0.25 s was employed, which corresponds to a radial patch dimension of about 200 m. The range to a particular scattering patch was estimated by $r = c_{av}t/2$ where c_{av} was computed from the measured sound speed profile.

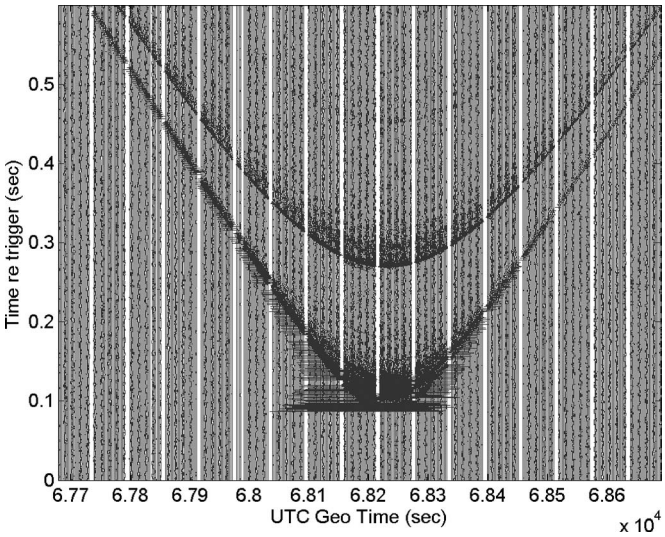


FIG. 3. Raw time series from the seabed reflection experiment. The first group of arrivals (minimum arrival time of 0.1 s) corresponds to the direct and single bottom bounce paths. The next group of arrivals (minimum arrival time of 0.28 s) corresponds to paths with one surface reflection.

For the measurement geometry there is slight range dependence in the bathymetry (3–4 m over 10 km) which is taken into account, however, the predicted reverberation is nearly identical (within a few tenths of a dB) to a range independent environment.

V. REFLECTION COEFFICIENT MEASUREMENT

The spherical wave reflection coefficient R_s was measured using a direct path (single-interaction) technique with a towed source and a single fixed omni-directional hydrophone. The source was marine seismic boomer (GeoAcoustics Uniboomer Model 5913B) which has a highly repeatable broadband pulse, pulsed once/s. The data have been decimated by a factor of 2 for the analysis here.

The receiver was a single sonobuoy 11 m above the seabed. The data were sampled at 48 kHz in the buoy and telemetered to the R/V Alliance. Both the source trigger and the data acquisition were driven by the same GPS clock to eliminate synchronization problems and to allow geometry reconstruction using time of flight. Details on the measurement and processing techniques can be found in Refs. 22 and 28.

The location of the reflection measurement is shown in Fig. 1. The raw reflection time series data are shown in Fig. 3. The first group of arrivals (minimum arrival time of 0.1 s) corresponds to the direct and single bottom bounce paths. The direct path data are processed to obtain source calibration as a function of angle and frequency. The first bottom bounce paths arrivals are integrated and the ratio (with geometry and transmission corrections) is the bottom reflection coefficient (see Ref. 28 for details in the processing). The measured reflection loss ($BL = -20 \log |R_s|$) is shown in Fig. 4.

In order to obtain an estimate of the geoacoustic properties, a time domain method[22] was employed that yields thickness and interval velocity in each layer that is temporally resolved by the pulse. Studies with synthetic data have shown that the velocity and thickness estimates obtained by

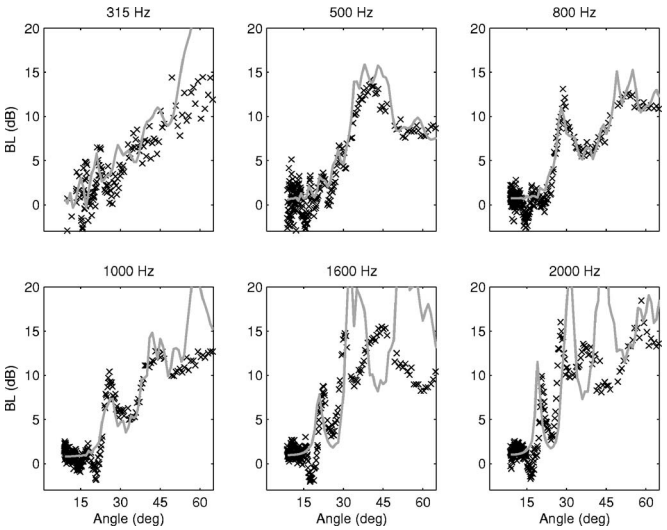


FIG. 4. Seabed reflection loss measurements (x) on the Malta Plateau (see Fig. 1 for location) processed in 1/3 octave bands. Model predictions (—) are also shown for the geoacoustic model in Table I.

fitting hyperbolas to the arrivals are typically within a few percent of the correct values. The data and the hyperbolic fits are shown in Fig. 5 for the upper three and upper six sediment layers where the data are plotted in reduced time (see Ref. 22) which essentially removes the hyperbolic dependence on range. The estimated interval velocities for each layer are given in Fig. 5(b) and Table I. In addition to layer thickness and velocity, density and compressional wave attenuation estimates are required. Density was obtained using empirical velocity-density relations.[29] Gravity core data indicated that the fine-grained mud layer, with a velocity of 1480 m/s, was about 25 cm layer at this site. This layer was included in the geoacoustic model (see Table I). The attenuation in the layers below the fine-grained mud layer was obtained by fitting the predicted reflection coefficient (below 15°) with the measured data. A single attenuation value was employed for all layers. A least-squares procedure was used and the sum of the variances over the three frequencies as a function of attenuation is shown in Fig. 6, where a clearly defined minimum is evident at 0.165 dB/m/kHz.

The measured data are spherical wave reflection coefficients, defined as the ratio of the reflected to the incident pressure from a point source

$$R_s(\theta) = \frac{D}{e^{ikD}} ik \int_0^{\frac{\pi}{2}-i\infty} J_o(kr\cos\theta) R(\theta) e^{-k(z+z_o)\sin\theta} \cos\theta d\theta, \quad (9)$$

where D is the path length along the seabed reflecting path, and k is the wave number. The plane wave reflection coefficient R was calculated from the geoacoustic model of Table I, and the resulting predicted spherical wave reflection coefficient is compared with the measurements in Fig. 4. The most important angles for long-range reverberation are at low angles, i.e., below the critical angle. The reflection coefficient model-to-data agreement is reasonably good. We use the geoacoustic model to generate the plane-wave reflection coefficient, Fig. 7, which will be needed later on (Sec. VII) for the reverberation model-to-data comparison.

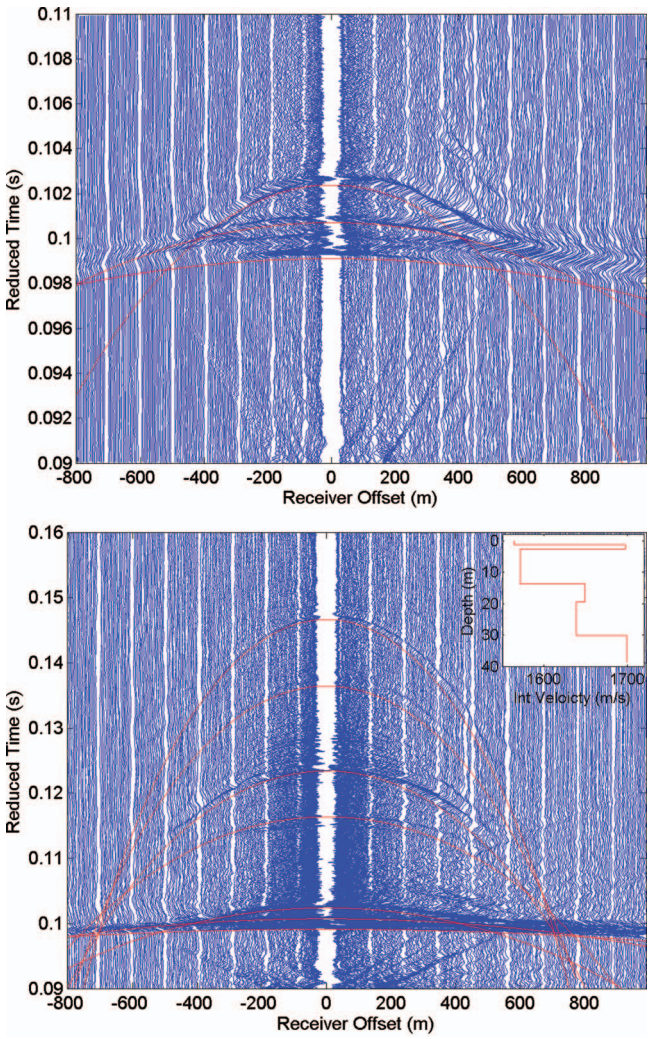


FIG. 5. Reflection time series (blue) and hyperbolic fits (red) for: (a) upper three layers, (b) upper six layers. The reducing velocity was 1512 m/s.

VI. SCATTERING STRENGTH MEASUREMENT

The scattering kernel M is determined here by direct path measurement. One of the challenges of measuring the direct path scattering strength in shallow water is the presence of multipaths. In order to control multipaths, vertical directionality is employed in both source and receiver.

The sources are spaced at $\lambda/2$ and driven in phase to produce a broad beam in the horizontal direction and nulls in

TABLE I. Geoacoustic model derived from the broadband reflection coefficient data. For the frequencies of greatest interest here, 900–1800 Hz, only the upper 3 m play a significant role in seabed reflection.

Thickness (m)	Sound speed (m/s)	Attenuation (dB/m/kHz)	Density (g/cm^3)
⋯	1512	0	1.03
0.25	1480	0.01	1.39
1.26	1564	0.165	1.65
1.4	1698	0.165	1.96
11	1571	0.165	1.67
5.8	1649	0.165	1.86
10.7	1639	0.165	1.84
⋯	1699	0.165	1.96

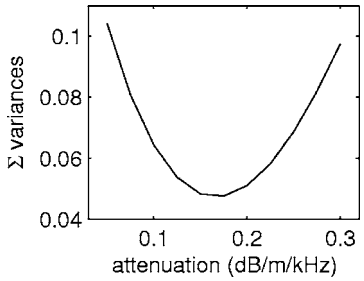


FIG. 6. Statistics of the fit between attenuation and the measured intensity reflection coefficient at the frequencies closest to the measured scattering strength (1000, 1600 and 2000 Hz). The sum of the variances is over frequency. There is a clear well-defined minimum at 0.165 dB/m/kHz using the velocities, densities and layer thicknesses of Table I.

the vertical. This helps reduce or eliminate so-called "fathometer" returns or returns from bottom or subbottom reflections. Two source pairs are used, Mod30s at 900 Hz and ITC 4001 at 1800 Hz. The receiver was a vertical array of 32 Benthos AQ-4 hydrophones with 0.18 m spacing; the data are sampled at 12 kHz. The vertical receive array and sources were deployed with the center of the receive array 34 m above the seabed and the vessel drifted at speeds of less than 0.5 m/s. Beam forming permits separation of seabed scattering from volume and/or sea surface scattering (within the spatial/temporal resolution of the experiment). A more detailed description of the experiment design, the data processing, and source calibration can be found in Ref. 5.

The reverberation theory requires the scattering kernel down to 0°, so the measurements must be extrapolated using a model; in this case an empirical is employed of the form of Eq. (6) with $m=n$. An important consideration when attempting to determine the scattering law for long-range reverberation is that only backscattering data at angles that contribute to the reverberation should be considered. This is important because the scattering mechanism (and hence the angular dependence) can change with grazing angle. For example, Ref. 30 showed a considerably different angular dependence above and below 20° due a change in the scattering mechanism (in that case, the low angles were governed by volume scattering and the higher angles by basement scattering). In order to quantify what angles are important, we can show [using Eq. (1)] that the distribution of intensity in the vertical is proportional to

$$I(\theta_o) \propto \sin^{n+m}\theta_o R(\theta_o)^{2r\theta_o/H}. \tag{10}$$

The predicted intensity distribution in angle for $n=m=1$ and $n=m=1/2$ is shown in Fig. 8. The difference in angular behavior for the two frequencies is due to the frequency dependence of the reflection coefficient (see Fig. 7). In order to not unduly limit the angles that might be important in the reverberation, the angular range for fitting the scattering data to model was ±16° at 900 Hz and ±13° at 1800 Hz corresponding to the 6 dB down points for the widest distribution ($n=1$).

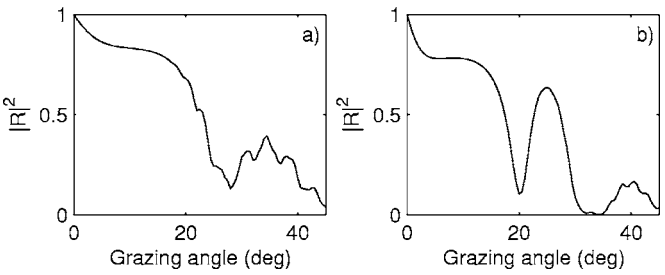


FIG. 7. Plane wave intensity reflection coefficient from the geoacoustic model (Table I) at (a) 900 and (b) 1800 Hz.

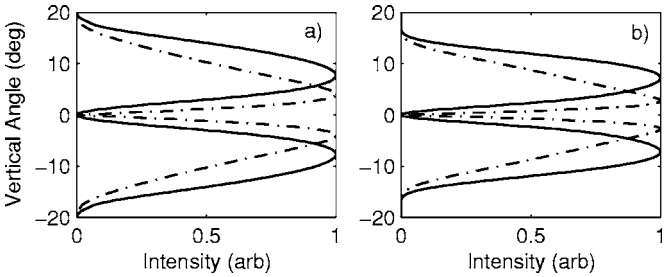


FIG. 8. Vertical angle distribution at 9 km, 135 m water depth, at (a) 900 Hz and (b) 1800 Hz. The solid and dashed-dot lines represent a scattered angle dependence of $n=1$ and $n=0.5$, respectively.

Backscattering measurements at the location in Fig. 1, averaged over 15 pings, are shown in Fig. 9. The data, averaged over 1°, were fit to a model $M(\theta_i=\theta_o)=\mu \sin^q\theta$, where $q=m+n$ and μ were obtained by a least-squares fit to the data over the angular range above ($\pm 16°, \pm 13°$). The resulting values are $q=0.8$ at 1800 Hz and $q=1.2$ at 900 Hz. However, due to the oscillations in the backscattering data, particularly a 900 Hz, these results were somewhat dependent on precisely which upper angle limit was used. For example, at an upper limit of ±14°, $q=0.92$ at 900 Hz. In general, what the fitting showed was that at low angles (of importance to reverberation, i.e., less than ~16°), the value of $q \sim 1$ for both frequencies. The resulting values of the scattering strength at 900 and 1800 Hz are $10\log_{10}\mu=-45$ dB and −33 dB, respectively, and the resulting scattering kernels are plotted in Fig. 9.

Given the backscattering kernel derived from the measurements of $M(\theta_i=\theta_o)=\mu \sin\theta$, we must decide on how to represent the bistatic scattering kernel, which could be $M=\mu \sin^{1/2}\theta_i \sin^{1/2}\theta_o$ or $M=\mu \sin\theta_o$ or a nonseparable kernel like the Lommel-Seeliger law [with $\mu=2\mu_{LS}$, Eq. (5)]. Any of these are possible and there is no convenient way to distinguish between them, or other candidate laws where $n+m=1$. However, it was shown in Sec. II B 2 that the reverberation is not very sensitive to n and m, but only their sum which we know is ~1. For example, the reverberation with $M=\mu \sin^{1/2}\theta_i \sin^{1/2}\theta_o$ is less than 1 dB higher than the lowest reverberation for the candidate models, and less than 1 dB lower than the highest reverberation for the candidate models. The highest candidate model is $M=\mu \sin\theta_o$ and the lowest is the Lommel-Seeliger law. Thus, accepting a minor uncertainty of ±1 dB, the scattering kernel $M=\mu \sin^{1/2}\theta_i \sin^{1/2}\theta_o$ is employed in the model-to-data comparison.

It is interesting to note that in Ref. 31, 800–1400 Hz reverberation simultaneously received on a vertical and horizontal array in this same geographic area was employed to try to distinguish between three scattering kernels, $m=n=0$, $m=n=1/2$ and $m=n=1$. While it was not possible in that

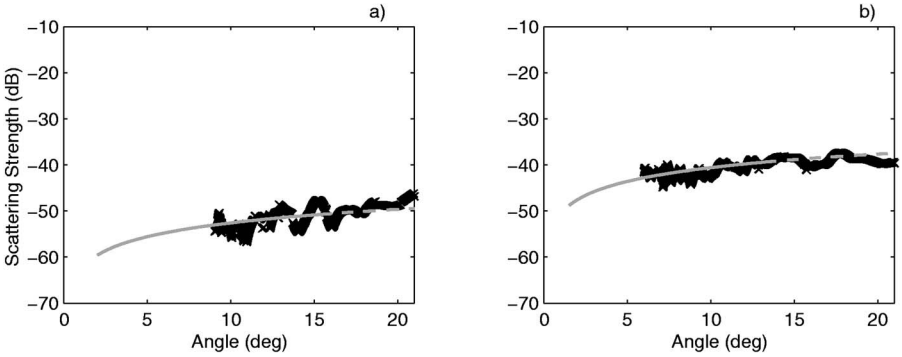


FIG. 9. Measured backscattering scattering strength (x) with model (solid gray line) obtained from least-squares fit to data at: (a) 900 Hz and (b) 1800 Hz over the predicted angle limits (6 dB down points) of the reverberation. The gray dashed line shows the fit extrapolated to higher angles.

study to definitively distinguish between these scattering kernels, some evidence was found to support the $m=n=1/2$ case or alternatively a hybrid Lambert/angle-independent scattering law.

VII. COMPARISON OF THEORY WITH OBSERVATIONS

In this section we compare the measured reverberation with the theoretical reverberation [Eq. 1] using the measured/extrapolated values of the reflection coefficient and scattering strength. Measured reverberation data at 1800 and 900 Hz are shown in Figs. 10 and 11 for each of the nine pings of Fig. 1. The time associated with the scattering patch is shown in the vertical dashed black line. The large arrival at about 20 s in the first few pings is associated with the Ragusa Ridge (Fig. 1).

Also shown in Fig. 10 are the theoretical predictions (dotted red line) based on the measured reflection coefficient and scattering kernel. The red solid line includes noise, which was taken from a 5 s average before the ping. It is useful to see the reverberation both with and without noise in order to ensure that the reverberation-to-noise level is sufficiently high that the comparison is meaningful. Note that the reverberation-to-noise level is high on all pings and the model predictions agree very closely with the measurements. The key point to note is that the agreement is good where the beam from the long-range reverberation data intersects with the measured scattering patch (indicated by vertical dotted line). The fact that the model to data agreement is good at other ranges means that the scattering strength and the reflection coefficient are nearly constant over this area.[32] Also shown in Fig. 10 are the predictions with the commonly used Lamberts law, [Eq. (6)] with $m=n=1$] where μ was fit to the scattering data of Fig. 9. Note that Lambert's law is not correct, and is about 5–6 dB too low.

Theoretical predictions at 900 Hz based on measured reflection and scattering $m=n=1/2$ also agree well with the observations (see Fig. 11). It is useful to examine the model-to-data comparison for the measured scattering kernel ping by ping to determine if there is any systematic dependence upon azimuth or range, i.e., determine if the scattering strength has any detectable azimuth dependence and/or the reflection coefficient has any detectable range/azimuth dependence. Figure 12(b) shows the model-to-data comparison at 900 and 1800 Hz for each ping. For each ping the predicted reverberation is subtracted from that measured, where the measured reverberation has been averaged over a patch size roughly comparable to that measured in the scattering strength experiment. At 900 Hz, the measured reverberation

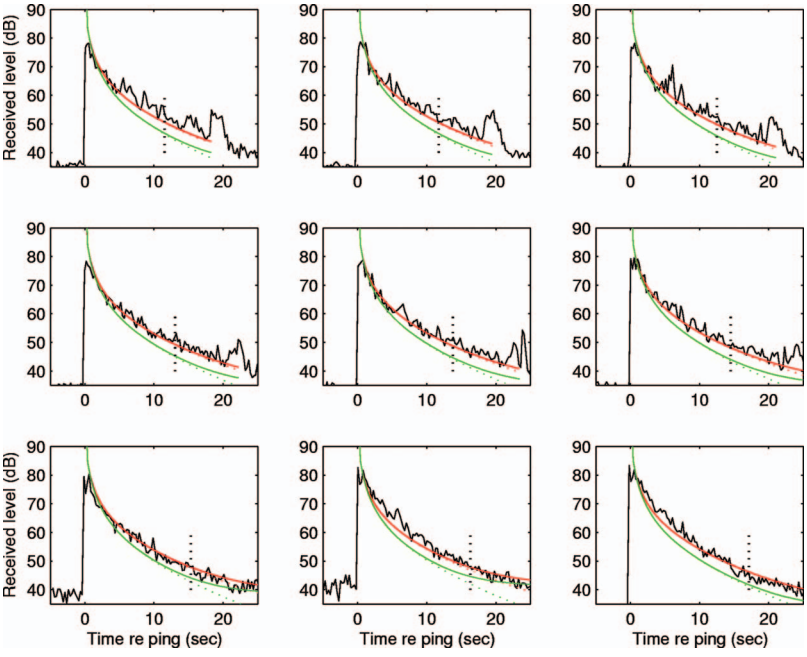


FIG. 10. Comparison of measured reverberation (black line) at 1800 Hz vs theoretical predictions using measured reflection coefficient and backscatter data. The nine panels represent the nine pings shown in Fig. 1 from south to north. The dotted red line is for the measured scattering kernel [Eq. (6) $m=n=1/2$], the dotted green line is for Lambert's law [Eq. (6) $m=n=1$]. The solid lines include the effect of noise and aid in determining where the data and model are dominated by noise. The location in time corresponding to insonification of the measured scattering site is shown by the vertical black dotted line.

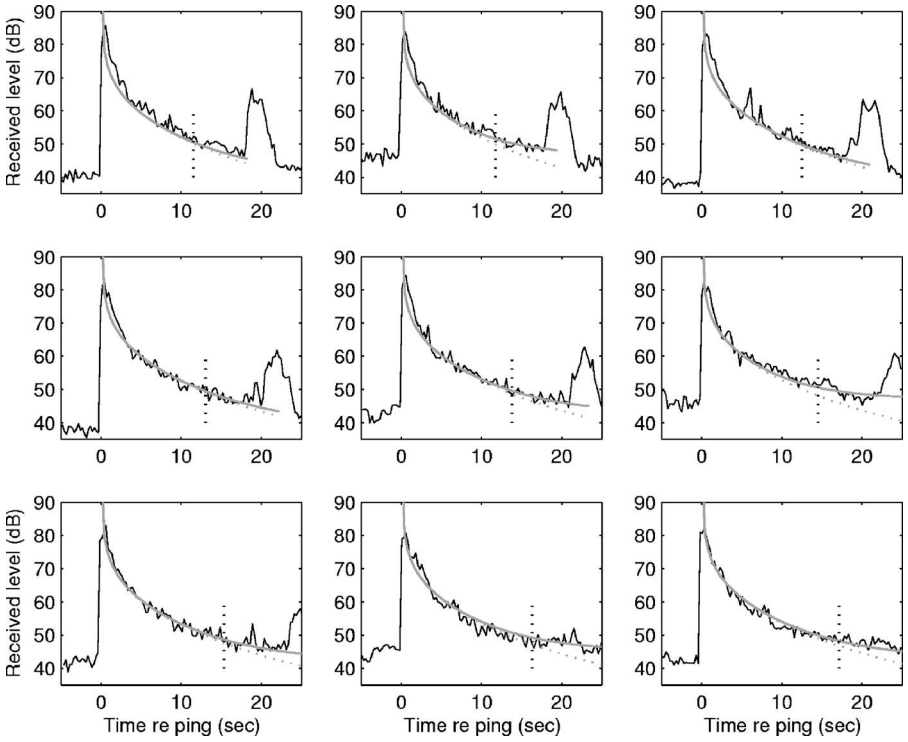


FIG. 11. Comparison of measured reverberation (black line) at 900 Hz vs theoretical predictions using measured reflection coefficient and backscatter data. See Fig. 10 for explanation, the gray line is for the measured scattering kernel [Eq. (6) $m=n=1/2$].

patch sizes are roughly 500×600 m and at 1800 Hz, 250×600 m. The main point in this plot is that the theory and measurement agreement is good at both frequencies. Also note that at 900 Hz the randomness of the model-data differences indicates that the scattering strength/reflection coefficient has no significant azimuthal/range dependence. At 1800 Hz, there is a possible trend between 140 and 160° azimuth, but the angular region is too small to make a definitive conclusion.

As a conservative estimate of uncertainties, we expect that the measured reverberation, scattering strength and reflection data all have roughly a 2 dB uncertainty (from source level and calibration uncertainties), the precise scattering law (unknown m, n) has another 1 dB. If it is assumed that the uncertainties are uncorrelated, we expect an overall ±3 dB uncertainty. The residual error generally is less than the expected uncertainty [see Fig. 12(b)].

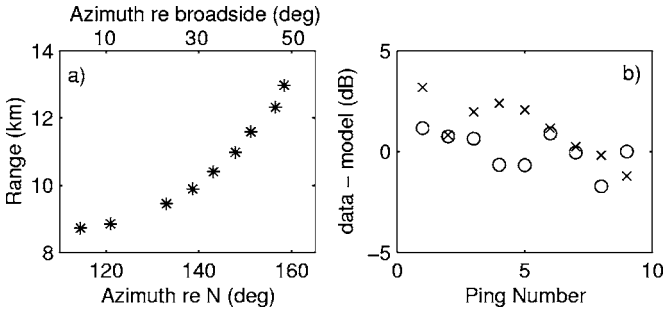


FIG. 12. Reverberation model to data comparison: (a) geometry, i.e., range and azimuth for each ping and (b) model-to-data differences (data minus model) at 900 Hz (o) and 1800 Hz (x) at the scattering measurement location.

VIII. DISCUSSION

A. Constrained comparison

In an "ideal" (but perhaps not experimentally viable) reverberation data-to-model comparison, the reflection coefficient and scattering kernel would be measured from 0° to the critical angle. In the case here, the reflection coefficient and scattering kernel are measured over part of that angular region and models are used to fit the data and extrapolate them across the entire angular range.

For the case of the reflection coefficient, R, a physics-based approach using a geoacoustic model was employed to extrapolate down to lower angles. One parameter in the geoacoustic model (attenuation) was obtained by least-squares fitting to the observed R. For the scattering kernel, an empirical model was used and parameters q and μ were obtained by a least-squares fit to the scattering data. Thus in both the reflection coefficient and the scattering kernel model there are free parameters that are obtained from the respective measurements. However, the key point is that neither model employed reverberation data as part of the parameter estimation. Thus, we term the comparison of predicted and observed reverberation constrained in the sense that none of the reverberation model inputs (reflection coefficient or scattering kernel parameters) were obtained using the reverberation data. It is also in this sense that we say that there were no "free" parameters, it means that in the reverberation model-to-data comparisons the inputs were fully determined independently of the reverberation data.

B. Physics-based versus empirical models

The choice of whether to use a physics-based model or an empirical model depends in part on the amount of infor-

mation available about the required inputs. For example, a physics-based model could be used for M, but it is not clear if fitting the (generally unknown) parameters controlling volume scattering and roughness would lead to a more satisfactory fit (or any new information) than contained in the empirical model used here.

Generally, however, a physics-based model may be preferable, since the extrapolation in angle may be more robust than for an empirical model. Another advantage of a physics-based scattering model (that treats a multilayered seabed), is the potential for a self-consistent description of the environment. In other words, the scattering model includes the same underlying physics (i.e., sediment layer reflectivity, absorption) as the reflection model. We see this as an important future goal of reverberation measurement and modeling studies.

IX. SUMMARY AND CONCLUSIONS

Comparisons are made of theoretical predictions of monostatic waveguide reverberation with measured directional reverberation in the Straits of Sicily. The theory requires knowledge of the environmental variables including the sound speed profile, water depth, seabed scattering strength and seabed reflection coefficient which are all obtained independently. The theoretical predictions show good agreement with the measured data (see Figs. 10–12), i.e., within the uncertainty bounds.

This result is important because it is a fundamental step for both the modeling and environmental measurement communities to help demonstrate the validity of the theory and the experimental techniques. It is clearly also a foundational step for the development of inverse methods (such as rapid environmental assessment and through-the-sensor environmental mapping strategies) inasmuch as it opens the door for a method to measure the "success" or "failure" of such methods.

Though ostensibly the reverberation model requires knowledge of the vertical bistatic angle dependence of scattering strength, it was shown that knowledge of the backscattering strength is sufficient to characterize the scattering kernel within relatively small error (about ±1 dB). In this region, the backscattering strength follows a $M(\theta)=\mu \sin\theta$ and not Lambert's law and is a strong function of frequency, −45 and −33 dB at 900 and 1800 Hz, respectively.

ACKNOWLEDGMENTS

The author gratefully acknowledges the NATO Undersea Research Centre (under whose auspices the experiments were conducted) and the Office of Naval Research Ocean Acoustics Program, Code OA321 whose support made this analysis possible.

APPENDIX: SEAWATER ATTENUATION IN THE ENERGY FLUX MODEL

The effect of attenuation in seawater is generally treated as an exponential outside the integral, as in the first term of Eq. (1), (for example, see Ref. 14). However, it is clear that this is an approximation, inasmuch as individual rays have different amounts of attenuation depending upon their angle. Here we provide the "exact" solution for attenuation in the water column. Assuming that the reflection coefficient can be represented by $R^2 \sim \exp(-\alpha\theta)$, the intensity decay as a function of range can be written as:

$$e^{-(\alpha\theta+2\beta 2H/\sin\theta)r/r_c}, \tag{A1}$$

where the factor of 2 in front of β comes from the fact that it is in units of pressure (nepers/m), not intensity, and r_c is the cycle distance

$$r_c = 2H/\tan\theta. \tag{A2}$$

Using the small angle approximation, $\sec\theta = 1+\theta^2/2$, Eq. (2) can be written

$$I = I_o \exp(-4\beta r)\frac{N}{H^2 r}\frac{c\tau}{2}\int_0^{\theta_c}\int_0^{\theta_c}\int_{\psi-\gamma/2}^{\psi+\gamma/2} M(\theta_i,\theta_o,\phi) \times \exp(-\bar{\alpha} r\theta_i^2/2H)\exp(-\bar{\alpha} r\theta_o^2/2H)\,d\theta_i\,d\theta_o\,d\phi, \tag{A3}$$

where

$$\bar{\alpha} = \alpha + 2\beta H. \tag{A4}$$

In other words, the effect of the seawater attenuation can be expressed as two factors: one that depends only on horizontal range, r, and a perturbation to the reflection slope α. Thus, the results in Eq. (7) are correct (as are results for transmission loss and reverberation in Refs. 4 and 7) simply by replacing α with $\bar{\alpha}$. At the frequencies, water depths, and sediment characteristics of interest in this study, α is of order 10^0 and $2\beta H$ is of order 10^{-2}, so that to very good approximation, $\bar{\alpha} \sim \alpha$. Note that Eq. (1) uses this same approximation, i.e., $\alpha \gg 2\beta H$.

[1] C. R. Eyring, R. J. Christensen, and R. W. Raitt, "Reverberation in the sea," J. Acoust. Soc. Am. **20**, 462–475 (1948).
[2] R. J. Urick, "Side scattering in shallow water," J. Acoust. Soc. Am. **32**, 351–355 (1960).
[3] K. V. Mackenzie, "Bottom reverberation for 530- and 1030 cps sound in deep water," J. Acoust. Soc. Am. **33**, 1498–1504 (1961).
[4] J.-X. Zhou, D. Guan, E. Shang, and E. Luo, "Long-range reverberation and bottom scattering strength in shallow water," Chin. J. Acoust. **1**, 54–63 (1982).
[5] C. W. Holland, R. Hollett, and L. Troiano, "A measurement technique for bottom scattering in shallow water," J. Acoust. Soc. Am. **108**, 997–1011 (2000).
[6] K. V. Mackenzie, "Long-range shallow water reverberation," J. Acoust. Soc. Am. **34**, 62–66 (1962).
[7] C. H. Harrison, "Closed-form expressions for ocean reverberation and signal excess with mode stripping and Lambert's law," J. Acoust. Soc. Am. **114**, 2744–2756 (2003).
[8] H. P. Bucker and H. E. Morris, "Normal-mode reverberation in channels or ducts," J. Acoust. Soc. Am. **44**, 827–828 (1968).
[9] D. Ellis, "A shallow-water normal-mode reverberation model," J. Acoust. Soc. Am. **97**, 2804–2814 (1995).
[10] H. Weinberg, "Generic sonar model," Naval Underwater Systems Center, New London, CT, TD 5971D (1985).
[11] F. D. Tappert, Physics of the PE Reverb Model, ONR-ARSRP Symposium, La Jolla, CA, 23–25 March (1993).
[12] R. J. Urick, "Reverberation-derived scattering strength of the shallow sea bed," J. Acoust. Soc. Am. **48**, 392–397 (1970).
[13] P. G. Cable, K. D. Frech, J. C. O'Connor, and J. M. Steele, "Reverberation-derived shallow-water bottom scattering strength," IEEE J. Ocean. Eng. **22**, 534–540 (1997).
[14] J. Zhou and X. Z. Zhang, "Shallow-water reverberation and small-angle bottom scattering," *International Conference on Shallow-water Acoustics*,

edited by X. Z. Zhang and J.-X. Zhou, No. 315–322, Beijing, China, 21–25 April 1997.
[15]J. R. Preston, D. D. Ellis, and R. C. Gauss, "Geoacoustic parameter extraction using reverberation data from the 2000 Boundary Characterization Experiment on the Malta Plateau," IEEE J. Ocean. Eng. **30**, 709–732 (2005).
[16]C. W. Holland, "On errors in estimating bottom scattering strength from acoustic reverberation," J. Acoust. Soc. Am. **118**, 2787–2790 (2005).
[17]K. D. LePage and C. H. Harrison, "Bistatic reverberation benchmarking exercise: BiStaR versus analytic formulas," J. Acoust. Soc. Am. **113**, 2333–2334 (2003).
[18]D. E. Weston, "Intensity-range relations in Oceanographic Acoustics," J. Sound Vib. **18**, 271–287 (1971).
[19]Note that in Ref. 4 α is defined in terms of the pressure reflection coefficient and thus is a factor of 2 smaller than the definition used here.
[20]While these kinds of empirical laws are frequently seen in the literature, they are not physically possible since they do not obey reciprocity. For purposes of the discussion, they are useful though, since it helps show the dependence of reverberation on the vertical bistatic angle.
[21]P. C. Hines, D. V. Crowe, and D. D. Ellis, "Extracting in-plane bistatic scattering information from a monostatic experiment," J. Acoust. Soc. Am. **104**, 758–768 (1998).
[22]C. W. Holland and J. Osler, "High resolution geoacoustic inversion in shallow water: A joint time and frequency domain technique," J. Acoust. Soc. Am. **107**, 1263–1279 (2000).
[23]C. W. Holland, "Mapping seabed variability: Rapid surveying of coastal regions," J. Acoust. Soc. Am. **119**, 1373–1387 (2006).
[24]M. D. Max, A. Kristensen, and E. Michelozzi, "Small scale Plio-Quaternary sequence stratigraphy and shallow geology of the west-central Malta Plateau," SACLANT Centre Report No. SR-209 (1993).
[25]J. Osler and O. Algan, "A high resolution seismic sequence analysis of the Malta Plateau, NATO Undersea Research Centre Report No. SR-311 (1999).
[26]N. R. Chapman, "Source levels of shallow explosive charges," J. Acoust. Soc. Am. **84**, 697–702 (1988).
[27]At 1800 Hz the first grating lobe is at 56° relative to broadside; the data show a systematic increase in beam levels of about 3 dB from about 40–55°. This correction was included at 1800 Hz for the three (out of nine) beams in that angular range.
[28]C. W. Holland, "Seabed reflection measurement uncertainty," J. Acoust. Soc. Am. **114**, 1861–1873 (2003).
[29]R. Bachman, "Acoustic and physical property relationships in marine sediment," J. Acoust. Soc. Am. **78**, 616–621 (1985).
[30]C. W. Holland and P. Neumann, "Sub-bottom scattering: A modeling approach," J. Acoust. Soc. Am. **104**, 1363–1373 (1998).
[31]M. K. Prior, "Experimental investigation of the angular dependence of low-frequency seabed reverberation," IEEE J. Ocean. Eng. **30**, 691–699 (2005).
[32]Strictly speaking, the good agreement at ranges other than the scattering patch (dotted line) means that the ratio μ/α is constant, but the simplest explanation is that μ and α are each constant.

Validation of statistical estimation of transmission loss in the presence of geoacoustic inversion uncertainty

Chen-Fen Huang, Peter Gerstoft, and William S. Hodgkiss
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

(Received 21 March 2006; revised 10 July 2006; accepted 10 July 2006)

Often the ocean acoustic environment is not well known and sonar performance prediction will be affected by this uncertainty. Here, a method for estimating transmission loss (TL) is proposed which incorporates these environmental uncertainties. Specifically, we derive an approach for the statistical estimation of TL based on the posterior probability density of environmental parameters obtained from the geoacoustic inversion process. First, a Markov chain Monte Carlo procedure is employed in the inversion process to sample the posterior probability density of the geoacoustic parameters. Then, these sampled parameters are mapped to the transmission loss domain where a full multidimensional probability distribution of TL as a function of range and depth is obtained. In addition, TL is also characterized by its summary statistics including the median, percentiles, and correlation coefficients. The approach is illustrated using a data set obtained from the ASIAEX 2001 East China Sea experiment. Based on the geoacoustic inversion results, the predicted TL and its variability are estimated and then compared with the measured TL. In general, there is a good agreement with the percentage of observed number of data points inside the credibility interval.
 [DOI: 10.1121/1.2261356]

PACS number(s): 43.30.Pc, 43.60.Pt [AIT] Pages: 1932–1941

I. INTRODUCTION

Statistical estimation of geoacoustic parameters from acoustic field data has been an active research topic for more than a decade.[1–7] This paper proposes to use the parameter uncertainties obtained during the geoacoustic inversion process to make a statistical estimation of transmission loss (TL). The transmission loss domain is important as it can be used in connection with sonar performance prediction (e.g., Ref. 8 and in particular the paper by Abbot and Dyer[9]).

Analytical approaches to transfer uncertainties have been adopted by several authors. Reference 10 derived an analytical expression for quantifying the uncertainty in predicted acoustic fields produced by environmental uncertainties. Reference 11 describes how uncertainty can be embedded into ocean acoustic propagation models through expansions of the input parameter uncertainties in orthogonal polynomials. The disadvantage of these approaches is that they are less flexible computationally and so far have only been used on simple problems.

Monte Carlo methods for sampling the environmental variability have been studied by several authors (e.g., Refs. 12 and 13). Random realizations of the acoustic environment are propagated via the deterministic wave equation to produce realizations of the acoustic field. Mean and higher moments are used to characterize the acoustic variability. While thousands of simulations generally are required, these computations are fast, simple, and thus not seen as a problem.

A Markov chain Monte Carlo (MCMC) method is used to first sample the probability distributions of the geoacoustic parameters. Unlike previous research, the results of the geoacoustic inversion are only an intermediate goal. Subsampling of the multidimensional model parameter distribution is then used to map parameter uncertainties to the TL domain. In Ref. 14, exhaustive grid sampling was used to obtain the geoacoustic uncertainties and map these to the TL domain. This was feasible because only 4 model parameters were explored. However, a more realistic inversion will have a large number of parameters. In this paper we invert for a total of 13 model parameters and also validate the estimated transmission loss with at-sea observations.

Figure 1 summarizes the estimation of TL (usage domain $\mathcal{U}$) from ocean acoustic data observed on a vertical or horizontal array (data domain $\mathcal{D}$).[14] The geoacoustic inverse problem is solved as an intermediate step to obtain the posterior distribution of environmental parameters $p(\mathbf{m}|\mathbf{d})$ (environmental domain $\mathcal{M}$). We are not directly interested in the environment itself but rather a statistical estimation of the TL field (usage domain $\mathcal{U}$). Based on the posterior distribution $p(\mathbf{m}|\mathbf{d})$, the probability distribution of the transmission loss $p(\mathbf{u}|\mathbf{d})$ is obtained via Monte Carlo integration. From this TL probability distribution, all relevant statistics of TL can be obtained, such as the median, percentiles, and correlation coefficients.

Both the experimental data $\mathbf{d}$ and the usage domain model $\mathbf{u}$ are related to $\mathbf{m}$ via forward models $\mathbf{D}(\mathbf{m})$ and $\mathbf{U}(\mathbf{m})$, respectively. Thus formally, if the data were error free and the mappings were unique, we would have $\mathbf{u}=\mathbf{U}(\mathbf{D}^{-1}(\mathbf{d}))$. It is assumed that the mappings $\mathbf{D}(\mathbf{m})$ and $\mathbf{U}(\mathbf{m})$ are deterministic and all uncertainties (including noise and modeling errors) are in the data. Due to the uncertainties in the data, the inverse mapping from $\mathbf{d}$ to $\mathbf{m}$ is formulated in a probabilistic framework where one also can include prior information. The forward mapping could be probabilistic as in the textbook by Tarantola[15] and in the papers by Mosegaard and Tarantola[16] and Rogers *et al.*[17]

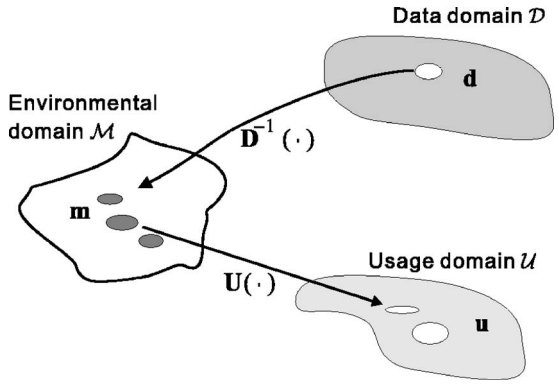


FIG. 1. An observation $\mathbf{d}$ ($\in\mathcal{D}$) is mapped into a distribution of environmental parameters $\mathbf{m}$ ($\in\mathcal{M}$) that potentially could have generated it. These environmental parameters are then mapped into the usage domain $\mathcal{U}$.

II. BAYESIAN INFERENCE

In the Bayesian paradigm, the solution to estimating parameters of interest $\mathbf{m}$ given an observation $\mathbf{d}$ is characterized by the posterior probability $p(\mathbf{m}|\mathbf{d})$. First, the prior information about the model parameter vector is quantified by the probability density function (pdf) $p(\mathbf{m})$. Then, this information is combined with the likelihood function $p(\mathbf{d}|\mathbf{m})$ provided by the combination of data and the physical model to give the posterior information of the model parameters $p(\mathbf{m}|\mathbf{d})$. A complete discussion of inverse theory from a probabilistic point of view may be found in the recent textbook by Tarantola.[15] The solution to the inverse problem is then

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d})} \propto \mathcal{L}(\mathbf{m})p(\mathbf{m}), \tag{1}$$

where $p(\mathbf{d})$ is a normalizing factor that makes the posterior probability density $p(\mathbf{m}|\mathbf{d})$ integrate to one. Since $p(\mathbf{d})$ does not depend on the environmental model $\mathbf{m}$, it typically is ignored in parameter estimation. Hence, as shown in the second representation, the normalization constant $p(\mathbf{d})$ is omitted and a brief notation $\mathcal{L}(\mathbf{m})$ is used to denote the likelihood function $p(\mathbf{d}|\mathbf{m})$.

Understanding and using the posterior distribution $p(\mathbf{m}|\mathbf{d})$ is at the heart of Bayesian inference. Specifically, one is interested in various features of the posterior distribution, such as the means, variances, and marginal distributions. These quantities can be written as expectations of functions $f(\mathbf{m})$ under $p(\mathbf{m}|\mathbf{d})$ as follows:

$$E[f(\mathbf{m})] = \int_{\mathcal{M}} f(\mathbf{m})p(\mathbf{m}|\mathbf{d})d\mathbf{m}. \tag{2}$$

For example, if the desired statistical quantity is the marginal posterior distribution of the parameter m_i, then

$$p(m_i|\mathbf{d}) = \int_{\mathcal{M}} \delta(m_i' - m_i)p(\mathbf{m}'|\mathbf{d})d\mathbf{m}'. \tag{3}$$

A. Data model

This section derives a likelihood function to be used in the probabilistic inversion following the same approach as described in Gerstoft and Mecklenbräuker.[2,18] At a single frequency, the relation between the observed complex-valued data vector $\mathbf{d}$ sampled at an N-element array and the modeled data $\mathbf{D}(\mathbf{m})$ is described by the model

$$\mathbf{d} = \mathbf{D}(\mathbf{m}) + \mathbf{e}, \tag{4}$$

where $\mathbf{e}$ represents the error term. The modeled data are given by $\mathbf{D}(\mathbf{m})=\mathbf{d}(\mathbf{m})s$, where the complex deterministic source term s is unknown. The transfer function $\mathbf{d}(\mathbf{m})$ is obtained using an acoustic propagation model for an environmental model $\mathbf{m}$.[19] For simplicity in the development below, data from only one frequency is assumed.

Assume the errors $\mathbf{e}$ to be Gaussian distributed with zero mean and covariance $\mathbf{C}_e$. The errors represent all features that are not modeled in the data such as noise, theoretical errors, and modeling errors.[2,7,15] Hence, the likelihood function is

$$\mathcal{L}(\mathbf{m},\mathbf{C}_e,s) = \frac{1}{\pi^N|\mathbf{C}_e|}\exp(-[\mathbf{d}-\mathbf{d}(\mathbf{m})s]^\dagger\mathbf{C}_e^{-1}[\mathbf{d} - \mathbf{d}(\mathbf{m})s]), \tag{5}$$

where N is the number of data points and superscript † denotes the complex conjugate transpose. Although in general not true, an independent and identically distributed (IID) error process $\mathbf{C}_e=\nu\mathbf{I}$ is assumed to describe the data uncertainty. The source term s can be estimated in closed form by requiring $\partial\log\mathcal{L}/\partial s=0$, whereby

$$s_{\mathrm{ML}} = \frac{\mathbf{d}^\dagger(\mathbf{m})\mathbf{d}}{\|\mathbf{d}(\mathbf{m})\|^2}. \tag{6}$$

It is seen that s depends on $\mathbf{m}$ but not on ν. After substituting s_{ML} back into Eq. (5), the likelihood function is then

$$\mathcal{L}(\mathbf{m},\nu) = \frac{1}{\pi^N\nu^N}\exp\left(-\frac{\phi(\mathbf{m})}{\nu}\right), \tag{7}$$

where

$$\phi(\mathbf{m}) = \|\mathbf{d}\|^2 - \frac{|\mathbf{d}^\dagger(\mathbf{m})\mathbf{d}|^2}{\|\mathbf{d}(\mathbf{m})\|^2} \tag{8}$$

is the objective function. Here, we treat the error variance ν as a nuisance parameter and eliminate it via integrating Eq. (7) weighted by a noninformative prior of ν $[p(\nu)=1/\nu]$ over its entire range[20]

$$\mathcal{L}(\mathbf{m}) = \int_0^\infty \mathcal{L}(\mathbf{m},\nu)p(\nu)d\nu. \tag{9}$$

Therefore, the likelihood function can be written as

$$\mathcal{L}(\mathbf{m}) = \frac{1}{\pi^N}\frac{(N-1)!}{\phi(\mathbf{m})^N}. \tag{10}$$

It is straightforward to extend the above formula to the multifrequency data set[20]

$$\mathcal{L}(\mathbf{m}) \propto \left[\frac{1}{\bar{\phi}^g(\mathbf{m})}\right]^{NJ} = [\Pi \phi_j(\mathbf{m})]^{-N}, \tag{11}$$

where J is the number of processed frequencies and $\bar{\phi}^g(\mathbf{m}) = \sqrt[J]{\Pi \phi_j(\mathbf{m})}$ is the geometric mean of the objective function over frequency.

The above derivation assumes that the errors are independent across both spatial samples of the acoustic field and frequencies. In practice these can be strongly correlated, for example, when the errors due to frequency-dependent modeling mismatch are the dominant source of error, the modeling error may not be independent across the frequencies used. Therefore, the number of independent samples NJ in Eq. (11) must be selected with care (see Sec. III B for details).

B. Prediction

A related problem is to infer what experimental values are likely to be observed given our knowledge of the environmental parameters. Thus, we are not just interested in the environment itself but also estimates in the information usage domain $\mathcal{U}$ (Fig. 1). In the present application, the usage domain is transmission loss (TL). The vector $\mathbf{u}$ is used to denote the transmission loss at I discrete positions, $u_i = u(r_i, z_i)$. For the example in Sec. III, we predicted the TL field on a 200×100 grid of range-depth cells, $I = 200 \times 100 = 20\,000$, inferred from a 13-dimensional model $\mathbf{m}$.

Probability density functions that describe yet unobserved events are referred to as predictive distributions. Based on the posterior distribution $p(\mathbf{m}|\mathbf{d})$, the posterior predictive distribution $p(\mathbf{u}|\mathbf{d})$ is obtained from the joint posterior pdf of $\mathbf{u}$ and $\mathbf{m}$ given $\mathbf{d}$,

$$p(\mathbf{u}|\mathbf{d}) = \int_{\mathcal{M}} p(\mathbf{u},\mathbf{m}|\mathbf{d}) d\mathbf{m} = \int_{\mathcal{M}} p(\mathbf{u}|\mathbf{m},\mathbf{d}) p(\mathbf{m}|\mathbf{d}) d\mathbf{m}, \tag{12}$$

where the second equation follows from the definition of conditional probability. Since all uncertainties are assumed to be in the data $\mathbf{d}$ and all information in $\mathbf{d}$ has been mapped into $\mathbf{m}$ (see Fig. 1 and the discussion in the last paragraph of Sec. I), conditioning on $\mathbf{d}$ adds no information in our prediction of $\mathbf{u}$. Therefore,

$$p(\mathbf{u}|\mathbf{m},\mathbf{d}) = p(\mathbf{u}|\mathbf{m}). \tag{13}$$

The conditional probability density $p(\mathbf{u}|\mathbf{m})$ is used to describe uncertainties in the forward mapping due to imperfect knowledge of the environment (e.g., parametrization).[15–17] Here, the forward mapping is assumed exact: a functional relationship $\mathbf{u} = \mathbf{U}(\mathbf{m})$ gives the transmission loss $\mathbf{u}$ exactly for each value of $\mathbf{m}$. Note that $\mathbf{u} = \mathbf{U}(\mathbf{m})$ is a short notation for the set of equations $u_i = U_i(\mathbf{m})$, $i = 1, \ldots, I$. Therefore, the probability density is

$$p(\mathbf{u}|\mathbf{m}) = \delta(\mathbf{U}(\mathbf{m}) - \mathbf{u}), \tag{14}$$

where the vector delta function is defined as the product of the delta functions for the elements[15,21] as in

$$\delta(\mathbf{U}(\mathbf{m}) - \mathbf{u}) = \prod_{i=1}^{I} \delta(U_i(\mathbf{m}) - u(r_i, z_i)). \tag{15}$$

The posterior predictive distribution of $\mathbf{u}$ for a set of discrete ranges and depths given the observed acoustic data $\mathbf{d}$ is obtained by integrating the values of the TL with respect to the posterior distribution of the model parameters

$$p(\mathbf{u}|\mathbf{d}) = \int_{\mathcal{M}} \delta(\mathbf{U}(\mathbf{m}) - \mathbf{u}) p(\mathbf{m}|\mathbf{d}) d\mathbf{m} \tag{16}$$

which has the same form as Eq. (2). As shown in the Appendix, this is a generalization of the transformation of random variables using the properties of the Dirac delta function. However, in the present case, neither the roots nor the derivatives are known, and thus it is easier to implement Eq. (16) directly as described in Sec. II C.

The posterior distribution $p(\mathbf{u}|\mathbf{d})$ carries all the information about the TL in the presence of the geoacoustic inversion uncertainties. As the predictive distributions are not necessarily Gaussian, it is preferable to characterize the distributions with medians and distance between the 5th and 95th percentiles instead of means and standard deviations. Note that the median corresponds to the 50th percentile of the distribution. The βth percentile of the TL distribution at a given position, denoted by $u^{\beta\%}$, is computed by finding the TL value that satisfies

$$\int_{-\infty}^{u^{\beta\%}} p(u|\mathbf{d}) \mathrm{d}u = \beta/100. \tag{17}$$

In addition to summarizing the statistics of TL at any particular point, the covariance structure of the TL at two points might also be of interest in uncertain acoustic environments. From Eq. (16), the covariance and correlation coefficient between the TL at two positions u_i and u_j can be computed, respectively, by

$$\mathrm{cov}(u_i, u_j) = E[u_i u_j] - E[u_i] E[u_j] \tag{18}$$

and

$$R_{ij} = \frac{\mathrm{cov}(u_i, u_j)}{\sqrt{\mathrm{cov}(u_i, u_i)}\sqrt{\mathrm{cov}(u_j, u_j)}}. \tag{19}$$

To compute the above statistical quantities of TL, one needs to evaluate the high-dimensional integral of Eq. (16). The integral can be approximated numerically as described in the next section.

C. Markov chain Monte Carlo method

Monte Carlo methods can evaluate integrals in high-dimensional space efficiently.[22] In particular, Markov chain Monte Carlo (MCMC) algorithms have been found to be well suited for problems of Bayesian inference. The commonly used MCMC methods are the Metropolis-Hastings algorithm, which was introduced first in Ref. 23, and Gibbs sampling which was developed originally in Ref. 24 where it was applied to image processing. MCMC are extensively

used in various fields of inverse problems, such as geophysics,[16,25] ocean acoustics,[7,26,27] and electromagnetics.[28] MCMC algorithms consist of a random walk in the parameter space where the next parameter value depends only on the current value. After an initial "burn-in" period in which the random walker moves toward the high posterior probability region, the chain samples a desired posterior pdf, that is, it returns a number of parameter vectors that are distributed as in the posterior pdf.

In the MCMC, samples are generated from the posterior distribution $p(\mathbf{m}|\mathbf{d})$. The difficult part is to create a Markov chain which converges rapidly. As noted by many authors,[7,29,30] parameter coupling frequently is encountered in ocean acoustics. High correlation between parameters can slow down the convergence of a MCMC sampler considerably. Thus, a parameter covariance matrix estimated from the sampled models during the initial "burn in" period[7] is used for determining appropriate coordinate rotations.

MCMC convergence was established by collecting two independent runs in parallel and periodically comparing the marginal distributions of the parameters estimated from each run.[7] The procedure is terminated when the maximum difference between two cumulative marginal distributions for all parameters is less than 0.05. A good introduction to MCMC methods is in Ref. 31, which also contains many applications in statistical data analysis.

The integral in Eq. (16) is the expectation of function $\delta(\mathbf{U}(\mathbf{m})-\mathbf{u})$ with respect to the posterior distribution of the model parameters. This and other expectations can be approximated by using the MCMC samples $\{\mathbf{m}^{(t)}\}$ drawn from the posterior distribution of model parameters $p(\mathbf{m}|\mathbf{d})$

$$p(\mathbf{u}|\mathbf{d}) = \frac{1}{T}\sum_{t=1}^{T} \delta(\mathbf{U}(\mathbf{m}^{(t)}) - \mathbf{u}), \tag{20}$$

where the superscript t is used to label the sequence of states in a Markov chain and T denotes the total length of the sequence. To implement Eq. (20), a numerical approximation is made by binning the calculated TL values. The bin width is selected small enough to have negligible effect on the distribution. Here a 1 dB bin width is used.

Using all samples from MCMC runs can consume a large amount of storage to save all $\mathbf{m}^{(t)}$ and computation time to compute $p(\mathbf{u}|\mathbf{d})$. It has been suggested in the statistical literature[31–33] that inferences should be based on a subsampling of each sequence, with a subsampling factor high enough that successive draws of $\mathbf{m}$ are approximately independent. The strategy is known as *subsampling*.[33] This can save a large amount of storage and computation time for using the MCMC samples in inference. This subsampling reduces the number of samples needed to calculate $p(\mathbf{m}|\mathbf{d})$ and thus translates into a large saving in computer time for calculating $p(\mathbf{u}|\mathbf{d})$. Practically, we use a Monte Carlo (random) subsampling of the MCMC samples $\{\mathbf{m}^{(t)}\}$ and monitor the convergence of the maximum difference between the marginal cumulative distributions estimated from subsamples and from all MCMC samples. This maximum difference should be less than 0.05 for all parameters. Then these subsampled model parameter vectors are used to compute $p(\mathbf{u}|\mathbf{d})$.

All results presented in this paper are generated by SAGA,[34] which implements the method described in Ref. 28.

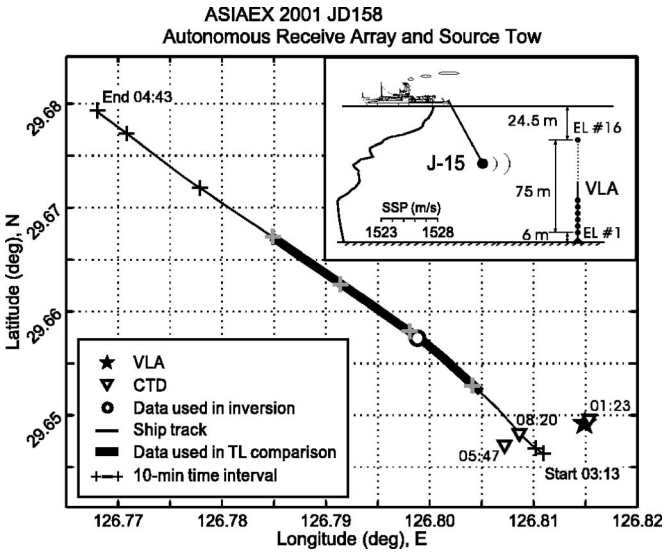


FIG. 2. Track of R/V *Melville* during the ASIAEX 2001 East China Sea experiment. The experimental geometry is shown in the upper right-hand corner of the figure.

III. RESULTS AND DISCUSSION

Data from the ASIAEX 2001 East China Sea experiment[35] are used to illustrate the approach. Figure 2 shows a map of the region where the acoustic measurements were taken. On Julian Day (JD) 158, acoustic energy was transmitted from the J-15 source towed near 48 m depth by R/V *Melville* with a speed of about 3 knots. The ship track is indicated by the line in the figure on which the distances between the source and the receiver range from 0.5 to 6 km. The experiment geometry is illustrated schematically in Fig. 2. A 16-element, 75-m aperture, autonomous recording vertical line array (VLA) was moored up from the seafloor at location 29°38.927′ N, 126°48.892′ E where the measured water depth was approximately 105.5 m. The lowermost element (element 1) was about 6 m above the bottom. Element 4 failed during deployment.

For oceanographic measurements, the current profile in the water column from 30 to 100 m was obtained by a ship-mounted ADCP system. During the acoustic transmissions, there exists a strong tidal current with magnitude greater than 0.5 m/s around the middle of the water column. Three sound speed profiles were measured by CTDs on JD 158. As shown in Fig. 2, typical summer sound speed profile characteristics were observed with significant fluctuations in the thermocline.

A general bathymetric and geological survey has indicated that in the neighborhood of the experimental site, the environment is nearly range independent. Additional details of the seismic and oceanographic experiments can be found in Ref. 35.

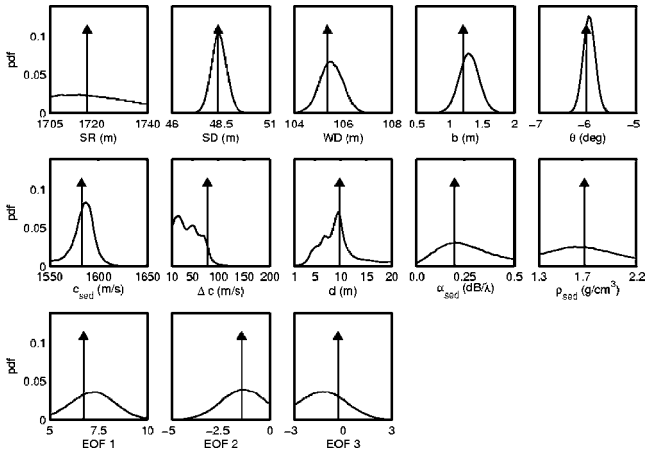


FIG. 3. 1D marginal posterior probability densities of the model parameters using the measured data obtained at approximately 1.7 km from the source. Arrows indicate the estimated optimum values of the parameters.

A. Baseline model

The baseline model is assumed to be range-independent and consists of an ocean layer overlying a uniform sediment layer on top of a subbottom. The model parameters were divided into three subsets: geometrical, geoacoustic, and ocean sound speed parameters. The geometrical parameters include source range SR, source depth SD, water depth WD, and the array shape (array tilt θ and bow b). The geoacoustic parameters include sediment compressional speed c_{sed}, density ρ_{sed}, attenuation α_{sed}, and thickness d, and increment of subbottom compressional speed from the top sediment layer Δc (subbottom density and attenuation are fixed at 2.4 g/cm^3 and 0.01 dB/λ, respectively). The ocean sound speed was modeled by a linear combination of empirical orthogonal functions (EOFs). An empirical orthogonal function (EOF) analysis at the experimental site shows that the first 3 EOFs contain about 90% of the energy. Therefore, the number of representative EOFs was set to three in the inversion.

An environmental domain of 13 parameters with their search bounds is indicated in Fig. 3, including (from upper to lower panels) geometrical, geoacoustic, and ocean sound speed EOF coefficients.

B. Posterior distributions for the model parameters

Matched-field (MF) geoacoustic inversion using the selected frequencies 195, 295, and 395 Hz was carried out with the measured data obtained at approximately 1.7 km from the source (the circle in Fig. 2). The MCMC algorithm along with the normal-mode propagation model SNAP (Ref. 19) is employed to sample the posterior probability density in model domain $\mathcal{M}$.

Figure 3 shows the marginal posterior distributions of the model parameters using the likelihood function, Eq. (11) with the number of independent samples, NJ, found as follows. First, the data error covariance matrix $\mathbf{C}_e$ is estimated using a maximum-likelihood approach. It is based on an ensemble average of the residual vectors (residual field between the observed and the modeled field generated from the optimum values of the parameters) from multiple inversions of vertical array data from a source tow.[36] For each source range, the residual vectors of processed frequencies are concatenated creating an error vector consisting of $15\times 3=45$ entries (15 hydrophone elements by 3 frequencies). A total of 98 error vectors are used to estimate $\mathbf{C}_e$. Then, to find the number of independent samples NJ, the eigenvalue analysis is performed on the estimated $\mathbf{C}_e$. The result shows there are 30 significant eigenvalues, containing 99.9% of the energy, in the error covariance matrix. Therefore, the number of independent samples, NJ, for this analysis, is 30.

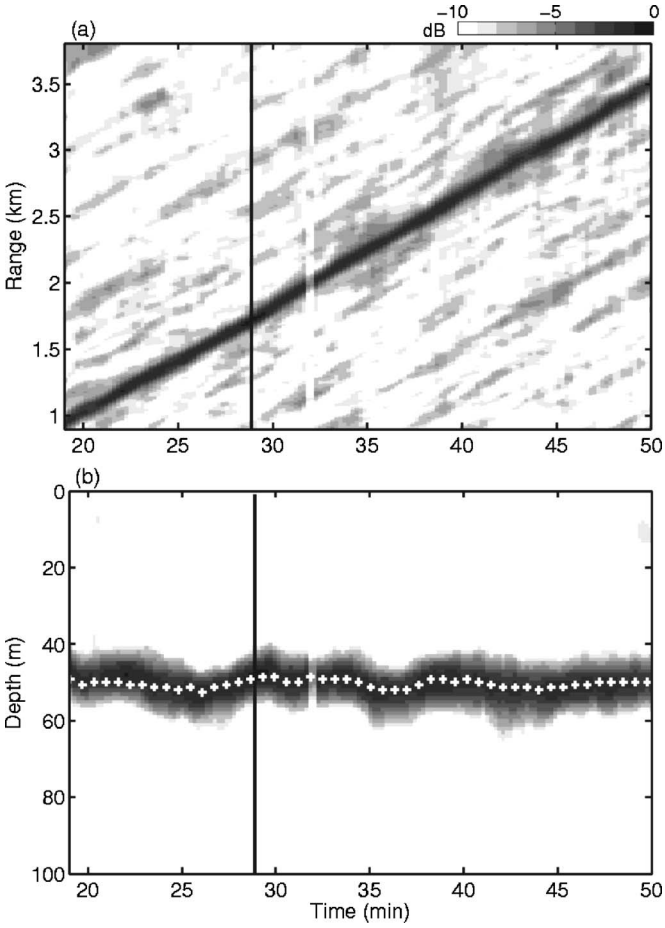


FIG. 4. MF-derived (a) source-receiver range and (b) source depth over the time interval from 19 to 50 min. The contour plots shows the MF output (dB) where the best match for each time sample is 0 dB. The vertical line on each plot indicates where the environmental model is estimated. Plus signs indicate the true measured source depths.

Figure 4 shows the MF-derived source position over the time interval from 19 to 50 min using the estimated optimum values of the parameters found from the above inversion. The source depths measured by the depth sensor are indicated by the plus signs. Compared with the GPS and the depth sensor measurements, MF-derived source position is consistent with the experimental configuration. Source localization based on the best-fit model tracks the actual source positions well.

C. Predictive distributions of transmission loss

With the posterior probability density of the environmental model parameters obtained from the inversion, we quantify the uncertainty mapped from the model parameters to the predicted transmission loss (TL). The posterior predic-

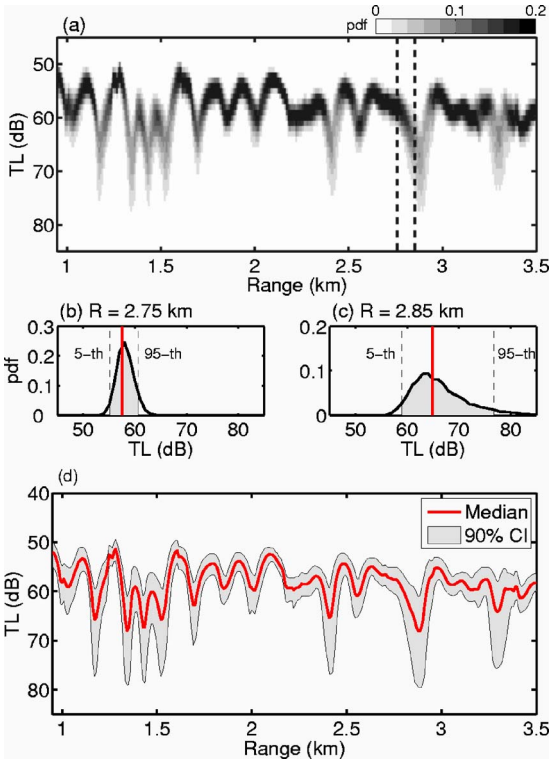


FIG. 5. (Color online) Posterior distribution of TL versus range for 295 Hz for array element 7 (at 69.5 m): (a) Contour of posterior distribution for TL versus range. (b) and (c) Posterior distributions of TL at two different ranges (2.75 km and 2.85 km). These corresponds to cuts (vertical dashed lines) through the contour. (d) Statistics of the predicted TL versus range. The solid line with gray area around shows the median and the 90% interval of posterior distribution.

tive distribution of TL for the position (r_i, z_i) is obtained by integrating the predictions of TL with respect to the posterior distribution of the model parameters, using Eq. (16).

Figure 5 shows the posterior distribution of the TL versus range at 295 Hz for array element 7 (at 69.5 m). Figure 5(a) shows the contour of the predictive distribution of the TL versus range. Gray levels represent the probability density. Darker shades mean higher probability of observing the predicted TL value. It is observed that at some ranges where the acoustic field is near a null (destructive interference), the predictive probabilities show large variations in the result.

Predictive distributions at two different ranges are shown in Figs. 5(b) and 5(c), which correspond to the points of constructive and destructive interferences, respectively. At the range of constructive interference [Fig. 5(b)], less variation of TL is observed. Therefore, the probability density concentrates in a smaller area. However, near the range of destructive interference [Fig. 5(c)], the probability density spreads in a larger area which indicates the acoustic field is more difficult to predict. Since the distribution of TL is often poorly approximated by a normal distribution, particularly near destructive interferences, the central tendency and spread of the TL distribution are indicated, respectively, by the median (heavy vertical line) and the distance between the 5th and 95th percentiles (gray area; referred to as the 90% Credibility Interval). Figure 5(d) summarizes the predictive

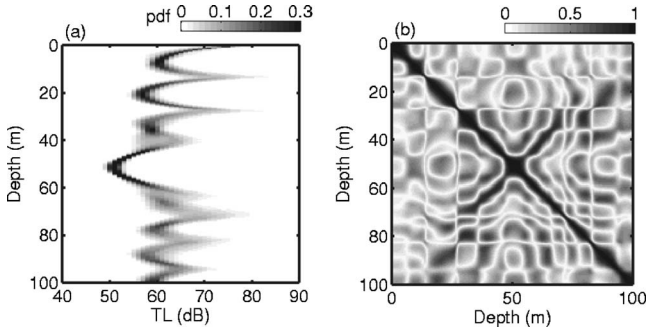


FIG. 6. Posterior probability distribution of TL versus depth for 295 Hz at the 2.85 km range: (a) Contour of posterior probability distribution for TL versus depth. (b) Magnitude of the correlation coefficient matrix for TL at all depths.

distributions by the median (heavy line) and the 90% CI (gray area). This is a practical way to convey the uncertainty in TL.

Figure 6 shows the posterior distribution of TL versus depth for 295 Hz at 2.85-km range. Figure 6(a) shows the contour of posterior probability distribution for TL versus depth. We see a similar constructive/destructive interference pattern as observed in the range contour of the TL distribution. The vertical covariance structure of the TL is examined in Fig. 6(b). Due to the interference of the normal modes, a chessboardlike correlation structure of the TL is observed. For regions near the constructive interference (for example, at a depth of 50 m), the TL is correlated more at neighboring depths. For regions near destructive interference (at 81 m depth), the correlation drops rapidly.

To demonstrate the correlation structure in detail, Fig. 7 shows the 2D posterior probability distributions between TL

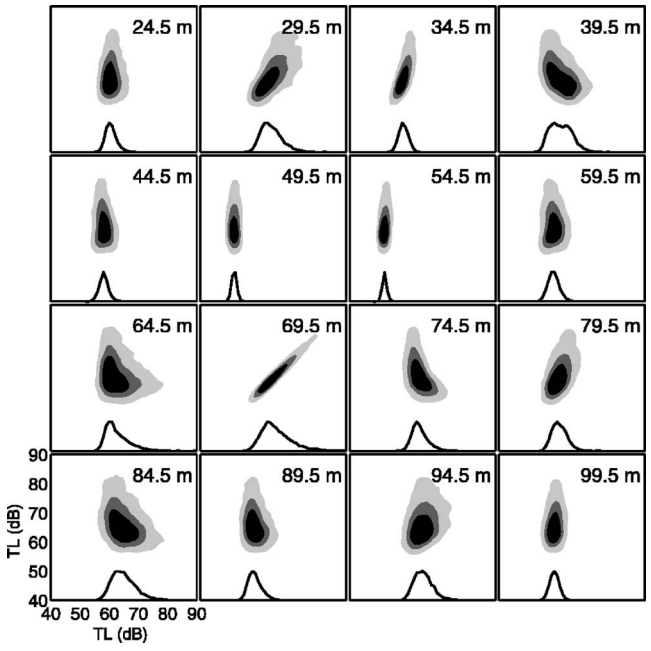


FIG. 7. 2D posterior probability distribution of TL versus depth for 295 Hz at the 2.85 km range. The vertical and horizontal axes indicate, respectively, the TL field at 69.5 m depth and that at the depth indicated on each panel. The gray-scale coloring from darkest to lightest represents 50%, 75%, and 95% highest posterior density (HPD) (Ref. 20). 1D posterior probability distribution of TL at that depth is also shown on the bottom of each panel.

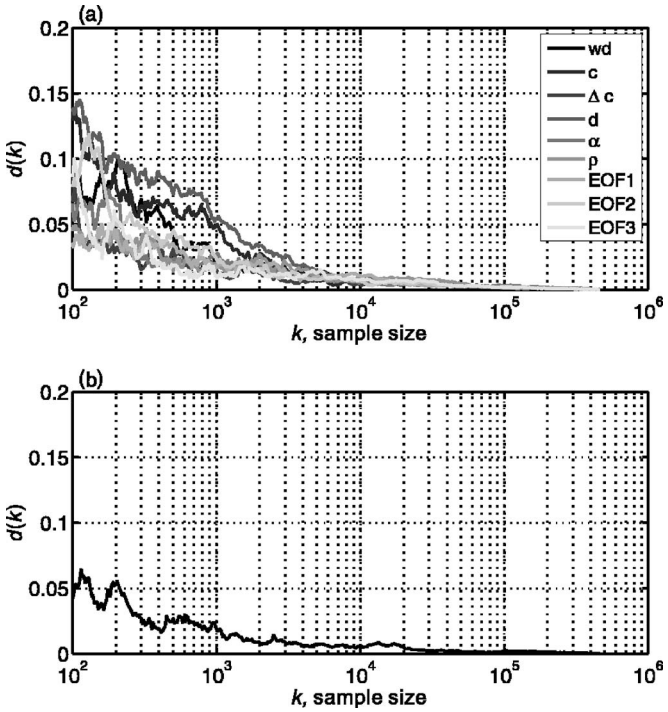


FIG. 8. Convergence of (a) the posterior probability distribution of each model parameter and (b) the predictive probability distribution of the TL at 69.5 m depth and 2.85 km range [see Fig. 5(c) for the distribution]. The vertical axis indicates the maximum difference between the cumulative distributions of k-length subsamples and the cumulative distribution of the full MCMC samples (the length of 480 000).

at 69.5 m depth (vertical axis) and TL at selected depths (horizontal). Gray levels represent the probability density; darker shade means higher probability density. The line plot on the bottom of each panel is the marginal distribution of TL at the corresponding depth, which corresponds a cut through Fig. 6(a) at that corresponding depth.

As discussed in Sec. II C, significant saving in both storage and computation time can be obtained by subsampling MCMC samples. Note that about 480 000 samples were required for the MCMC to converge. Figure 8(a) shows the convergence of the Monte Carlo subsampling for each model parameter. We find that about 10 000 samples are sufficient to characterize the marginal distributions of model parameters.

As an indication of convergence for TL distribution, we computed the marginal distributions of TL for the subsampled model parameter vectors. Figure 8(b) shows convergence for the marginal probability distributions of TL at 69.5 m depth and 2.85 km range, which corresponds to a long tail distribution as shown in Fig. 5(c). Similar to sampling the marginal distributions of model parameters, 10 000 samples can capture accurately the predictive distribution of the TL at this chosen position (with maximum error 0.025).

D. Experimental comparisons

We have demonstrated how to estimate the statistical properties of the TL in the presence of uncertainty embedded in the environmental model parameters. To further illustrate the versatility and usefulness of the predictive distributions of the TL, the resulting statistics are compared with actual TL observations.

Bayesian inference gives us the posterior distribution of the full parameter vector. To estimate the statistical properties of the TL, only the posterior distribution of geoacoustic parameters and ocean sound speed EOFs is required, but not the distribution of the geometric parameters. The environmental parameters are the geoacoustic parameters, ocean sound speed EOF coefficients, and water depth (water depth is included since it affects the number of propagating modes in the waveguide). Uncertainties in these parameters can be obtained easily by integration over the remaining geometric parameters, that is, simply removing these variables (SD, SR, b, and θ) from the parameter vector.

Source depth is an important parameter for predicting TL fields accurately. In this data set, the depth sensor measurement indicates that the source varied between 48 and 52 m. Since the measured and MF estimated source depths (as shown in Fig. 4) are virtually the same, the MF-derived time-varying source position is included in the TL prediction, referred to as the MFSD model.

Figure 9 compares the observed TL (dots) with the predicted TL statistics of the MFSD model (solid line with gray area) for the frequencies 195, 295, and 395 Hz (left to right) and for array elements 1, 7, and 16 (bottom to top; depths at 99.5, 69.5, and 24.5 m). We see that for 195 Hz the TL uncertainty band is about 5 dB near the ranges of constructive interference and is widened near the ranges of destructive interference. As frequency increases, larger spreads in TL predictions are observed. This is most pronounced near regions of destructive interference. In general, the predicted TL patterns using the MFSD model follow the trends of the measured TL well. Table I summarizes comparisons of measured and predicted TL from Fig. 9. As frequency increases, the uncertainty band of the predicted TL (number in dB) increases by approximately 3 dB and more of the observed TL points are within the 90% CI.

To investigate the effect of the source depth uncertainty, we have estimated the TL distributions for 295 Hz for the following three additional cases:

(1) The marginal posterior distribution of source depth obtained from the 1.7 km inversion is assumed for all ranges, referred to as the SD@1.7 km model.
(2) The range-dependent source depth variability is accounted for by the statistics of the measured source depth. The parameter SD is treated as a random variable having the mean value of 50.2 m and standard deviation of 1.5 m, estimated from the measured source depths. At each range, the source depth is a Gaussian distribution centered at 50.2 m with uncertainty band of 1.5 m, referred to as the MEAN+STD model.
(3) The uncertainty band in the MEAN+STD model is applied to the MFSD model. At each range, the source depth is assumed to be a Gaussian distribution centered at the MF-derived source depth with standard deviation of 1.5 m, referred to as the MEAN+STD model.

Figure 10 shows comparisons of predicted and measured TL for the above described source depth distribution models

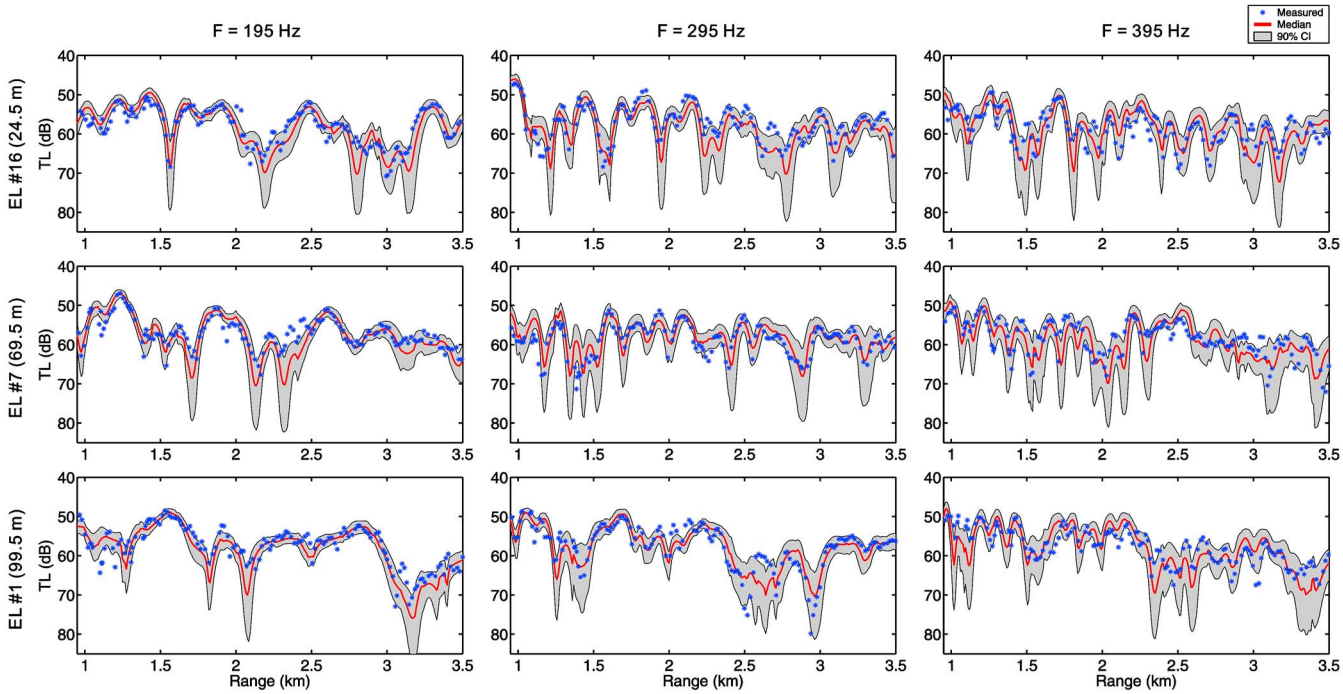


FIG. 9. (Color online) Predicted and measured TL (dots) for array elements 1, 7, and 16 and for frequencies 195, 295, and 395 Hz. The median of the predicted TL (solid line) is shown together with the 90% CI (gray area). The source depth is estimated from MF processing.

for a frequency of 295 Hz. The prediction quality is summarized in Table I. The results using the marginal posterior distribution of the source depth obtained at the 1.7 km range (left column) show that the prediction quality is rather poor. For instance, for array element 7 (middle row) the constructive interference region near 1.6 km and the destructive interference region near 2.2 km are not captured by the predictions. This is expected since the source depth at each range differs substantially from the estimated source depth at the range of 1.7 km. For the MEAN+STD model (middle column), we see that the fine scale features of the observed TL are matched by the predictions. Compared with the MFSD model (middle column in Fig. 9), the addition of source depth uncertainty results in the TL uncertainty band being much wider. The predicted TL patterns follow the trends of the measured TL well. For the MFSD+STD model (right column), it shows the highest percentage of the observed TL inside the CI. Table I shows that the median value of predicted TL spread increases by approximately 3 dB more than the MFSD alone, and about 7%–11% more of the observed TL points fall within the gray area.

We found that, in general, about 80% of the observed TL data falls within the predicted 90% CI. Since the predicted TL statistics are derived from uncertainty in geoacoustic parameters $p(\mathbf{m}|\mathbf{d})$ for the given environmental parameterization only (we assume a range-independent environment). Complicated environments, such as spatial and temporal fluctuations in the water column, sediment, sea surface, and water-sediment interface, are not modeled and this will increase the error. Further, all noise sources have not been taken into account. Therefore, the percentage of observed data points inside the computed CI is less than the predicted.

IV. CONCLUSION

This paper investigates the statistical estimation of TL based on the posterior probability density of environmental parameters obtained from the geoacoustic inversion process. First, a Markov chain Monte Carlo procedure is employed to sample the posterior probability density of the geoacoustic parameters. Then, these parameter uncertainties are mapped to the transmission loss domain where a full multidimensional probability distribution of the TL as a function of

TABLE I. Summary of TL prediction performance. Numbers in dB indicate the median value of the predicted TL spread over all range, while numbers in % represent the percentage of the measured TL points that lie inside the 90% CI.

	MFSD (Fig. 9)			F=295 Hz (Fig. 10)		
Element number (depth)	195 Hz	295 Hz	395 Hz	SD@1.7 km	MEAN+STD	MESD+STD
16 (99.5 m)	4.9 dB/73%	6.2 dB/80%	7.7 dB/82%	7.2 dB/61%	11 dB/88%	9.6 dB/91%
7 (69.5 m)	4.3 dB/73%	6.1 dB/85%	8.2 dB/79%	7.3 dB/69%	9.6 dB/93%	9.3 dB/92%
1 (24.5 m)	4.6 dB/69%	5.4 dB/80%	7.8 dB/81%	4.9 dB/67%	6.9 dB/85%	7.4 dB/91%

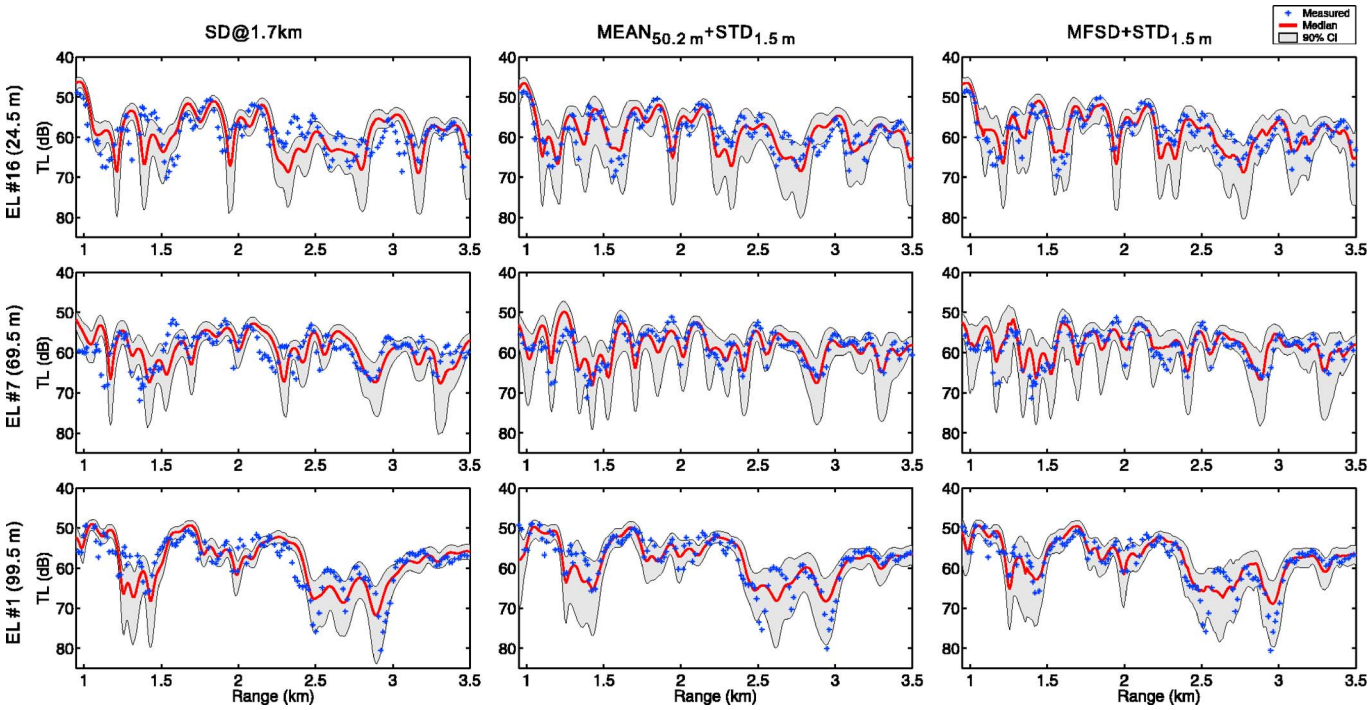


FIG. 10. (Color online) Predicted and measured TL for 295 Hz for various source depth distribution models. Left column (SD @ 1.7 km): the source depth is inferred from the inversion at 1.7 km; Middle column (MEAN+STD): fixed source depth (50.2 m) with the standard deviation (1.5 m); Right column (MFSD+STD): MF-derived source depth with 1.5-m standard deviation.

range and depth is obtained. The summary statistics of predicted TL including the median, percentiles, and correlation coefficients are considered.

A Monte Carlo subsampling technique is applied to subsample the full MCMC model parameter samples. A significant saving in both storage and computation time (a factor of 50) was observed using this technique.

The predicted TL statistics are compared with actual TL observations from the ASIAEX 2001 East China Sea experiment. In general, about 80% of the observed TL data falls within 90% of the range-varying predicted TL probability distribution. Thus, the geoacoustic inversion has captured most of the uncertainty in the environment.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under Grant No. N00014-05-1-0264.

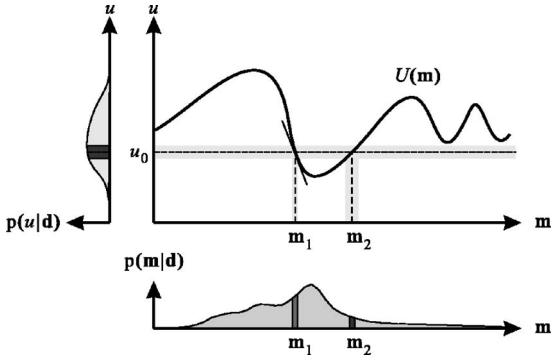


FIG. 11. Model-utility relationship: many-to-one transformations. In the center panel, horizontal axis represents the model domain, vertical axis represents the utility domain. The bottom and left panels indicate $p(\mathbf{m}|\mathbf{d})$ and $p(u|\mathbf{d})$, respectively.

APPENDIX: TRANSFORMATION OF RANDOM VARIABLES

Given that $u=U(\mathbf{m})$ and assuming that $U(\mathbf{m})$ is a monotonic function of $\mathbf{m}$ (monotonic assumption only true in this paragraph), the posterior pdf of $\mathbf{m}$ is related to the posterior pdf of u by the transformation of random variables

$$p(u|\mathbf{d}) = p(\mathbf{m}|\mathbf{d}) \left| \frac{\partial \mathbf{m}}{\partial U(\mathbf{m})} \right|, \tag{A1}$$

where $|\partial \mathbf{m}/\partial U(\mathbf{m})|$ is the absolute value of the Jacobian determinant, whose reciprocal represents the hypervolume in the $\mathcal{U}$ domain mapped out by the small hypercube region in the $\mathcal{M}$ domain.

Propagation of parameter uncertainties to the TL predictions by integration in the sense of Eq. (16) can be related to the transformation of random variables as in Eq. (A1) using the properties of the Dirac delta function. Suppose that $f(x)=0$ has N zeros $\{x_n\}$ and $\mathrm{d}f(x_n)/\mathrm{d}x \neq 0$, then $\delta(f(x))$ equals a sequence of impulses at $x=x_n$ of area $|\mathrm{d}f(x_n)/\mathrm{d}x|^{-1}$,[37] i.e.,

$$\delta(f(x)) = \sum_n \delta(x - x_n) \left| \frac{\mathrm{d}f(x_n)}{\mathrm{d}x} \right|^{-1}. \tag{A2}$$

Using Eq. (A2), Eq. (16) can be rewritten as

$$p(u|\mathbf{d}) = \int_{\mathcal{M}} \delta[U(\mathbf{m}) - u] p(\mathbf{m}|\mathbf{d}) d\mathbf{m} = \sum_{n=1}^{N} p(\mathbf{m}_n|\mathbf{d}) \times \left| \frac{\partial U(\mathbf{m}_n)}{\partial \mathbf{m}} \right|^{-1}, \tag{A3}$$

where $\{\mathbf{m}_n\}$ are the roots of the equation $U(\mathbf{m})-u=0$. Equa-

tion (16) is the generalization of Eq. (A1) to many-to-one transformations.

Equation (A3) can be explained intuitively using Fig. 11. For a nonmonotonic function $U(\mathbf{m})$, the probability mass of any specific value u_0 can be found by first solving for the roots of the equation $u_0 = U(\mathbf{m})$, then calculating the inverse of $|\partial U(\mathbf{m})/\partial \mathbf{m}|$ at the roots $\mathbf{m}_n$ weighted according to $p(\mathbf{m}_n|\mathbf{d})$, and finally summing all probability masses.

[1]P. Gerstoft, "Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions," J. Acoust. Soc. Am. **95**, 770–782 (1994).

[2]P. Gerstoft and C. F. Mecklenbräuker, "Ocean acoustic inversion with estimation of *a posteriori* probability distributions," J. Acoust. Soc. Am. **104**, 808–819 (1998).

[3]M. I. Taroudakis and M. G. Markaki, "Bottom geoacoustic inversion by broadband matched field processing," J. Comput. Acoust. **16**, 167–183 (1998).

[4]L. Jaschke and N. R. Chapman, "Matched field inversion of broadband data using the freeze bath method," J. Acoust. Soc. Am. **106**, 1838–1851 (1999).

[5]G. R. Potty, J. H. Miller, J. F. Lynch, and K. Smith, "Tomographic inversion for sediment parameters in shallow water," J. Acoust. Soc. Am. **108**, 973–986 (2000).

[6]D. P. Knobles, R. A. Koch, L. A. Thompson, K. C. Focke, and P. E. Eisman, "Broadband sound propagation in shallow water and geoacoustic inversion," J. Acoust. Soc. Am. **113**, 205–222 (2003).

[7]S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion I: A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).

[8]F. B. Jensen and N. Pace, *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance* (Kluwer Academic, The Netherlands, 2002).

[9]P. Abbot and I. Dyer, "Sonar performance predictions based on environmental variability," in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer Academic, The Netherlands, 2002), pp. 611–618.

[10]K. R. James and D. R. Dowling, "A probability density function method for acoustic field uncertainty analysis," J. Acoust. Soc. Am. **118**, 2802–2810 (2005).

[11]S. Finette, "Embedding uncertainty into ocean acoustic propagation models (l)," J. Acoust. Soc. Am. **117**, 997–1000 (2005).

[12]D. Tielburger, S. Finette, and S. Wolf, "Acoustic propagation through an internal wave field in a shallow water waveguide," J. Acoust. Soc. Am. **101**, 789–808 (1997).

[13]D. Rouseff and T. E. Ewart, "Effect of random sea surface and bottom roughness on propagation in shallow water," J. Acoust. Soc. Am. **98**, 3397–3404 (1995).

[14]P. Gerstoft, C.-F. Huang, and W. S. Hodgkiss, "Estimation of transmission loss in the presence of geoacoustic inversion uncertainty," IEEE J. Ocean. Eng. **31**, April issue, 2006.

[15]A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (SIAM, Philadelphia, 2005).

[16]K. Mosegaard and A. Tarantola, "Probabilistic approach to inverse problems," in *International Handbook of Earthquake & Engineering Seismology, Part A* (Academic, London, 2002), pp. 237–265.

[17]L. T. Rogers, M. Jablecki, and P. Gerstoft, "Posterior distributions of a statistic of propagation loss inferred from radar sea clutter," Radar Science **40**, 1–14 (2005).

[18]C. F. Mecklenbräuker and P. Gerstoft, "Objective functions for ocean acoustic inversion derived by likelihood methods," J. Comput. Acoust. **8**, 259–270 (2000).

[19]F. B. Jensen and M. C. Ferla, "SNAP: The SACLANTCEN normal-mode acoustic propagation model," SACLANT Undersea Research Centre, SM-121, La Spezia, Italy (1979).

[20]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Uncertainty analysis in matched-field geoacoustic inversions," J. Acoust. Soc. Am. **119**, 197–207 (2006).

[21]R. N. Bracewell, *The Fourier Transform and its Applications*, 3rd ed. (McGraw-Hill, New York, 2000).

[22]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in Fortran 77*, 2nd ed. (Cambridge University Press, London, 1992).

[23]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," J. Chem. Phys. **21**, 1087–1092 (1953).

[24]S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984).

[25]M. K. Sen and P. L. Stoffa, "Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion," Geophys. Prospect. **44**, 313–350 (1996).

[26]Z.-H. Michalopoulou and M. Picarelli, "Gibbs sampling for time-delay- and amplitude estimation in underwater acoustics," J. Acoust. Soc. Am. **117**, 799–808 (2005).

[27]D. Battle, P. Gerstoft, W. S. Hodgkiss, W. A. Kuperman, and P. Nielsen, "Bayesian model selection applied to self-noise geoacoustic inversion," J. Acoust. Soc. Am. **116**, 2043–2056 (2004).

[28]C. Yardim, P. Gerstoft, and W. S. Hodgkiss, "Estimation of radio refractivity from radar clutter using Bayesian Monte Carlo analysis," IEEE Trans. Antennas Propag. **54**, 1318–1327 (2006).

[29]M. D. Collins and L. Fishman, "Efficient navigation of parameter landscapes," J. Acoust. Soc. Am. **98**, 1637–1644 (1995).

[30]G. L. D'Spain, J. J. Murray, W. S. Hodgkiss, N. O. Booth, and P. W. Schey, "Mirages in shallow water matched field processing," J. Acoust. Soc. Am. **105**, 3245–3265 (1999).

[31]W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London, 1996).

[32]C. J. Geyer, "Practical Markov chain Monte Carlo (with discussion)," Stat. Sci. **7**, 473–483 (1992).

[33]C. P. Robert and G. Casella, *Monte Carlo Statistical Methods* (Springer-Verlag, New York, 1999).

[34]P. Gerstoft, SAGA Users guide 5.0, an inversion software package, An updated version of "SAGA Users guide 2.0, an inversion software package," SACLANT Undersea Research Centre, SM-333, La Spezia, Italy (1997).

[35]C.-F. Huang and W. S. Hodgkiss, "Matched field geoacoustic inversion of low frequency source tow data from the ASIAEX East China Sea experiment," IEEE J. Ocean. Eng. **29**, 952–963 (2004).

[36]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Data error covariance matrix for vertical array data in an ocean waveguide," J. Acoust. Soc. Am. **119**, 3272 (2006).

[37]A. Papoulis, *Systems and Transforms with Applications in Optics* (McGraw-Hill, New York, 1968).

Observations of biological choruses in the Southern California Bight: A chorus at midfrequencies

G. L. D'Spain[a] and H. H. Batchelor
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 93940-0701

(Received 30 May 2006; accepted 1 August 2006)

This paper describes the characteristics of an underwater biological chorus recorded in the midfrequency band (1–10 kHz) in the Southern California Bight. The recordings were made in July, 2002 by a large-vertical-aperture, 131-element, 2D billboard array. The chorus, observed predominantly on two consecutive nights during the 8-day experiment, is composed of two bands of energy centered around 1.5 kHz and between 4 and 5 kHz. It causes a complete reversal in the vertical directional characteristics of the mid-frequency ambient sound field between day and nighttime periods; whereas the vertical structure during the day shows a notch in the horizontal direction with levels more than 10 dB below those at higher angles, the nighttime levels in the horizontal can exceed those at other vertical angles by more than 10 dB. These nighttime sounds also are strongly anisotropic in azimuth; they appear to come mainly from a popular Southern California fishing spot where the water depths exceed 75 m. Vertical beam-to-beam coherence squared estimates suggest the chorus source region exists on spatial scales greater than the multipath interference wavelengths of this environment. Individual sounds comprising the chorus, although difficult to separate from the background din, have a fluttering, rasping character. [DOI: 10.1121/1.2338802]

PACS number(s): 43.30.Sf, 43.80.Ka, 43.30.Nb, 43.80.Ev [KGF] Pages: 1942–1955

I. INTRODUCTION

Biological sounds are an important component of the naturally occurring underwater sound field. In particular, biological choruses, the phenomenon where large numbers of animals call simultaneously over sustained periods of time, can have a remarkable impact on the sound field's levels and directional characteristics. The fact that marine animals create underwater sound has been reported in the scientific literature since at least the 1800's (see Tavolga, 1964b, for a short historical summary). In the U.S., the study of these sounds greatly intensified during World War II (e.g., Dobrin, 1947; Everest, Young, and Johnson, 1948; Johnson, 1948; Knudsen, Alford, and Emling, 1948; U.C. Div. War Research, 1946) and continued for the two decades or so afterwards (e.g., Tavolga, 1964a, 1967), motivated in part by their effects on military sonar performance. However, activity in this area of research declined probably because the U.S. Navy focus (and funding) turned to deep water problems. Several ongoing efforts have been conducted in various areas of the world to characterize the contribution of biological choruses to the ocean sound field (e.g., Cato, 1978, 1980; McCauley and Cato; Kelly, Kewley, and Burgess, 1985). With interest returning to shallow water environments, the issue of the potential impact of biological choruses on sonar operations has regained some attention. However, even greater attention is being focused on the converse problem; the potential impact of sonar operations and other types of manmade sound on the ocean's biological communities (e.g., National Research Council, 1994, 2000, 2003). The marine mammal population is of primary concern, but the impact of manmade sound on nonmammalian species also is an active area of research (e.g., Hastings, *et al.*, 1996, and references therein; McCauley, Fewtrell, and Popper, 2003, and references therein; Myrberg, 2002; Tavolga, 2002). Recently, interest has been renewed in the use of passive acoustical techniques for monitoring the locations and health of biological populations and essential fish habitats, as well as for understanding fish behavior and ecology (Rountree *et al.*, 2002; see also Bonacito *et al.*, 2002; Connaughton, Fine, and Taylor, 2002; Lobel, 1991; Luczkovich and Sprague, 2002; Mann and Lobel, 1995; Rountree *et al.*, 2003, and references therein).

Much remains to be learned about marine animal acoustics, especially the sounds of fish and invertebrates in their natural environment. Of the more than 25 000 species of fish that exist in the world today, the acoustic behavior in the wild of only a few hundred species is known to some extent (National Research Council, 2003). Even less is known about sounds from marine invertebrates, although significant results on snapping shrimp sounds recently have been obtained to supplement the insights learned during World War II (see National Research Council, 2003, for a list of references on snapping shrimp). Other marine invertebrates such as barnacles, mussels, and sea urchins are known to produce underwater sound (Fish, 1964; Cato, 1978), but little is known of the global spatial distribution and *in situ* characteristics of these sounds.

The purpose of this paper is to describe the temporal and spatial characteristics of underwater biological sounds recorded during an 8-day experiment in summer, 2002, in the Southern California Bight. The acoustic measurements were made with a large aperture, 131-element, two-dimensional

[a]Electronic mail: gld@mpl.ucsd.edu

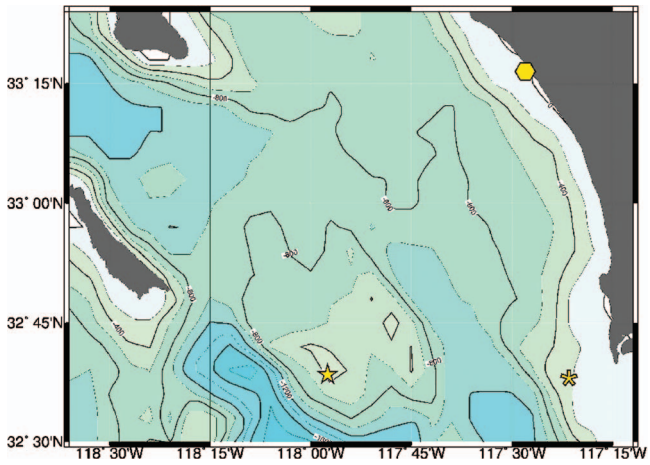


FIG. 1. Map of the Southern California offshore region showing the locations of the main experimental sites where biological choruses have been recorded. The location of the MFnoise-02b experiment discussed in this paper is marked with a star. Biological choruses at low frequencies (50–800 Hz) have been recorded at the sites plotted with an asterisk and hexagon, and are discussed elsewhere. Bathymetry contours are plotted every 200 m.

(2D) billboard hydrophone array that provided high resolution directional estimates of the sound field, particularly in the vertical direction. The next section of this paper provides a brief description of the experiment and the hydrophone array used to collect the underwater acoustic data. Section III outlines the data analysis approach, including the array processing methods used with the billboard array data. The presentation in Sec. IV is devoted to describing the properties of the biological sounds derived from processing the ocean acoustic data. Implications and speculation on the source(s) of these sounds are presented in Sec. V. This section also provides a comparison of the measurements with those made over an 8-day period in July, 1963 just to the west of the summer, 2002, site by Wenz, Calderon, and Scanlan (1965). Finally, Sec. VI summarizes the conclusions from this work.

II. DESCRIPTION OF THE EXPERIMENT AND HARDWARE

The locations of the three main experimental sites where low and midfrequency choruses have been recorded are plotted on a map of the Southern California offshore region in Fig. 1. The location of the experiment discussed in this paper is the westernmost site indicated by a star. The location of the set of Shallow Water Test Cell Experiments ("SWellEx") is plotted with an asterisk and that of the Adaptive Beach Monitoring ("ABM") experiments is shown as a hexagon. Low frequency biological choruses were recorded in these latter two sets of experiments (D'Spain *et al.* 1997).

The Mid-Frequency Ocean Noise Experiment (MFnoise-02b) occurred over a northwest/southeast-trending bathymetric ridge called the "40-Mile Bank," located about 60 km to the west of San Diego and 35 km southeast of San Clemente Island (Fig. 1). The deployment site (32° 38.5′ N, 117° 57.5′ W), is 2 km to the southeast of the "43-Fathom Spot," a popular fishing area in the Southern California Bight. The R/P FLIP, a 100-m manned spar buoy operated by the Marine Physical Laboratory (MPL), was placed in a three-point mooring in 175-m-deep water and acted as the central data recording platform during the July, 2002 experiment. A detailed bathymetry map of the site is shown in Fig. 2.

A 131-hydrophone, 2D billboard array with 12 kHz bandwidth per element (digitizing rate of 24 Ksamples/s) and interelement spacing of 0.2 m (equal to half a wavelength at 3.75 kHz) was deployed from FLIP during MFnoise-02b. A schematic of the array is shown in Fig. 3. The main part of the array is composed of 4 vertical staves with 32 elements each, all having equal 0.2-m spacing. Not shown in Fig. 3 are three additional hydrophones, two of which are placed collinear with, and 1 m to either side of the center of, the uppermost row of 4 hydrophones to provide larger horizontal aperture. The wet-end digitizing and telemetry hardware are located in the pressure cases at the top of the array.

The signal conditioning, digitizing, and telemetry hardware are located in the pressure cases at the top of the array. The signal conditioning hardware is composed of a low-electrical-noise operational amplifier that provides 60 dB of gain in the 400 Hz to 28 kHz frequency band. This amplification is required to make use of the full dynamic range of the 16-bit analog-to-digital (A/D) converters. These A/D converters are of the delta-sigma design (Candy and Temes, 1992); that is they use an initial 64 times oversampling, an anti-aliasing filter that operates on the digitized samples (Antoniou, 1979), and a decimator to reduce the output sampling rate to 24 Ksamples/s. The benefits of this delta-sigma design include the fact that the anti-aliasing filtering is performed digitally so that the properties of the filter do not change over time and that the filter has unity gain with a frequency-independent response up to nearly half the sampling frequency. The digitized data then are time-division multiplexed onto a common data bus and transmitted along a single mode fiber-optic cable by a laser diode to the recording system on R/P FLIP.

Above the pressure cases is a hydrodynamic, syntactic-foam float that provides 2224 N of positive buoyancy to keep the array vertical. The array itself is free to pivot at the single-point attachments at the top and bottom of the drawing in Fig. 3. Upon deployment, three tiltmeters were attached to the array support frame to measure heading, pitch, and roll of the array, as well as water temperature and depth. Additional details of the hardware are contained in Skinner *et al.* (2003).

The array was deployed about 12 m horizontal distance from FLIP's hull, off one of FLIP's booms. The weight at the lower end of the array was set on the ocean bottom, resulting in the geometric center of the array being located at about 170 m depth. The array recorded data from all 131 hydrophones nearly continuously over the 8-day period from 23 to 30 July, 2002.

III. DATA ANALYSIS APPROACH

A. Spectral analysis

To provide a pictorial guide to the acoustic data collected in the experiment, gray-scale plots showing the time and frequency dependence of the spectral levels ("spectro-

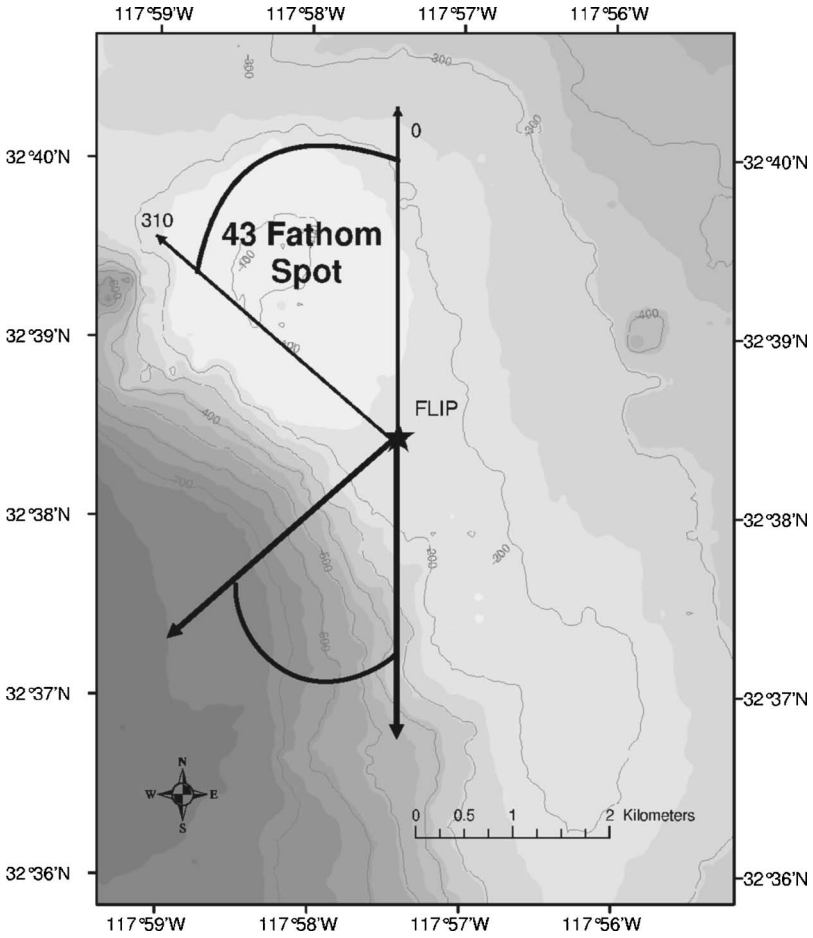


FIG. 2. Bathymetry map of the "40-Mile Bank" and the "43-Fathom Spot," with the mooring location of R/P FLIP indicated by a star. The azimuthal interval from 310° to 0° corresponds to the direction of arrival of the energy from the chorus discussed in this paper. Because of the 2D geometry of the billboard array used to make the measurements and the omnidirectionality of its individual hydrophone elements, the results of beamforming in azimuth have a reflection symmetry about the horizontal axis of the array. This symmetry is called the "left-right ambiguity." Since the array was oriented in the east-west direction when the chorus azimuthal directionality was measured, the chorus energy could equally well have come from a second, "ambiguous" azimuthal interval from 180° to 230°, as shown.

grams") are created with data from selected hydrophone elements. These spectrograms span the full 12 kHz array bandwidth and cover 2-min sequential periods of time over the complete duration of data recording. The spectrograms are created by dividing the time series into equal-length segments having 50% overlap, fast Fourier transforming (FFT) each segment after windowing with a Kaiser-Bessel window of α equal to 2.5 (the value of α determines the degree of window tapering on either side of the peak, with increasing values of α resulting in greater degrees of taper; Harris, 1978), and then properly calibrating and plotting the results. To obtain spectral density estimates, the frequency bin squared amplitudes are averaged over a sufficient number of fast Fourier transformed segments so that the 90% confidence limits typically are within 1 dB about the estimate.

B. Beamforming

Both conventional plane wave beamforming (abbreviated "CBF") and white-noise-constrained adaptive plane wave beamforming ("ABF") methods were used in the analysis of the underwater acoustic array data. White-noise-constrained beamformers are designed to provide some of the higher spatial resolution and interferer cancellation capabilities of minimum variance adaptive beamforming while maintaining a portion of the robustness of conventional beamforming (Cox, Zeskind, and Owen, 1987; Gramann, 1992). The constraint value is a free parameter that allows the beamformer to be "tuned" to the properties of a given signal and noise structure given deviations from the underlying assumptions made in the processing (e.g., that the element responses are identical and their positions are known exactly, that the propagation across the array is plane wave in nature, that correlated multipath does not exist, etc.). The white noise constraint, in effect, restricts the maximum length of the element weight vector used in weighting the individual sensor data prior to the phase-advance/delay-and-sum process. Setting the constraint value equal to $10\log(N)$, where N is the number of array elements, is equivalent to conventional beamforming, and the beamformer becomes increasingly more adaptive with decreasing constraint value. The beamforming implementations used in this paper operate in the frequency domain so that other parameters set *a priori* in the beamforming process are those associated with estimating the data cross spectral matrices (i.e., the FFT length, the percentage overlap in the time series segments used in obtaining realizations of the cross spectral matrix, and the number of realizations incoherently averaged).

Conventional beamforming was used exclusively for processing data from physical aperture in the vertical direction because of the problem of correlated multipath with adaptive beamformers (Cantoni and Gondara, 1980). Because of this correlated multipath issue for ABF, and because of limited aperture in the horizontal direction, a combined CBF/ABF technique was used with the 2D midfrequency billboard array data. The complex pressure recorded by each element at a given frequency was spatially fast Fourier trans-

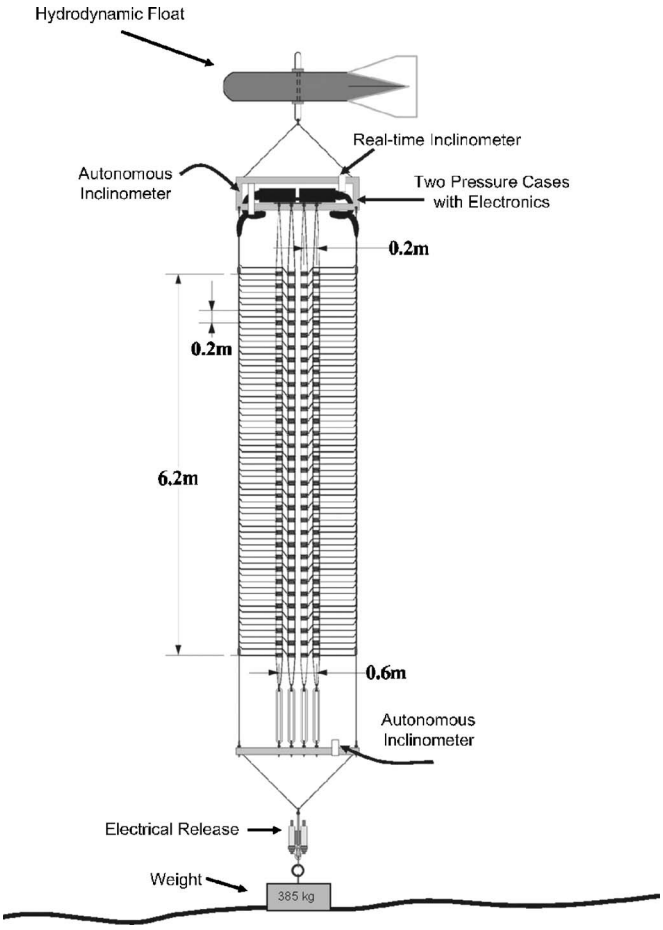


FIG. 3. Schematic of the 131-element, 2D billboard array. Only the 128 hydrophones comprising the 4 vertical staves of 32 elements each are shown.

formed across each vertical stave after spatial windowing with a Kaiser-Bessel window of α equal to 1.5 (Harris, 1978). The corresponding spatial frequency bin for each vertical stave was extracted, thereby creating an equivalent 4-element horizontal array at the array mid-depth (about 170 m) with individual elements having a vertical directivity given by the plane wave response of the single-stave FFT beamformer. White-noise-constrained ABF then was used with this horizontal array of 4 vertically directional elements to perform beamforming in azimuth. (Conventional beamforming also was used at this stage in some cases). This process was repeated for all vertical spatial frequency bins to yield the matrix of vertical/azimuthal directionality estimates. The vertical directionality results in Sec. IV are plotted as a function of fractional vertical spatial sampling frequency, where the spatial sampling frequency is the inverse of the interelement spacing of 0.2 m, to avoid the nonlinear distortion at angles away from broadside caused by converting from wave number to physical angle.

C. Beam-to-beam coherence squared estimates

Coherence squared estimates between the various pairs of vertical beam output time series are presented in Sec. IV as one indication of the possible spatially diffuse nature of the source(s) contributing to the midfrequency chorus. The coherence squared estimate at a given frequency is a statistical measure of the degree of linear relatedness between two time series at that frequency (Bendat and Piersol, 1986). For example, a coherence squared estimate of 0.5 at a specific frequency indicates that 50% of the variability in one time series can be predicted from the variability in the other time series at that frequency. A spatially concentrated source in a multipath environment such as the ocean creates arrivals in various vertical directions that are coherent with respect to one another. In contrast, the multipath arrivals from a spatially diffuse source region become mutually incoherent due to the process of summation of the multipath contributions over the various portions of the source region (see D'Spain *et al.* 2001, and the references therein). In order to put confidence limits on the coherence squared estimates, the probability density function, or cumulative distribution function, of the estimate must be known. The actual cumulative distribution function of the coherence squared estimate for two stationary Gaussian time series is given in Table 1 of Carter, Knapp, and Nuttall (1973) in terms of a finite sum of Gauss hypergeometric series. The expression simplifies to the following finite geometric series if the true population coherence squared is assumed to be zero:

$$P(|\hat{\gamma}|^2) = 1 - [1 - |\hat{\gamma}|^2]^{N-1}. \quad (1)$$

The value of N is the number of statistically independent estimates of the cross and auto spectra which are averaged together to calculate $|\hat{\gamma}|^2$. Equation (1) can be used to determine the value of $|\gamma_0|^2$ such that

$$\text{Prob}(|\hat{\gamma}| \geq |\gamma_0|^2) \leq 1 - P(|\gamma_0|^2), \quad (2)$$

assuming that the true $|\gamma|^2 = 0$. The value of $|\gamma_0|^2$ corresponding to the 95% confidence level is used as the lowermost value on the color scales of the beam-to-beam coherence squared plots presented in the next section.

IV. OBSERVATIONS OF NIGHTTIME CHORUSES IN THE MIDFREQUENCY BAND

The remarkable effect of the nighttime choruses on the vertical directionality of the midfrequency ocean noise field is illustrated in Fig. 4. It shows an increase of 20 dB in the horizontal beam levels at the array's design frequency of 3.75 kHz. The daytime profiles at two selected azimuths of arrival have notches in the horizontal direction that are nearly 15 dB lower in level than at higher angles of arrival due to the fact that the predominant noise sources (e.g., wind generated ocean surface wave activity, surface ships) occur at or near the ocean surface. However, the pattern of the nighttime profiles is completely reversed; levels in the horizontal direction exceed those at higher angles by more than 10 dB. The frequency content of the choruses is shown in the upper curve in Fig. 5. The lower curve in the plot is the horizontal beam spectral density during a quiet afternoon period for comparison. The chorus sounds are composed of two broad peaks, centered at 1.5 kHz and between 4 and 5 kHz. The peak in the upper curve at 7.5 kHz marks the frequency above which spatial aliasing occurs.

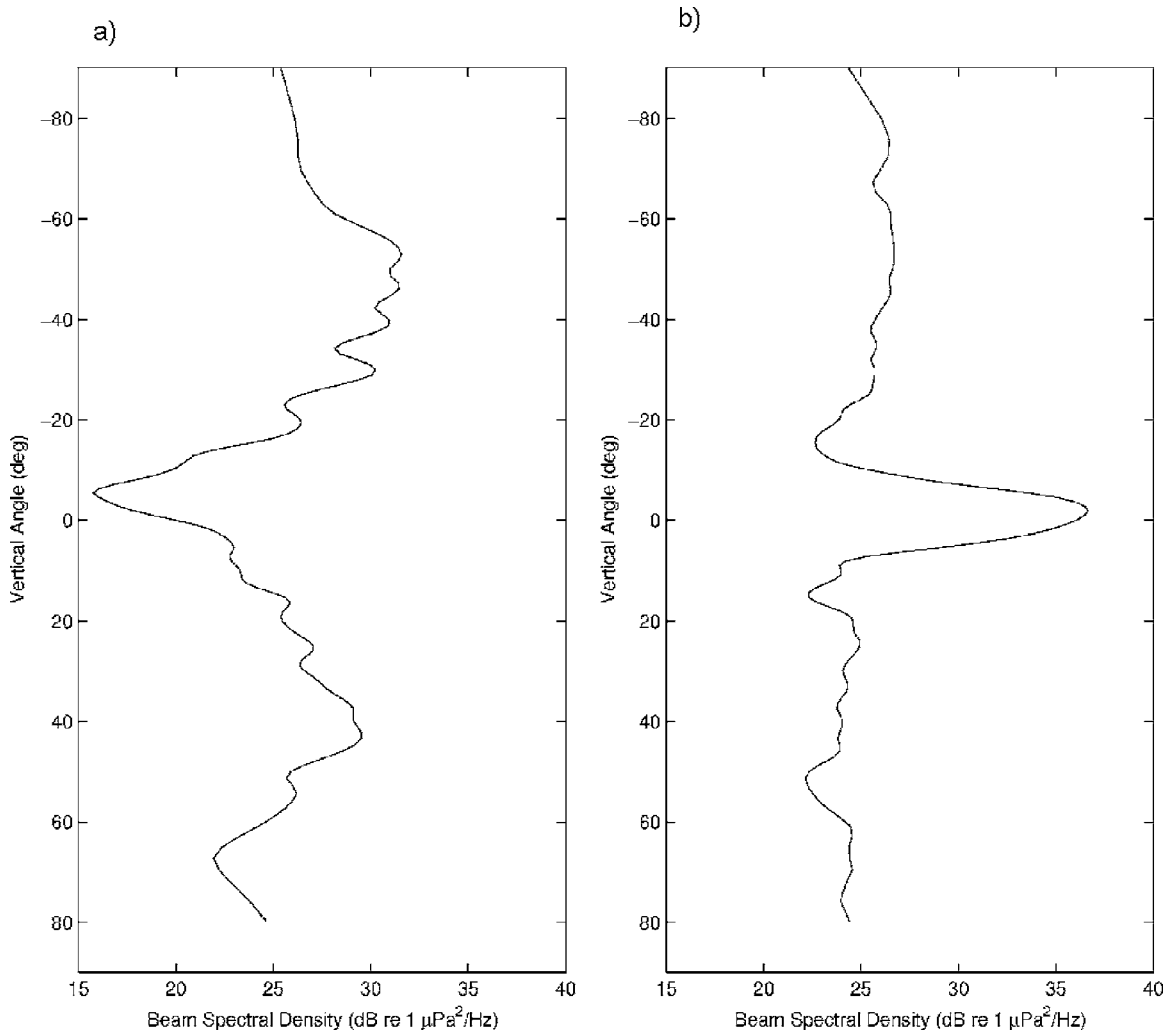


FIG. 4. A comparison of the vertical directionality of the ocean noise field at 3750 Hz for a given azimuth measured under calm conditions (wind speed less than 4 m/s) during the day (a) and at night (b) during the summer, 2002 Mid-Frequency Noise experiment ("MFnoise-02b"). Negative vertical angles are upward-looking with-90° toward the surface, +90° toward the ocean bottom, and 0° in the horizontal.

As stated in the following, individual sounds with the frequency content of the chorus are not clearly discernible in the data regardless of the type of processing employed. Rather, all the evidence from the analyses of the data suggests that the received level is the sum of the energies of the individual calls from all the animals comprising the chorus. However, if one assumes for the sake of argument that the received level is due solely to an individual animal at 2 km range, then the source level for this animal can be estimated. Calculations using the CASS/GRAB Gaussian ray bundle code (Weinberg and Keenan, 1996) show that the transmission loss in the 4–5 kHz band over a 2 km range is 65–70 dB for

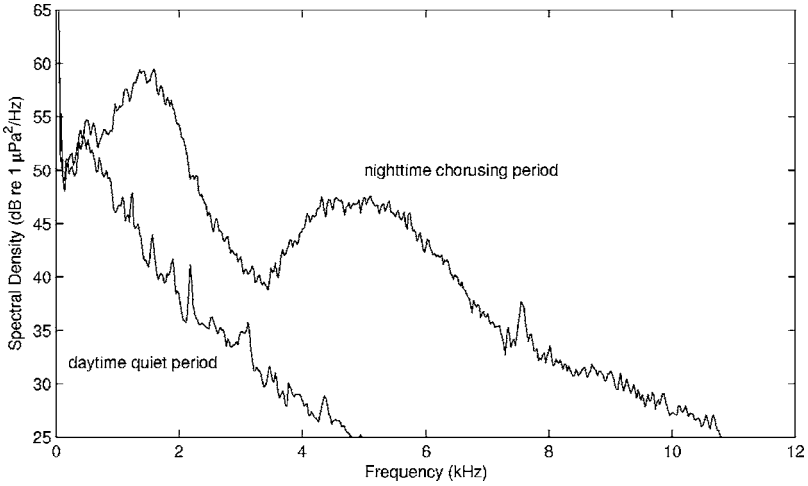


FIG. 5. Calibrated spectral densities from the horizontal beam output of a single vertical stave of the MF billboard array. The upper curve was estimated from data recorded at 2 a.m. local time (09:00 GMT), 24 July, and the lower curve is from data recorded during a low-noise period (00:46 GMT, 25 July). The high-pass-filtering effects of two resistor-capacitor filters with corner frequencies at 350 and 500 Hz, and the low-pass response of the anti-aliasing filter above 11 kHz have not been taken into account.

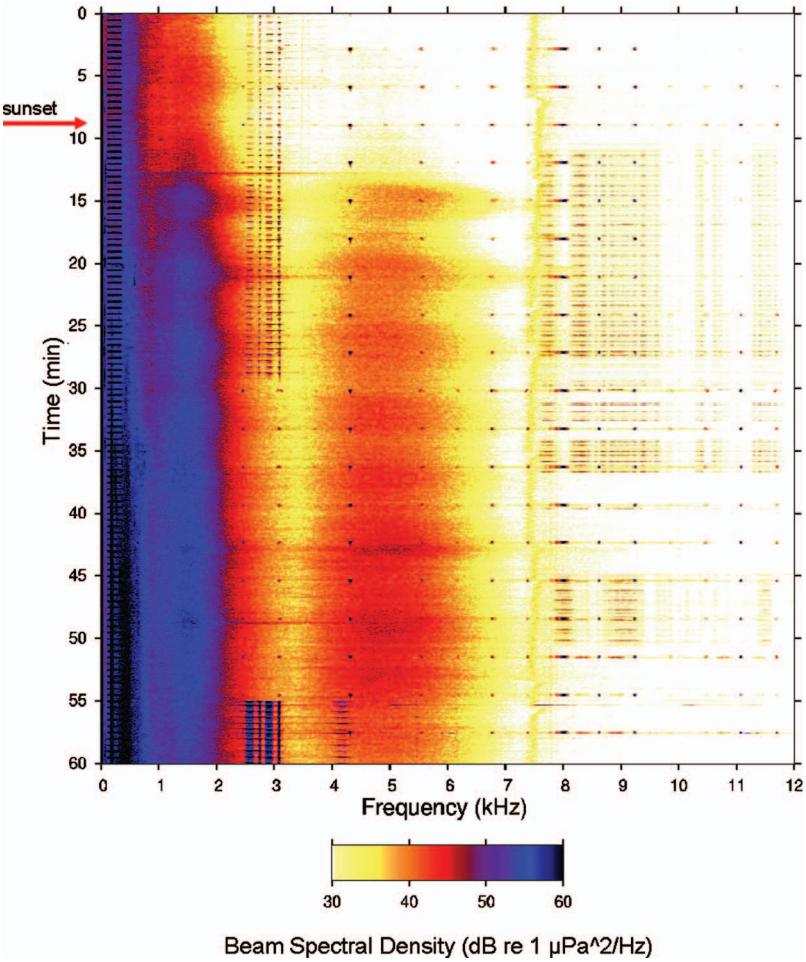


FIG. 6. Single vertical stave horizontal beam spectrogram over the 12 kHz band for a 1-h period starting at 02:50 GMT 24 July (19:50 PDT 23 July). Local sunset occurs just prior to the 10-min mark as indicated on the plot.

a shallow source (3–15 m) and is as little as 60 dB for a deeper source (50–70 m). The horizontal beam spectral density level in Fig. 5 is 45–50 dB re 1 μPa^2/Hz. Therefore, the source spectral density is 105 to 120 dB re 1 μPa^2/Hz@1 m. Integrating this spectral density over a frequency band 1 kHz wide, assuming the spectral density levels are approximately constant over this 1 kHz band, adds 30 dB (equal to 10^* log 10(1000 Hz)), resulting in an estimated source level of 135–150 dB re 1 μPa@1 m. This estimated source level is in the realistic range for an individual fish or other type of marine animal. Again, however, all the evidence indicates that the received level is associated with a chorusing behavior of a collection of animals and not just one animal.

The temporal variability of the chorus energy is presented in Sec. IV A followed by a discussion of its azimuthal directional characteristics.

A. Temporal character

To examine the temporal character of the choruses, spectrograms of a single stave's horizontal beam output were created over 1-h periods. Figure 6 presents the 1-h beam spectrogram starting about 10 min before local sunset on 23 July, 2002. The horizontal "hatch" marks in the 100–400 Hz band that occur at regular $\frac{1}{2}$-min intervals, the signals in the 2.5–3.1 kHz band, and those above 8 kHz are recordings of active signals transmitted during the experiment. In addition, the dots most clearly seen at 3.1, 4.2, 5.5, and 6.8 kHz that occur once every 3 min are generated by the submarine beacon on the hull of FLIP. Spatial aliasing exists above 7.5 kHz; the broadside beam of an equally spaced hydrophone line array is spatially aliased at frequencies above twice the array design frequency. The most prominent feature of the plot is the appearance of the "clouds" of energy around 1.5 kHz and 3.5–6.5 kHz shortly after sunset. These two clouds display identical variability in level over time, indicating that they most likely come from the same source(s). Over the first half-hour or so, the levels oscillate with a period of 5.5–6 min, becoming fairly steady at the end of the hour period. A narrow band of energy at 400 Hz begins to appear at the 35-min mark and a second, narrower band at 200 Hz starts somewhat earlier. Both of these low frequency bands are believed to also be biological in origin, although from a different source than the midfrequency energy since they have different temporal dependence.

The low frequency bands do not appear later in the evening, and are absent in the hour-long horizontal beam spectrogram starting a half-hour before local midnight; Fig. 7. However, the energy centered at 1.5 kHz and in the interval 3.5–6.5 kHz still is the prominent feature of the plot. In fact, the levels of these midfrequency sounds were higher during this time interval than at any other time over the 8

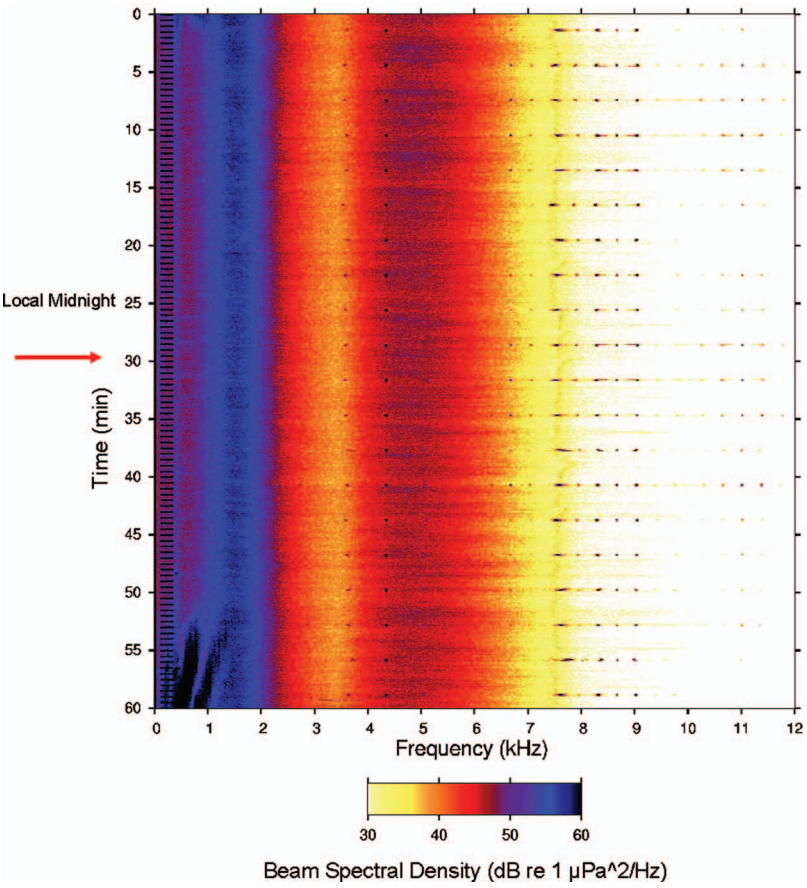


FIG. 7. One-hour, single stave, horizontal beam spectrogram starting at 06:30 GMT, 24 July (23:30 PDT 23 July). The time of local midnight is indicated on the plot.

-day period of data recording. The broadband interference pattern below 1 kHz of an approaching surface ship can be seen in the final 5 min of the plot.

As sunrise approaches, the midfrequency energy disappears as shown by the hour-long spectrogram starting at 12:02 GMT (5:02 PDT) in Fig. 8. Interference from a surface ship in the periods between 10 and 22 min and after 35 min cover the detailed temporal behavior of the biological energy as it decreases in level. However, it clearly has decayed to near-background levels by the end of the 1-h period, just after sunrise.

The variations in chorus levels in the 5 kHz bin over the nights of 24 July and 25 July are shown in Fig. 9. The 5.5–6 min oscillatory nature of the levels at the beginning of the chorusing is clearly evident between hours 3 and 4 on the 24th and to some extent on the 25th. The large-amplitude points just after hour 4 on the 24th are due to active signal transmissions as are the spikes between 07:00 and 09:00 GMT (hours 7 and 9). The spikes most evident in the last hour of the upper curve are caused by transmissions from the submarine beacon on FLIP. Gaps in the lower curve around 06:00, just prior to 07:30, and after 11:30 are due to temporary stoppages in the data recording system. The chorus levels themselves show a steady increase to maximum levels in the hour just before local midnight between 06:00 and 07:00 GMT, followed by a slower decline in levels over the remainder of the nighttime period. The regular nature of this variation is punctuated by half to one-hour periods where the levels increase (e.g., just before hour 10, just after hour 11, between hours 12 and 13 in the upper curve) or decrease (around the 4-h mark in both curves). This type of diurnal dependence—absent during the day and prevalent much of the nighttime period—is a common feature of the biological choruses identified in the ocean acoustic measurements in the Southern California Bight (D'Spain *et al.*, 1997; Johnson, 1948; Fish, 1964).

Horizontal beam spectrograms of 1- and 2-min duration were created to search for individual calls in these biological choruses. No individual signals are apparent, except possibly in one or two 2-min periods. By prewhitening the spectra (i.e., normalizing by an averaged background spectrum) and plotting the results with a small dynamic range after band-filtering between 500 Hz and 7.4 kHz (re Fig. 10), individual signatures predominantly in the 4–5 kHz bands become visible. These signals, when played back through a speaker, have a fluttering, rasping character.

The lack of visually distinguishable individual calls in the beam spectrograms may be the result of a combination of significant distance between the source region and receiving array and the creation of the sounds by large numbers of individuals. Support for the spatially diffuse nature of the sources is provided by coherence squared estimates between the vertical beams. Figure 11 shows a comparison between the vertical beam-to-beam coherence squared estimates at 4470 Hz obtained during a period when the midfrequency biological chorus was the dominant received sound (a) and at the time that a low-level tone at this frequency was transmitted from a controlled acoustic source (an electroacoustic projector) at 52 m depth and 2.0 km range (the same distance as the 43-Fathom Spot, but in the opposite direction). Both the

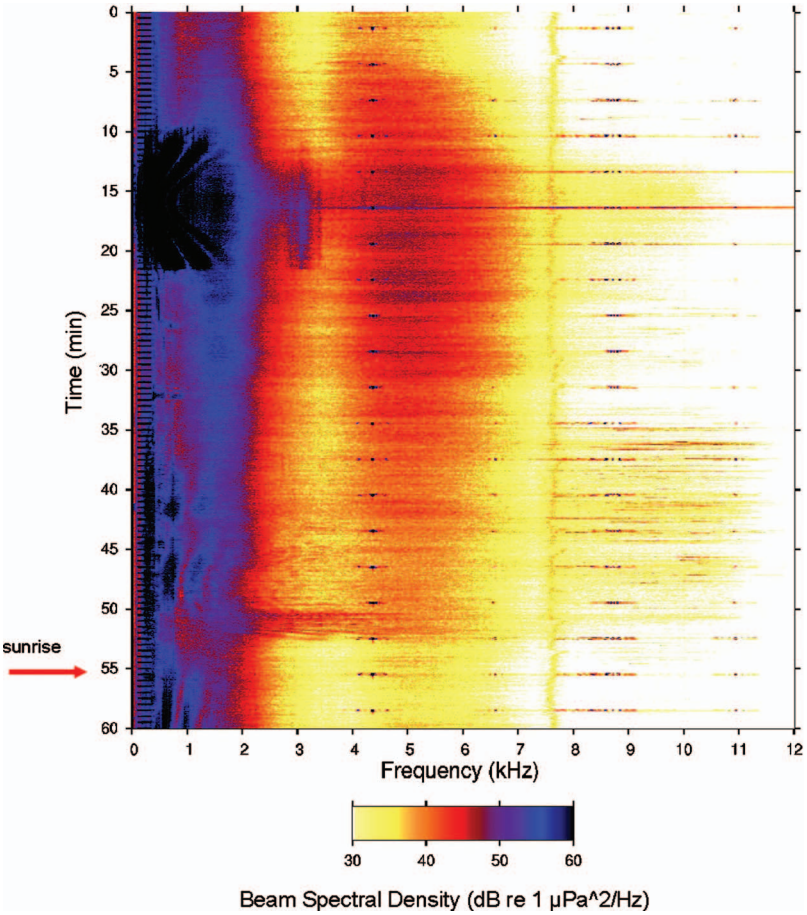


FIG. 8. One-hour, single stave, horizontal beam spectrogram starting at 12:02 GMT (05:02 PDT 24 July) with the time of local sunrise occurring around the 55-min mark on the vertical axis.

received chorus levels and the received levels from the controlled source are much greater than the background (wind-generated) noise in the 4470 Hz bin. The statistically significant coherence squared values in the off-diagonal bins in panel (b), caused by correlated multipath arrivals, are indicative of the spatially concentrated nature of the controlled source. None of the off-diagonal coherence squared estimates in panel (a) during the period of high-level chorusing of Fig. 11 are statistically significant because the range integration resulting from a spatially distributed source region effectively decorrelates the multipath components (D'Spain *et al.*, 2001).

The midfrequency chorus sounds were recorded continuously over several-hour periods on two consecutive nights (24 and 25 July) of the 8-day experiment in 2002. Single element spectrograms over the entire 8-day experiment showed no evidence of the chorus from any other night's data except for a half-hour period just before sunrise on 27 July. Horizontal beam spectrograms about this time period (created in the same way as those in Figs. 6–8) indicate that the chorusing was detectable for about 3 h, but was quite weak except during this half-hour period. Likewise, at other nighttime periods during the 8-day experiment, the array gain provided by the billboard array permits the chorus

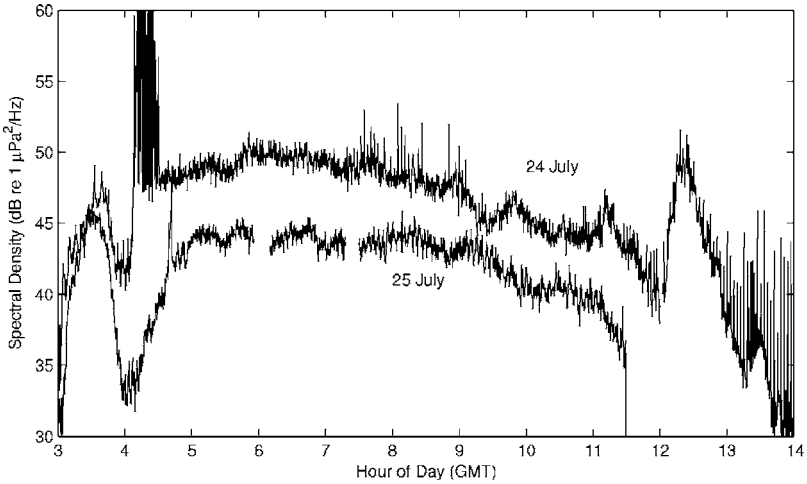


FIG. 9. Time series of the horizontal beam spectral density level at 5 kHz over two consecutive nights. The upper curve is for 24 July, covering the 11-h period from 03:00 to 14:00 GMT (20:00 to 07:00 PDT), whereas the lower curve covers the 8.5-h period on 25 July from 03:00 to 11:28 GMT.

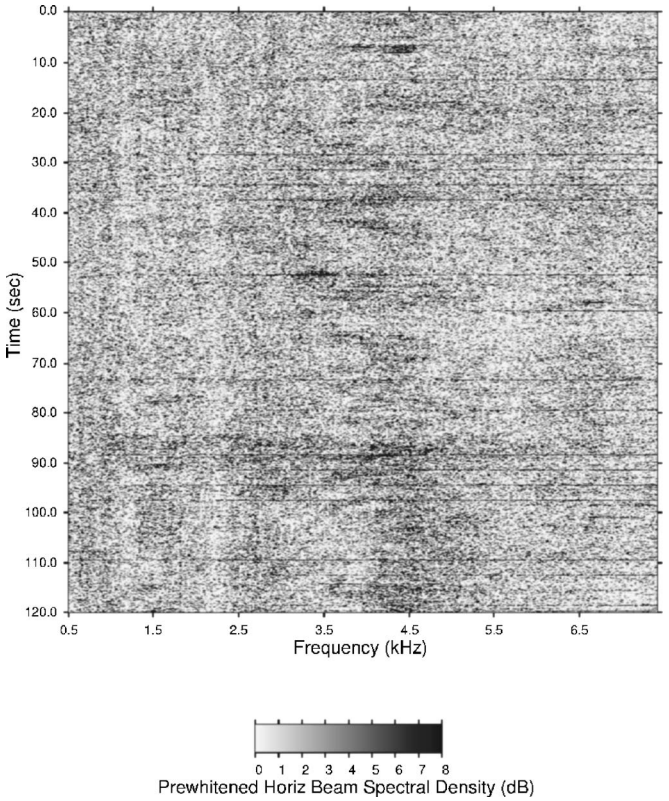


FIG. 10. Two minute horizontal beam spectrogram from 500 Hz to 7.4 kHz starting at 11:52 GMT (04:52 PDT), 24 July, after normalizing by an averaged and smoothed background spectral density curve.

to be detected in the horizontal direction at certain azimuths. However, it clearly is not detectable on the horizontal beams in most other nighttime intervals. This ephemeral nature of the nighttime chorus probably is due to temporal variability in the level of sound production and/or as the spatial mobility of the source region(s).

B. Azimuthal directionality

The four vertical staves of the MF billboard array (Fig. 3) provide only modest horizontal aperture. In addition, the planar geometry of the array creates a left-right ambiguity in the horizontal direction. Nonetheless, this horizontal aperture has proven to be extremely valuable in the analysis of the wind-driven ocean noise field (reported elsewhere) as well as the biological sounds. Figure 12 shows a plot of the beamformer output spectral density at 1500 Hz as a function of vertical angle and azimuth. The combined CBF in the vertical and ABF in the horizontal (Sec. III), with a white-noise constraint 3 dB down from conventional, was used to create the plot. The arriving energy is concentrated in one interval of azimuths relative to endfire (endfire is defined as the direction where a source is collinear with each of the 32 4-element horizontal line subarrays that make up the 2D billboard array). Using the heading data from the tiltmeters mounted on the array (Fig. 3), one of two possible physical azimuthal intervals is indicated in Fig. 2 by the radial lines at 0° and 310°. As Fig. 2 shows, this azimuthal interval encompasses the 43-Fathom Spot. Therefore, with the caveat that

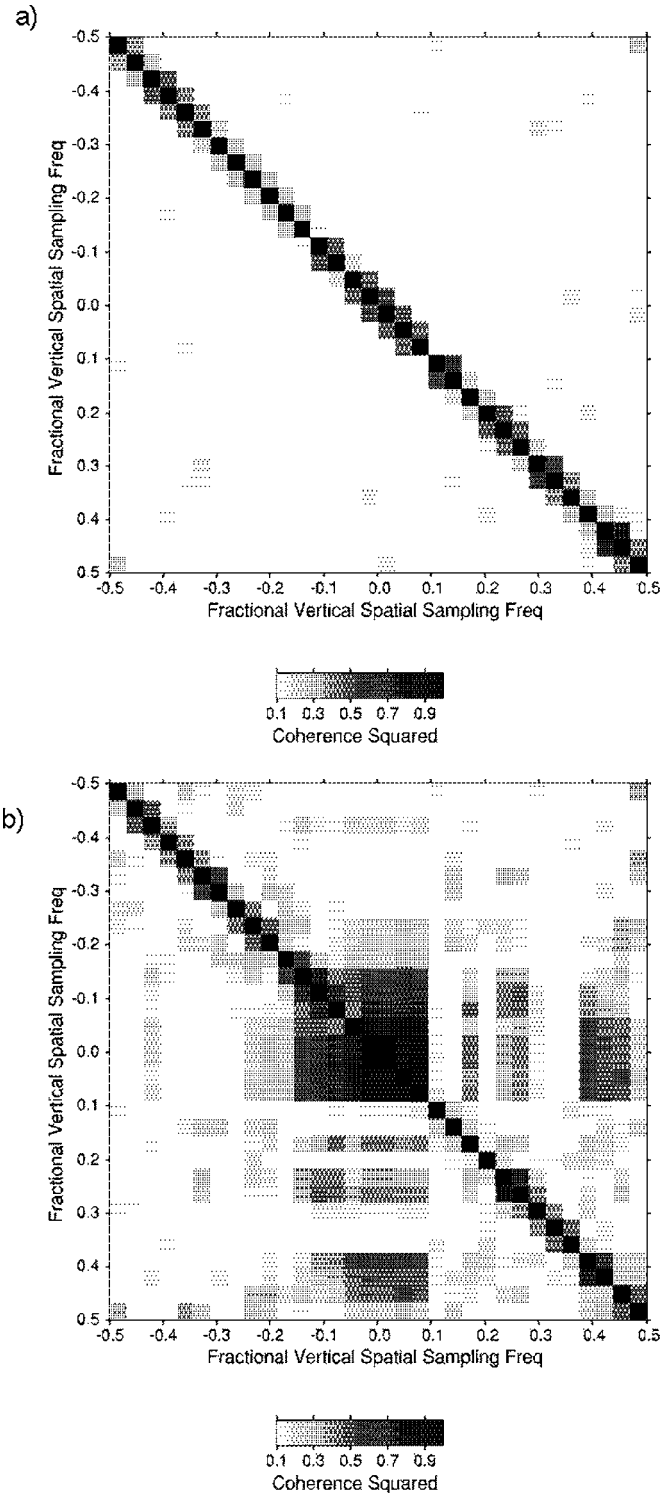


FIG. 11. Matrices of the vertical-beam-to-vertical-beam coherence squared estimates at 4470 Hz from a single vertical stave's data at two different time periods. The data in (a) were recorded at 06:32 GMT (23:30 PDT), 24 July, when the mid-frequency biological chorus was the dominant sound, and those in (b) were recorded 4 days later at 03:00 GMT 27 July (20:00 PDT 26 July) when a controlled source at 52 m depth and 2 km range was transmitting a low-level tone. The lowermost value of the gray-scale in the plots is the 95% confidence level for the coherence squared estimates assuming that the true population coherence squared is zero (re Sec. II).

the billboard array suffers from a left/right ambiguity, the biological sounds are associated with the 43-Fathom Spot.

A frequency of 1500 Hz was used to create Fig. 12 because it is the frequency of the highest-level peak in the

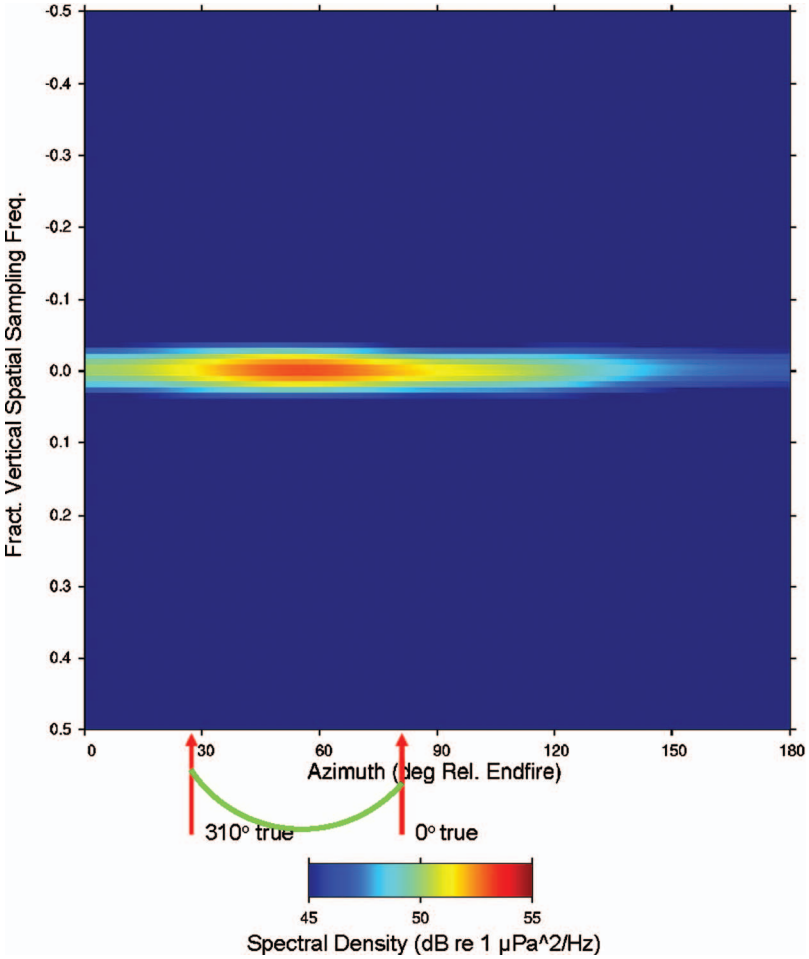


FIG. 12. Beamformer output spectral density at 1500 Hz as a function of fractional vertical spatial sampling frequency and azimuth relative to array endfire, based on 5.5 s of data starting at 09:00 GMT (02:00 PDT), 24 July, the same segment of time series as that used for the right-hand panel in Fig. 4.

biological spectral density in Fig. 5. Finer horizontal resolution can be obtained by beamforming on the energy at higher frequencies. Therefore, processing was performed at each of the frequency bins in the third-octave band centered at 4500 Hz, i.e., from 4000 to 5050 Hz, and the resulting beamforming outputs were inherently summed. These frequencies are above the design frequency of the array so that spatial aliasing occurs. However, as mentioned earlier in this section, the broadside beam for an equally spaced hydrophone line array is not contaminated by aliasing until the frequency is increased to almost twice the array design frequency. Therefore, the horizontal beam in the vertical direction is not affected by spatial aliasing. In contrast, the beams in azimuth do suffer from spatial aliasing since the array is not necessarily aligned broadside to the biological source azimuth. By summing across frequency, the main lobe energy, which is fixed in direction with respect to changes in frequency, is enhanced whereas the grating lobe and sidelobe energy is smeared since its location in azimuth changes with frequency. Figure 13 shows the results of this processing for the horizontal beam in the vertical direction as a function of azimuth and of time over a 50-min time period starting at 06:30 GMT, 24 July (the same start time as Fig. 7). Most of the energy on the right side of the plot at azimuth angles relative to endfire greater than 130° is due to spatial aliasing of the energy arriving in the angle interval of 30°–60°. Therefore, the plot indicates two distinct source regions, one around 90° azimuth relative to endfire and one between 30° and 60° azimuth relative to endfire, for this biological energy over this time period. Over this time period, 0° relative to endfire corresponds to 280° true and 180° relative to endfire to 100° true. Therefore, broadside at 90° relative azimuth corresponding to the arrival direction from one source region is either 10° or 190° true, due to the left/right ambiguity of the 2D array. The other arrival direction of 30°–60° relative to endfire corresponds either to 310° to 340° true, or 220° to 250° true. Therefore, the true azimuths of arrival of both of the bands of energy during this time period are associated with the direction towards the 43-Fathom Spot, again with the caveat of the left/right ambiguity of the planar array.

Beamforming in azimuth at other time periods indicate that the chorus energy does not always arrive from the direction of the 43 Fathom Spot. Rather, it can come from directions to the east or west of that location.

V. DISCUSSION

As presented in Sec. IV, the azimuths of arrival of the midfrequency biological chorus correspond to the direction toward the 43-Fathom Spot to the northwest of the FLIP site during the period when the received chorus levels reached their maximum values in the 8-day experiment. Over this time period, the chorus emanates from two separate source regions (Fig. 13). The energy arrives in vertical angles concentrated about the horizontal either because the source(s) themselves are located at significant depth in the water col-

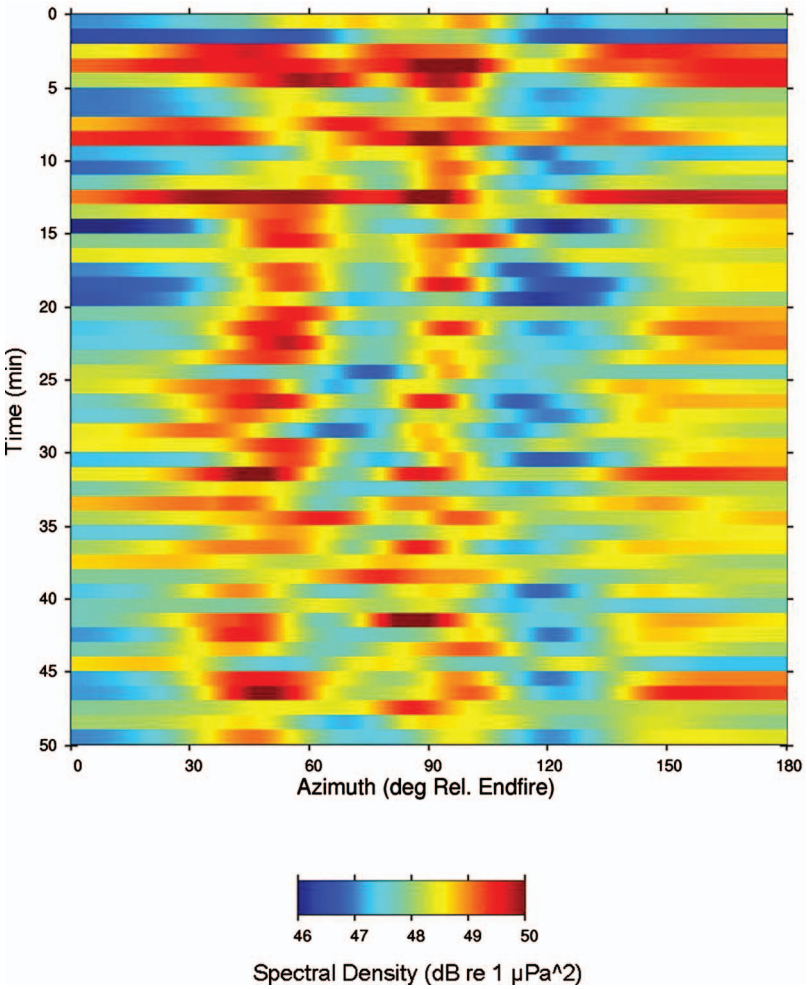


FIG. 13. Azimuthal directionality of the vertical beam energy at and near the horizontal direction as a function of time over a 50-min period starting at 06:30 GMT 24 July. The beamforming output amplitude squared was integrated over a 1/3-octave band centered at 4500 Hz.

umn (providing strong evidence of their biological nature) or because of downslope conversion of higher angle energy due to interaction with the southeast-facing flank of the bathymetry, or both. The vertical beam-to-beam coherence squared results suggest that the chorus source region is spatially diffuse, i.e., that it exists on spatial scales significantly greater than the multipath interference wavelengths.

During the 8-day experiment, the biological chorus was predominant on two consecutive nights, 24 and 25 July, 2002. This time period corresponds to the time of a full moon. Previous work on biological sounds has shown a relationship between the level of chorusing activity in some species and the level of moonlight (Fish, 1964). Therefore, changes in moonlight levels may be related to the changes in the chorusing observed in MFnoise-02b. Alternatively, changes in lunar cycle result in changes in tidally driven ocean currents which may have caused the observed changes in biological activity. Ocean currents in the lower third of the water column, as measured by a downward-looking acoustic Doppler current profiler (ADCP), were particularly strong during the full moon period, reaching speeds up to 0.3 m/s (Terrill, 2003). In any case, a significantly longer time series of underwater acoustic measurements is required to determine if an actual association exists between lunar cycle and the midfrequency chorus levels or is coincidental.

The 2002 Mid-Frequency Noise experiment was not designed to record biological sounds. In fact, identification of the midfrequency chorus as biological in nature occurred only during the analysis of the data after the experiments. Therefore, no information was collected during these experiments specifically to aid in determining the species creating the sounds. However, as part of another program largely funded by the National Geographic Society, scientists from the Scripps Institution of Oceanography and elsewhere deployed a bottom-tethered instrument containing a camera and strobe light on the 40-Mile Bank. The 18-h period of deployment encompassed the nighttime hours of 24 July, when the midfrequency chorus levels were highest. Hardy *et al.* (2002) provides one of the images collected by the digital camera system (their Fig. 6). The bottom-dwelling animals contained in the picture include red crabs (*Galathea californica*), Pacific hagfish (*Eptatretus stoutii*), a pink urchin (*Allocentrotus fragilis*), a short-spined combfish (*Zaniolepis frenatus*), and an amphipod swarm (identified by E. Kisfaludy, Scripps Institution of Oceanography, 2002). It is unknown whether or not any of these species produce biological choruses.

The 43-Fathom Spot 2 km to the northwest of FLIP's location is a popular Southern California fishing site. The catch there includes bluefin and bigeye tuna, yellowtail, marlin, albacore, skipjack, and dorado (Eschmeyer, Herald, and Hammann, 1983). Fishing is best during times of strong easterly-flowing currents. The diet of these game fish includes small fishes (e.g., anchovies) and squid. Additionally, more than twenty-five species of rockfish have been found in

the area of the 43-Fathom Spot (Mary Yoklavich, NOAA Fisheries, 2005) at least one of which, the bocaccio, is known to produce sound. The frequency content of the secondary broad spectral peak of the midfrequency chorus (re Fig. 5) is quite high for fish-generated sounds (Fish and Mowbray, 1970). High level sounds are created by fish through amplication by their swim bladders (Tavolga, 1964b) and the resonance frequencies of swim bladders typically are in the few hundred hertz range, an order of magnitude lower than the center frequency of the 4–5 kHz energy in Fig. 5. However, other mechanisms such as caudal fin cavitation (M. Ball, Scripps Aquarium, 2003) or stridulation (Tavolga, 1964b) may account for the midfrequency nature of the chorus.

The 4–5 kHz spectral peak in Fig. 5 actually is more representative of those measured near colonies of snapping shrimp (University of California Division of War Research, 1946). However, snapping shrimp colonies are found in waters less than 60 m in depth, which are located more than 30 km from the site (Fig. 1). In addition, snapping shrimp sounds show a smaller degree of diurnal variability (Everest, Young, and Johnson, 1948; University of California Division of War Research, 1946) and tend to vary little on weekly and monthly time scales because of the sessile nature of the colonies (Everest, Young, and Johnson, 1948; Fish, 1964). Other types of invertebrates are known to generate sound (Fish, 1964; Cato, 1978). In fact, the spectrum of New Zealand evening choruses associated with the sea urchin *Evechinus chloroticus* (from R. I. Tait as reported in Fish, 1964) shows a peak at 1.5 kHz, exactly in correspondence with the lower frequency portion of the midfrequency chorus in the upper curve in Fig. 5. However, a spectral energy peak in the 4–5 kHz band is not present in Tait's data. As mentioned earlier, it is unknown whether or not the pink urchin seen in the photograph of the ocean bottom of the 40-Mile Bank creates a midfrequency chorus. Additional invertebrates identified in benthic community surveys at the 43-Fathom Spot include numerous sea stars, species of squid and octopus, several sea urchins, and many species of crab (Mary Yoklavich, NOAA Fisheries, 2005), although none of these species are confirmed sound producers.

An underwater acoustic experiment with the same 8 -day duration and at almost the same time of year as the MFnoise-02b experiment (16–23 July versus 23–30 July) was conducted on the west side of San Clemente Island (around 32° 55′ N, 118° 34′ W in Fig. 1) in 1963 (Wenz, Calderon, and Scanlan, 1965). Two omnidirectional hydrophones were deployed near the ocean bottom in water depths of 110 and 823 m. Results from averaging by eye the 1/3-octave band analog filter outputs (with "…some discrimination against large short-term transients"), measured for 15 min at the beginning of each hour over the 8-day recording period are presented by Wenz, Calderon, and Scanlan (1965). Their results indicated large diurnal variability at 1/3-octave center frequencies from 40 Hz to 3.15 kHz. Two biological choruses, one in the 100–250 Hz band and a second from 160 to 1000 Hz, were identified by these authors, and then further subdivided into three low frequency choruses by Thompson (1965). This work by Naval Electronics Laboratory personnel represents a significant addition to the characterization of the biological contributions to the Southern California Bight ocean acoustic field, initiated during World War II at the University of California Division of War Research (U.C. Div. War Research, 1946; Everest, Young, and Johnson, 1948; Johnson, 1948). Although a chorus at midfrequencies is not identified, the 1/3-octave band spectral levels at center frequencies of 2 and 3.15 kHz show diurnal-scale changes of up to 10 dB. Much of this variability may be associated with diurnal changes in wind speed, as pointed out by the authors. However, on some days the variations in level have a bimodal character, possibly suggesting that more than one process contributed to the daily fluctuations in ambient noise levels. The 16–23 July, 1963 period did not include a full moon, but did contain a new moon when tidally driven current flow also can be high. In any case, measurements with omnidirectional hydrophones may not allow various naturally occurring processes with similar scales of temporal variability to be differentiated.

Future studies would ideally include the ability to identify sound-producing organisms at the time that chorusing is recorded. One example would be to deploy a hydrophone array from FLIP moored near the 43-Fathom Bank and process the acoustic data in near real time. When the chorusing sounds are identified, an autonomous underwater vehicle (AUV) would be deployed to the area where the chorus sounds are coming from. The AUV would simultaneously record acoustic data and video data of the organisms in the region. Another approach might be to capture some of the animals in the area. Sound recordings of these captive animals could be made. Having examples of biological sounds produced by organisms found in the study area may lead to the identification of specific sounds recorded in the 2002 experiment.

Individual calls with the spectral content of the biological choruses cannot be seen in the beam spectrograms. However, other individual sounds of apparent biological origin are visible and can be heard clearly when the time series is played back through a loudspeaker. For example, sounds from barking sea lions (sea lions vocalize underwater as well as in air (Schusterman *et al.*, 2000) were present on most of the days of data recording. (California sea lions, *Zalophus californianus*, are the most prevalent pinniped off the Southern California coast; Carretta *et al.* 2002). In addition, calls that sound somewhat like a combination of a gobbling turkey and a squealing pig occurred repeatedly over 1- to-2-min periods on a few of the nights. The source of these sounds, coming from much shallower depths than the array deployment depth of 170 m, is unknown.

VI. CONCLUSIONS

A biological chorus with energy in two broad spectral peaks centered around 1.5 and 5 kHz occurred predominantly over two consecutive nights corresponding to a full moon period during an 8-day experiment in July, 2002. The large-aperture, multi-element aspect of the 131-hydrophone 2D billboard array deployed during the experiment provides increased detection capability of this midfrequency chorus in

the presence of interferers (mostly ships) and background noise as well as high resolution estimates of its directionality. The azimuths of arrival of the chorus energy correspond mainly to the direction toward the 43-Fathom Spot, a popular southern California fishing spot 2 km to the northwest of the experiment site. Beamforming results at the highest possible frequencies show that the chorus emanates from two separate source regions during the time period when the received chorus levels were at their highest values. The spatially diffuse nature of the source region is supported by the lack of statistically significant coherence between the various vertical beam outputs during the chorus-dominated periods. Individual sounds that contribute to the chorus are difficult to isolate from the background noise of the chorus itself, suggesting that the source region(s) is located at some distance from the experiment site. However, for one or two short periods, individual signals do appear to be barely distinguishable (either visually in beam spectrograms or audibly when the time series are played through a loudspeaker); they have a fluttering, rasping character.

The sound producers could be marine mammals, a number of fishes, several invertebrates, or a combination thereof. At this time, the biological species contributing to the chorusing are unknown.

In the vertical direction, the occurrence of the chorus results in an extraordinary reversal in the directional characteristics of the midfrequency ambient sound field between day and nighttime periods. That is, during the day, background noise levels near the horizontal are nearly 15 dB lower than at higher angles of arrival because the dominant midfrequency noise sources (surface ships and wind-generated ocean surface wave activity) are located at, or near, the ocean surface and because the summertime water conditions during the experiment were such that the sound speed profile was strongly downward-refracting. During the nighttime periods when the biological chorus sounds are present, the levels in the horizontal can exceed those at other vertical angles of arrival by more than 10 dB.

These results have significant implications for midfrequency sonar system performance. In particular, for systems with large vertical aperture, a modest amount of horizontal aperture permits discrimination against these kinds of interfering biological noise.

ACKNOWLEDGMENTS

Jeff Skinner and Greg Edmonds at MPL designed the midfrequency billboard array wet-end sampling system and dry-end data acquisition system, and Dick Harriss of MPL developed the array's mechanical structure and deployment strategy. Bill Hodgkiss helped with overall coordination of the program. Pam Scott and Dave Ensberg at MPL helped with the array construction and at-sea data collection. Dave also performed much of the data archival and quick-look analyses. The assistance of Capt. Bill Gaines at MPL and the crew of the R/P FLIP is greatly appreciated. Katherine Kim also played an important role in conducting the experiment. Thanks to Kevin Hardy and Art Yayanos at Scripps Institution of Oceanography for providing us with the underwater photographs from their deep ocean digital camera system. This work was supported by the Office of Naval Research, Code 321(US).

Antoniou, A. (**1979**). *Digital Filters: Analysis and Design* (McGraw-Hill, New York), pp. 124–128.

Ball, M. (private communication, **2003**).

Bendat, J. S., and Piersol, A. G. (**1986**). *Random Data: Analysis and Measurement Procedures* (Wiley, New York), 2nd ed..

Bonacito, C., Costantini, M., Picciulin, M., Ferrero, E. A., and Hawkins, A. D. (**2002**). "Passive hydrophone census of *Sciaena umbra* (Sciaenidae) in the Gulf of Trieste (Northern Adriatic Sea, Italy)," Bioacoustics 12, 292–294.

Candy, J. C., and Temes, G. C. (**1992**). *Oversampling Methods for A/D D/A Conversion, Oversampling Delta-Sigma Converters* (IEEE Press, Englewood Cliffs, NJ).

Cantoni, A., and Godara, L. C. (**1980**). "Resolving the directions of sources in a correlated field incident on an array," J. Acoust. Soc. Am. **67**, 1247–1255.

Carretta, J. V., Muto, M. M., Barlow, J., Baker, J., Forney, K. A., and Lowry, M. (**2002**). "U.S. Pacific marine mammal stock assessments: 2002," NOAA Tech. Memo. NOAA-TM-NMFS-SWFSC-346, Southwest Fisheries Science Center, NMFS/NOAA, La Jolla, CA, pp. 290.

Carter, G. C., Knapp, C. H., and Nuttall, A. H. (**1973**). "Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing," IEEE Trans. Audio Electroacoust. **AU-21**(4), 337–344.

Cato, D. H. (**1978**). "Marine biological choruses observed in tropical waters near Australia," J. Acoust. Soc. Am. **64**(3), 736–743.

Cato, D. H. (**1980**). "Some unusual sounds of apparent biological origin responsible for sustained background noise in the Timor Sea," J. Acoust. Soc. Am. **68**(4), 1056–1060.

Connaughton, M. A., Fine, M. L., and Taylor, M. H. (**2002**). "Use of sound for localization of spawning weakfish in Delaware Bay (USA) and effects of fish size, temperature, and season on sound parameters," Bioacoustics 12, 294–296.

Cox, H., Zeskind, R. M., and Owen, M. M. (**1987**). "Robust adaptive beamforming," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**(10), 1365–1376.

D'Spain, G. L., Berger, L. P., Kuperman, W. A., and Hodgkiss, W. S. (**1997**). "Summer night sounds by fish in shallow water," in *Shallow Water Acoustics*, edited by R. Zhang and J. Zhou (China Ocean Press, Beijing), pp. 379–384.

D'Spain, G. L., Berger, L. P., Kuperman, W. A., Stevens, J. L., and Baker, G. E. (**2001**). "Normal mode composition of earthquake T phases," Special issue on Hydroacoustics of the CTBT in Pure and Appl. Geophys. **158**(3), 475–512.

Dobrin, M. B. (**1947**). "Measurements of underwater noise produced by marine life," Science, **105**, 19–23.

Eschmeyer, W. N., Herald, O. W., and Hammann H. (**1983**). *Peterson Field Guides: A Field Guide to Pacific Coast Fishes North America* (Houghton Mifflin, Boston), pp. 336.

Everest, F. A., Young, R. W., and Johnson, M. W. (**1948**). "Acoustical characteristics of noise produced by snapping shrimp," J. Acoust. Soc. Am. **20**(2), 137–142.

Fish, M. P. (**1964**). "Biological sources of sustained ambient sea noise," in *Marine Bio-Acoustics*, edited by W. Tavolga (Pergamon, New York), pp. 175–194.

Fish, M. P., and Mowbray, W. (**1970**). *Sounds of Western North Atlantic Fishes* (The Johns Hopkins Press, Baltimore), pp. 207.

Gramann, R. A. (**1992**). "ABF algorithms implemented at ARL-UT," ARL Tech. Letter ARL-TL-EV-92-31, Applied Research Laboratories, The University of Texas at Austin, Austin, TX.

Hardy, K., Olsson, M., Yayanos, A. A., Prsha, J., and Hagey, W. (**2002**). "Deep ocean visualization experimenter (DOVE): Low-cost 10 km camera and instrument platform," IEEE Oceans 2002 Conference, pp. 1234–1238.

Harris, F. J. (**1978**). "On the use of windows for harmonic analysis with the discrete Fourier transform," Proc. IEEE **66**(1), 51–83.

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**(3), 1759–1766.

Johnson, M. W. (**1948**). "Sound as a tool in marine ecology, from data on

biological noises and the deep scattering layer," J. Mar. Res. **7**(3), 443–458.
Kelly, L. J., Kewley, D. J., and Burgess, A. S. (**1985**). "A biological chorus in deep water northwest of Australia," J. Acoust. Soc. Am. **77**(2), 508–511.
Knudsen, V. O., Alford, R. S., and Emling, J. W. (**1948**). "Underwater ambient noise," J. Mar. Res. **7**, 410–429.
Lobel, P. S. (**1991**). "Sounds produced by spawning fishes," Environ. Biol. Fish. **33**, 351–358.
Luczkovich, J. J., and Sprague, M. W. (**2002**). "Using passive acoustics to monitor estuarine fish populations," Bioacoustics 12, 289–291.
Mann, D. A., and Lobel, P. S. (**1995**). "Passive acoustic detection of sounds produced by the damselfish, *Dascyllus albisella* (Pomacentridae)," Bioacoustics 6, 199–213.
McCauley, R. D., and Cato, D. H. (**2000**). "Patterns of fish calling in a nearshore environment in the Great Barrier Reef," Philos. Trans. R. Soc. London, Ser. B **355**(1401), 1289–1293.
McCauley, R. D., Fewtrell, J., and Popper, A. N. (**2003**). "High intensity anthropogenic sound damages fish ears," J. Acoust. Soc. Am. **113**, 1–5.
Myrberg, A. A. Jr. (**2002**). "Fish bioacoustics and behavior," Bioacoustics 12, 107–109.
National Research Council (**1994**). *Low-Frequency Sound and Marine Mammals: Current Knowledge and Research Needs* (National Academy Press, Washington, DC), pp. 75.
National Research Council (**2000**). *Marine Mammals and Low-Frequency Sound* (National Academy Press, Washington, DC), pp. 146.
National Research Council (**2003**). *Ocean Noise and Marine Mammals* (National Academy Press, Washington, DC), pp. 192.
Roundtree, R. A., Goudey, C., Hawkins, T., Luczkovich, J., and Mann, D. (**2003**). "Listening to Fish: Passive acoustic applications in marine fisheries," Sea Grant Digital Oceans, Massachusetts Institute of Technology Sea Grant College Program, MITSG 0301, pp. 36.
Rountree, R., Goudey, C., Hawkins, T., Luczkovich, J. J., and Mann, D. (**2002**). "Listening to Fish: An International Workshop on the Applications of Passive Acoustics to Fisheries," Sea Grant College Program, Massachusetts Institute of Technology, Cambridge, MA, pp. 34.
Schusterman, R. J., Kastak, D., Levenson, D. H., Reichmuth, C. J., and Southall, B. L. (**2000**). "Why pinnipeds don't echolocate," J. Acoust. Soc. Am. **107**, 2256–2264.
Skinner, J. D., Edmonds, G. L., Ensberg, D. E., D'Spain, G. L., and Hodgkiss, W. S. (**2003**). "A high speed acoustic data acquisition system using mostly COTS components," Oceans 2003 Conference, San Diego, pp. 8.
Tavolga, W. N., ed. (**1964a**). *Marine Bio-Acoustics* (Pergamon, New York), pp. 413.
Tavolga, W. N. (**1964b**). "Sonic characteristics and mechanisms in marine fishes," in *Marine Bio-Acoustics*, edited by W. Tavolga (Pergamon, New York), pp. 195–211.
Tavolga, W. N., ed. (**1967**). *Marine Bio-Acoustics* (Pergamon, New York), Vol. **2**, pp. 353.
Tavolga, W. N. (**2002**). "Fish bioacoustics: A personal history," Bioacoustics 12, 101–104.
Terrill, E. (private communication **2003**).
Thompson, P. O. (**1965**). "Marine biological sound west of San Clemente Island; Diurnal distributions and effects on ambient noise level during July, 1963," NEL Rept. 1290, Naval Electronics Laboratory (now part of the Space and Naval Warfare Center, San Diego), San Diego, pp. 42.
University of California, Division of War Research (**1946**). "Underwater noise caused by snapping shrimp," Contract NObs-2074, UCDWR No. U337, University of California, Division of War Research, San Diego, CA.
Weinberg, H., and Keenan, R. E. (**1996**). "Gaussian ray bundles for modeling high-frequency propagation loss under shallow-water conditions," J. Acoust. Soc. Am. **100**(3), 1421–1431.
Wenz, G. M., Calderon, M. A., and Scanlan, T. F. (**1965**). "Underwater acoustic ambient-noise and transmission tests west of San Clemente Island, July, 1963," NEL Rept. 1260, Naval Electronics Laboratory (now part of the Space and Naval Warfare Center, San Diego), San Diego, 46 pp. plus 5 app., declassified in 1977.
Yoklavich, M. (private communication, **2005**).

Acoustic detection of North Atlantic right whale contact calls using the generalized likelihood ratio test

Ildar R. Urazghildiiev[a] and Christopher W. Clark
Bioacoustics Research Program, Cornell Laboratory of Ornithology, Ithaca, New York 14850-1999

(Received 3 January 2006; revised 12 June 2006; accepted 29 June 2006)

This paper addresses the problem of passive acoustic detection of contact calls produced by the highly endangered North Atlantic right whale *Eubalaena glacialis*. The proposed solution is based on using a generalized likelihood ratio test detector of polynomial-phase signals with unknown amplitude and polynomial coefficients observed in the presence of locally stationary Gaussian noise. The closed form representation for a minimal sufficient statistic is derived and a realizable detection scheme is developed. The receiver operation characteristic curves are calculated using empirical data recordings containing known right whale calls. The curves demonstrate that the proposed detector provides superior detection performance as compared with other known detection techniques for northern right whale contact calls. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2257385]

PACS number(s): 43.30.Wi, 43.80.Ev [WWA] Pages: 1956–1963

I. INTRODUCTION

Passive acoustic detection of acoustically active animals has become an increasingly important area of research. The successful application of automatic acoustic detection algorithms is particularly critical for monitoring endangered species or in cases where sound is the most effective detection modality.[1–6] Inherent in biological detection is the challenge of coping with natural variation in signal features. Although some animal acoustic signals are highly stereotyped, all biological signals contain variability and in many, such variability is considered adaptive.[7] For highly adaptive signals such as long-distance contact calls, selection often favors sounds optimized for a habitat's sound transmission and ambient noise characteristics, which leads to constraints on basic acoustic features such as call duration, bandwidth, and center frequency.[8] Thus, for example, the contact call of the highly endangered northern right whale *Eubalaena glacialis* is a simple, frequency-modulated (FM) upsweep in the 50–400 Hz frequency band lasting about 1 s.[6,9] Within these constraints, contact calls exhibit variability in initial frequency, FM rate, duration, and bandwidth. Given these considerations, a detection algorithm must achieve a balance between incorporating enough variability so as to be successful at detecting the signal of interest, while limiting the extent to which variability is included so as to be efficient.

In its simplest form, the problem of signal detection can be cast in the framework of binary hypothesis testing, in which an observation must be assigned to one of two possible outcomes; presence or absence. Optimal detectors minimizing the average number of detection errors can be developed using statistical decision theory. There are many optimal methods or strategies that could be applied, such as Bayesian, Neyman-Pearson, maximum likelihood (ML), min-max, constant false alarm rate (CFAR), and others.[10–12] The choice of the method depends on the goals and conditions existing in a particular application. However, all of the optimal strategies are similar in the sense that they use the likelihood ratio test. The likelihood ratio is a minimal sufficient statistic for the whole class of optimal hypothesis testing methods. Therefore, the structure of the detector as well as its performance is completely specified by the statistic being used.

In this paper, the problem of passive acoustic detection of North Atlantic right whale (NARW) contact calls using a single hydrophone as a sensor is considered. The main goal of the paper is to derive a closed form representation for the minimal sufficient statistic. To solve this problem, a statistical model of the observed process is developed. The NARW contact calls are represented as a family of polynomial-phase signals (PPS) with unknown amplitude and polynomial coefficients. To model ambient noise, a locally stationary Gaussian random process is used. Based on these models, the closed form representation for the generalized likelihood ratio test (GLRT) is derived.

The second goal of this paper is to design a realizable GLRT-based detection scheme. We show that a GLRT-based statistic can be calculated using a bank of matched filters (MF) with frequency responses specified by the polynomial coefficients of the PPS. The polynomial coefficients can be chosen from a training data set consisting of a number of NARW contact calls selected by a human operator. We design a robust and computationally efficient algorithm for calculating the MF output. The algorithm is reduced to dividing the data recording into chunks that are 8–16 s long. For each chunk of data, the fast Fourier transform (FFT) spectrum is calculated, the noise power spectrum density (PSD) is estimated, and the data are normalized. To obtain a robust PSD estimate, median filtering of a data spectrogram is proposed. Then the MF output is computed as the inverse FFT of the product of the normalized data spectrum and the frequency

[a]Electronic mail: iru2@cornell.edu

 0001-4966/2006/120(4)/1956/8/$22.50

response of the MF. The outputs of P MFs are compared, and the maximum is taken. This results in the GLRT-based statistic.

It is important to note that ambient noise can include transient sounds from biological and man-made sources. This can be particularly true in the western North Atlantic for portions of the year when singing humpback whales occur in areas frequented by right whales.[5] Humpback transients were not a concern in these data analyses since no singers were evident in any of the recordings. Because of noise transients the accuracy of the proposed noise model can be seriously degraded. However, as the tests demonstrate for the data considered here, the Gaussian model is acceptable for about 90% of the observed ambient noise conditions. As a result, the proposed GLRT detector exhibits optimality properties in empirical situations even when the probability of occurrence of transient noise is high.

The detection performance of the GLRT detector was compared with the performances of some nonparametric and heuristic NARW detection techniques described in the literature. This included comparisons to the square-law "energy" detector,[11] the spectrogram cross-correlation detector,[13–15] the "edge" detector,[16] and a technique based on the Kolmogorov-Smirnov test.[17] The detection performance was tested using empirical data recordings that included 1283 NARW contact calls occurring over a continuous 31-day period and two 24-h data subsets with the lowest and highest impulsive noise rate. The resultant receiver operating characteristics (ROC) were used as the basis for performance comparisons. These comparisons demonstrate that in the region of the detection probability 0.75 and more, the GLRT detector reveals the optimality property that it provides the highest probability of detection for a given false alarm probability.

II. OBSERVATION MODEL AND THE MINIMAL SUFFICIENT STATISTIC

It is assumed that the sensor provides a digitized series of real-valued samples $x(t)$, $t=1,2,\ldots$, with sampling rate greater than twice the highest frequency in a right whale contact call. The observed process can be represented as

$$x(t)=\sum_i A_i s_i(t-\tau_i)+w(t), \tag{1}$$

where $s_i(t)$ is the wave form of the ith signal of interest (i.e., NARW contact call), A_i is the unknown positive nonrandom scalar representing the amplitude, τ_i is the time-of-arrival (TOA) of the ith signal, and $w(t)$ is ambient noise. It is assumed that signal durations are limited by the value of N such that an arbitrary i-th wave form can be represented as

$$s(t)=\cos[\theta(t)], \quad t=1,2,\ldots,N, \tag{2}$$

where $\theta(t)$ is the phase. The process of signal detection is reduced to calculating some statistic $z(t)=z[x(t)]$ as a function of samples $x(t)$ and comparing the statistic to a critical threshold. The data segment specified by the time window of N samples with start time t is denoted as a N-dimensional vector $\boldsymbol{x}(t)=[x(t),x(t+1),\ldots,x(t+N-1)]^T\in E^N$, where E^N is Euclidean N-dimensional space and the superscript symbol "T" denotes the transpose. Then the statistic can be represented as

$$z(t)=z[\boldsymbol{x}(t)]. \tag{3}$$

To derive the transformation (3) in closed form, a parametric model for the vector $\boldsymbol{x}(t)$ is introduced.

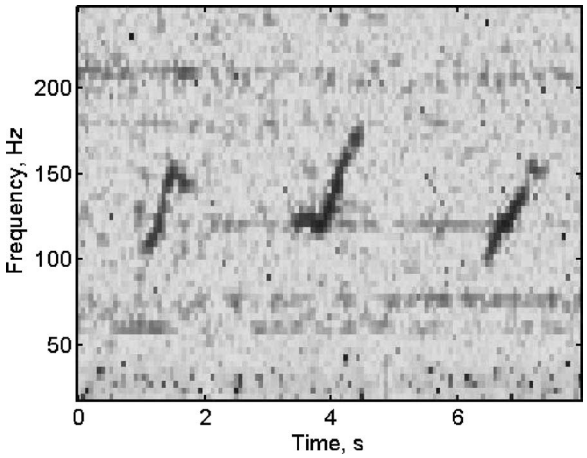


FIG. 1. A spectrogram of three NARW contact calls.

The model of signals can be developed based on the observation that the frequency modulations in NARW contact calls (Fig. 1) are similar to those of frequency-modulated (FM) upswept signals.[9,13–16] Therefore, the signal model is represented by a class of polynomial-phase signals. Using this model, the phase of an arbitrary PPS can be written as[18]

$$\theta(t)=\sum_{m=0}^{M} a_m t^m, \quad t=1,2,\ldots,N, \tag{4}$$

where a_m is the mth polynomial coefficient. Let us introduce the notation $f_m=(m+1)a_{m+1}/(2\pi)$, $m=0,1,\ldots,M-1$. In this notation, the instantaneous frequency of the PPS can be represented as

$$f(t)=\frac{1}{2\pi}\frac{d\theta(t)}{dt}=\sum_{m=0}^{M-1} f_m t^m. \tag{5}$$

The initial phase a_0 is assumed to be a random value uniformly distributed over the interval $(0,\ldots,2\pi)$. It is a nuisance parameter with respect to the detection problem. Therefore, for the signal model parameters we introduce the nonrandom vector

$$\boldsymbol{\lambda}=(f_0,\ldots,f_{M-1})^T\in U_\lambda, \tag{6}$$

where U_λ is the set of model parameters. We define U_λ as a finite discrete set consisting of P elements, $U_\lambda=\{\lambda_1,\lambda_2,\ldots,\lambda_P\}$. The admissible set of λ is defined by the time-frequency characteristics of NARW contact calls. Note that for an arbitrary PPS, $s(t)$, the instantaneous frequencies $f(t)$ approximate the positions of the peak absolute values of the short-time Fourier transforms (STFT) of $s(t)$. This property is important in practice since the visual analysis of the spectrogram is one of the primary methods of signal detection in bioacoustics. For a quadratic-phase signal, the first three polynomial coefficients can be interpreted as follows:

f_0 is the start frequency, and f_1 and f_2 represent the slope and the curvature, respectively.

Thus, as a model of the signal wave form, we introduce the N-dimensional vector $\boldsymbol{s}(\boldsymbol{\lambda})=[s(1,\boldsymbol{\lambda}),s(2,\boldsymbol{\lambda}),\ldots,s(N,\boldsymbol{\lambda})]^T$ with the elements $s(t,\boldsymbol{\lambda})=\cos[\theta_\lambda(t)+a_0]$, where $\theta_\lambda(t)=2\pi\Sigma_{m=1}^{M}(m)^{-1}f_{m-1}t^m$ is obtained by replacing the coefficients a_m by f_m in Eq. (4). Then the detection problem can be formulated in terms of testing the following hypothesis:

$$H_0{:}\boldsymbol{x}(t)=\boldsymbol{w}(t),\quad H{:}\boldsymbol{x}(t)=A\boldsymbol{s}(\boldsymbol{\lambda})+\boldsymbol{w}(t), \tag{7}$$

where $\boldsymbol{w}(t)=[w(t),w(t+1),\ldots,w(t+N-1)]^T\in E^N$ is the vector of noise and A is a positive scalar representing the signal amplitude. The signal parameters $\{A,\boldsymbol{\lambda}\}$ are assumed to be unknown and nonrandom. The null hypothesis H_0 represents the case of signal absence, and the alternative hypothesis H corresponds to the case of signal presence. The alternative hypothesis is a composite. For the hypothesis H_0 and H, a uniformly most powerful test[10–12] does not exist. Therefore a suitable strategy is to implement a minimally sufficient statistic such as the generalized likelihood ratio[11,12]

$$z(t)=\frac{\max_{A,\boldsymbol{\lambda}}W[\boldsymbol{x}(t)|H]}{W[\boldsymbol{x}(t)|H_0]}, \tag{8}$$

where $W(\boldsymbol{x}|H_0)$, $W(\boldsymbol{x}|H)$ are the probability density functions of the vector $\boldsymbol{x}$ under the two hypotheses. A maximum in Eq. (8) is reached when substituting the actual signal parameters A, $\boldsymbol{\lambda}$ in the likelihood ratio. However, since these parameters are unknown, the ML estimates can be used instead such that the statistic (8) can be written as

$$z(t)=\frac{W[\boldsymbol{x}(t)|H,\hat{A},\hat{\boldsymbol{\lambda}}]}{W[\boldsymbol{x}(t)|H_0]}, \tag{9}$$

where $\hat{A}$, $\hat{\boldsymbol{\lambda}}$ are the ML estimates of the parameters A and $\boldsymbol{\lambda}$, respectively.

An attractive model for ambient noise is a Gaussian random process. However, the presence of transient noise results in an empirical distribution of ambient noise that can be different from Gaussian. To verify the adequacy of the Gaussian model, statistical analyses using empirical data recordings were carried out. The description of the data set used as well as the experiments conducted are given in Sec. IV. The results of the experiments have demonstrated that for approximately 90% of data chunks with the length from 8 to 16 s, statistical properties are similar to those of a stationary Gaussian process. Therefore, a colored, locally stationary Gaussian random process is introduced as the model of ambient noise. We suppose that on any time interval $\Omega(q)=\{qN_R+1,qN_R+2,\ldots,(q+1)N_R\}$, $q=0,1,2,\ldots$, consisting of N_R samples each ($N_R\gg N$), variations in the covariance matrix are negligibly small. Therefore, the noise vector in Eq. (7) is assumed to be Gaussian distributed with zero mean and covariance matrix $\boldsymbol{R}(t)=\boldsymbol{R}(q)$, so that

$$\boldsymbol{w}(t)\in N[\boldsymbol{0},\boldsymbol{R}(q)],\quad t\in\Omega(q). \tag{10}$$

The value N_R is taken to ensure the duration of $\Omega(q)$ between 8 and 16 s. Based on Eq. (10) as a statistical model of the observed data, we introduce the following probability density functions:

$$\ln W[\boldsymbol{x}(t)|H_0]=-\boldsymbol{x}(t)^T\boldsymbol{R}(q)^{-1}\boldsymbol{x}(t)/2+b$$

$$\ln W[\boldsymbol{x}(t)|H]=-[\boldsymbol{x}(t)-A\boldsymbol{s}(\boldsymbol{\lambda})]^T\boldsymbol{R}(q)^{-1}[\boldsymbol{x}(t)-A\boldsymbol{s}(\boldsymbol{\lambda})]/2+b \tag{11}$$

$$\boldsymbol{\lambda}\in U_\lambda,\quad t\in\Omega(q),\quad q=0,1,2,\ldots,.$$

Here b is a constant that does not depend on the signal.[11] It follows from Eqs. (9) and (11) that the sufficient statistic is fully specified by the ML estimates $\hat{A}$, $\hat{\boldsymbol{\lambda}}$. Note that since the signal vector depends on the random initial phase, the probability density function $W(\boldsymbol{x}|H)$ should be averaged over the parameter a_0. Therefore, it is necessary to use the in-phase and quadrature components of the signal.[11] Let us introduce the complex vector $\boldsymbol{c}(\boldsymbol{\lambda})=[c(1,\boldsymbol{\lambda}),c(2,\boldsymbol{\lambda}),\ldots,c(N,\boldsymbol{\lambda})]\in E^N$ with the elements $c(t,\boldsymbol{\lambda})=\exp[j\theta_\lambda(t)]$. Then the ML estimates $\hat{A}$, $\hat{\boldsymbol{\lambda}}$ can be found from Eq. (11) by maximization of $\ln W[\boldsymbol{x}(t)|H]$ over the parameters $A,\boldsymbol{\lambda}$, as

$$\hat{\boldsymbol{\lambda}}=\arg\max_{\lambda\in U_{\boldsymbol{\lambda}}}|\hat{A}\boldsymbol{x}(t)^T\boldsymbol{R}(q)^{-1}\boldsymbol{c}(\boldsymbol{\lambda})| \tag{12}$$

$$\hat{A}=\frac{|\boldsymbol{x}(t)^T\boldsymbol{R}(q)^{-1}\boldsymbol{c}(\hat{\boldsymbol{\lambda}})|}{|\boldsymbol{c}(\hat{\boldsymbol{\lambda}})^T\boldsymbol{R}(q)^{-1}\boldsymbol{c}(\hat{\boldsymbol{\lambda}})|}. \tag{13}$$

The denominator in Eq. (13) does not depend on the data $\boldsymbol{x}(t)$, and therefore does not affect the structure of the detector. Taking Eqs. (11)–(13) into account, the minimal sufficient statistic $z(t)$ can be rewritten as

$$z(t)=\max_{\lambda\in U_{\boldsymbol{\lambda}}}|\boldsymbol{x}(t)^T\boldsymbol{R}(q)^{-1}\boldsymbol{c}(\boldsymbol{\lambda})|^2. \tag{14}$$

The decision to accept or reject the null hypothesis is made by comparing the statistic $z(t)$ with a critical threshold C. The threshold is determined by applying optimality criteria, which are introduced based on the goals of the experiment as well as on *a priori* information. The probabilities of the null hypothesis $p(H_0)$ and its alternative $p(H)$ are important *a priori* information, but are unknown in practice. However, it is known that the maximum rate of occurrence of NARW contact calls is about 100–150 per hour, while the duration of calls is about 1 s.[6,9,13–16] Therefore, the following condition is satisfied:

$$p(H_0)\gg p(H). \tag{15}$$

For such a condition, the Neyman-Pearson criterion minimizing a false alarm probability for a given detection probability (or vice versa) can be applied.

III. CALCULATING THE STATISTIC USING A BANK OF MATCHED FILTERS

Since signal TOAs are unknown, the statistic $z(t)$ should be calculated for all possible data segments, $\boldsymbol{x}(t)$, $t=1,2,\ldots,$. Direct implementation of Eq. (14) is impractical since implementation of such an algorithm is computation-

ally expensive. Therefore, we propose calculating the GLRT-based statistic using a bank of realizable linear filters.

It follows from Eqs. (10) and (14) that the only time-varying parameter in our model is the noise covariance matrix $\boldsymbol{R}(q)$. This matrix should be estimated for each chunk of data $x(t), t \in \Omega(q)$. To design the realizable detection scheme, it is necessary and sufficient to design the detector for any chunk of data N_R samples long. Without loss of generality, we consider an arbitrary data chunk $x(t), t \in \Omega(q)$, in which the index q specifying the time position of the chunk is omitted.

Let us introduce the complex variables $h(n, \boldsymbol{\lambda})$, $n = 0, 1, \ldots, N-1$ as the coordinates of the vector $\boldsymbol{h}(\boldsymbol{\lambda}) = \hat{\boldsymbol{R}}^{-1}\boldsymbol{c}(\boldsymbol{\lambda}) \in E^N$ where $\hat{\boldsymbol{R}} \in E^{N \times N}$ is a positive defined matrix obtained as a certain estimate of $\boldsymbol{R}(q)$. Then, given $\boldsymbol{\lambda}_p$, $p = 1, 2, \ldots, P$, the complex output of a linear FIR filter with impulse response $h(n, \boldsymbol{\lambda}_p)$ can be represented as a convolution

$$u(t, \boldsymbol{\lambda}_p) = \sum_{n=0}^{N-1} x(t-n) h(n, \boldsymbol{\lambda}_p), \quad t \in \Omega(q). \tag{16}$$

This filter can be referred to as a matched filter[11] (MF) for the signal $\boldsymbol{s}(\boldsymbol{\lambda}_p)$. The estimate $\hat{\boldsymbol{\lambda}}$ can then be calculated using a bank of P matched filters as

$$\hat{\boldsymbol{\lambda}} = \arg \max_p |u(t, \boldsymbol{\lambda}_p)|. \tag{17}$$

The combination of Eqs. (14), (16), and (17) results in

$$z(t) = \max_p |u(t, \boldsymbol{\lambda}_p)|^2. \tag{18}$$

Equation (18) represents a realizable scheme for calculating the GLRT-based statistic. This scheme requires $M_T = O(PN_RN)$ complex multiplications per chunk of data. A decrease in M_T can be achieved by calculating the MF output in the frequency domain using the FFT. In the frequency domain, the convolution relation Eq. (16) corresponds to a multiplication of the respective Fourier transforms. Let N_R be equal to a power of 2, symbol $\Im\{\cdot\}$ denote the FFT, and symbol $\Im^{-1}\{\cdot\}$ denote the inverse FFT. Then

$$|u(t, \boldsymbol{\lambda}_p)| = |\Im^{-1}\{U(\omega_i, \boldsymbol{\lambda}_p)\}|, \quad i = 0, \ldots, N_R - 1, \tag{19}$$

where

$$U(\omega_i, \boldsymbol{\lambda}_p) = X(\omega_i) C(\omega_i, \boldsymbol{\lambda}_p) / B(\omega_i) \tag{20}$$

is the FFT spectrum of the MF output, $X(\omega_i) = \Im\{x(t)\}$ is the FFT spectrum of the data chunk $x(t)$, $t \in \Omega(q)$, calculated at the point $\omega_i = 2\pi i / N_R$, $C(\omega_i, \boldsymbol{\lambda}_p) = \Im\{c(t, \boldsymbol{\lambda}_p)\}$ is the frequency response of the pth MF computed from the samples $c(t, \boldsymbol{\lambda}_p)$ padded by zeros, and $B^2(\omega_i)$ is the noise PSD. Applying the FFT to the whole bank of matched filters [see Eqs. (18) and (19)] requires $M_F = O(PN_R \ln N_R)$ complex multiplications per chunk of data. Observe that $M_F \ll M_T$ for typical values of $N > 1000$ and $N_R = (8, \ldots, 16)N$. To implement the FFT-based algorithm in Eq. (19), the noise PSD should be estimated for each chunk of data.

The standard Bartlett or Welch methods of PSD estimaton[19] provide an acceptable accuracy if noise has no outliers. However, data recordings often include powerful signals and impulsive noise, which degrade the accuracy of conventional noise spectrum estimators. To cope with the problem of outliers, median filtering is widely used (see, e.g., Refs. 21 and 22). Therefore, to calculate the noise PSD estimate, we propose a robust technique based on median filtering of a spectrogram. Let $G(\tilde{\omega}_k, n)$, $k = 0, 1, \ldots, K-1$, and $n = 1, 2, \ldots, N_G$ denote the spectrogram calculated for the data chunk $\Omega(q)$ using a sliding short-time window of K samples ($K \ll N$) and overlapped by K_{ov} samples. Here $\tilde{\omega}_k = 2\pi k / K$ is a new set of discrete angular frequencies and $N_G = (N_R - K_{ov})/(N - K_{ov})$ is the number of time frames in the spectrogram. As follows from Eq. (10), for any $t \in \Omega_R(q)$ the variations of noise PSD are negligibly small. Then the noise PSD estimate $\hat{B}(\tilde{\omega}_k)$ can be calculated as a median of the N_G samples of the spectrogram on a frequency-by-frequency basis

$$\hat{B}(\tilde{\omega}_k) = \text{med}\{G(\tilde{\omega}_k, 1), G(\tilde{\omega}_k, 2), \ldots, G(\tilde{\omega}_k, N_G)\}. \tag{21}$$

The values of $B(\omega_i)$ can be estimated from Eq. (21) by interpolation of $\hat{B}(\tilde{\omega}_k)$ at the frequencies ω_i. Computation of the spectrogram $G(\tilde{\omega}_k, n)$ using the FFT requires $M_G = O[N_R K \ln K / (N - K_{ov})]$ complex multiplications. In practice $(N - K_{ov}) \gg 1$ and $K \ll N_R$. Hence, $M_G \ll M_F$, i.e., calculating the MF outputs, Eq. (19), requires most of the computations.

Thus Eqs. (18)–(21) represent a robust and computationally efficient scheme for calculating the GLRT-based statistic of polynomial-phase signals with unknown polynomial coefficients in the presence of locally stationary Gaussian noise. Note that any two values of $z(t)$ and $z(t+n)$, $0 < n \leq N$ are correlated since overlapping data segments are used for calculating them. Consequently, the number of samples of $z(t)$, Eq. (18), used for the subsequent analysis can be reduced without losing essential information. On the other hand, the under sampling of $z(t)$ should not result in missed signal peaks. Our observations show that in the vicinity of a signal's occurrence, the peak of the function $z(t)$ can be very sharp so that simple uniform down sampling does not provide a good tradeoff between reduced samples and the number of missed signal peaks. To avoid this difficulty, we propose a nonuniform down-sampling algorithm. According to this algorithm, each chunk is partitioned into N_{seg} nonoverlapping segments consisting of N samples so that $N_{\text{seg}}N = N_R$. To avoid the edge effect, only the first $(N_{\text{seg}} - 1)$ segments of each chunk are used to calculate the sufficient statistics. (This, in turn, requires the chunks to be overlapped by N samples). For each jth segment $1 \leq j < N_{\text{seg}}$ the sufficient statistic z_j is calculated as the maximum value of $z(t)$ in the segment. To avoid taking two highly correlated values of the statistic corresponding to a pulse of noise or signal located between two adjacent segments, an additional rule is applied: The interval between the time indexes corresponding to the statistics z_j, z_{j+1} in the adjusted segments must not be less than $0.5N$. Thus, this algorithm decreases the number of samples on the MF bank outputs by a factor of N without missing peaks for any two signals separated by $0.5N$ samples or more.

IV. TESTING

The main goal of the testing analysis was to quantitatively compare the performances of the proposed GLRT-based detector with several published methods for detecting NARW contact calls. Detection performance was estimated for two noise conditions corresponding to the day with the lowest impulsive noise rate and the day with the highest impulsive noise rates. The data used for the tests were taken from 31 days of continuous recordings for the period from December 18, 2002 to January 18, 2003 in Cape Cod Bay, MA, collected at a 2 kHz sample rate using bottom-mounted hydrophones.[16,20] The detection performances were judged based on the comparison of the ROC curves.

In the first part of the analysis, the hypothesis that local ambient noise had a Gaussian distribution was tested. The nonparametric Kolmogorov-Smirnov (KS) test implemented in MATLAB 7.0 R14, at a 0.05 significance level was used. The data were divided into 299,839 "short chunks" of length 8.192 s and into 149,919 "long chunks" of length 16.384 s. Before testing, all chunks were bandpass filtered using a Chebyshev type-1 IIR filter with cut-off frequencies of 40 and 250 Hz, and normalized so that the standard deviation was 1. The upper cut-off frequency was chosen based on the range of energy distribution of more than 99% of the NARW contact calls. The results indicate that the hypothesis was accepted for 90.5% of the "short chunks" and for 86.7% of the "long chunks." (Note that when testing a Gaussian random process, the percentage of positives is asymptotically equal to 95% for a 0.05 significance level.) The main reasons for rejecting the hypothesis were due to the presence of strong NARW contact calls or different kinds of impulsive noise. Although it is recognized that this test does not control for all factors that might influence recordings of ambient noise (e.g., time of year, sensor location), we believe that such factors for the Cape Cod Bay region would have little influence on the outcome, and we conclude that these test results support the conclusion that ambient noise can be modeled as Gaussian over periods of less than 17 s. Analysis also shows that changes in noise variance are negligibly small within the time intervals less than several minutes. Therefore, a colored, locally stationary Gaussian random process can be accepted as the model of ambient noise. It is interesting to note that during the data period from December 18, 2002 through January 18, 2003 wave heights in the area were typically between 0.5 and 4 m, with one occurrence of 4–5 m.

In the second part of the analysis, the empirical distributions of the model parameters were computed using a training data set of 721 NARW contact calls with a high signal-to-noise ratio (SNR). The training calls were collected in the period from 2001 to 2003 at Cape Cod Bay and Great South Channel regions. The histogram of signal duration for this sample set is shown in Fig. 2 (top frame). Based on this distribution, signal duration was chosen as 1.024 s so that $N=2048$. Note that from Fig. 2, the duration of a NARW contact call may be longer than 1 s, especially in cases where the initial portion of the call is relatively constant in frequency or downward inflected. However, our observations show that the duration of the upswept part of signal, the key feature of contact calls, is close to 1 s.

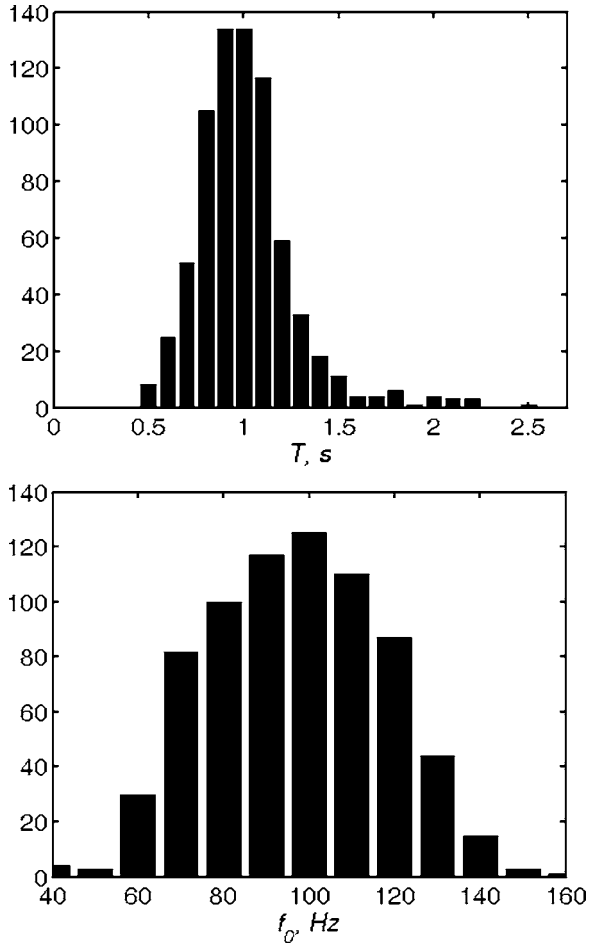


FIG. 2. The one-dimensional histograms of the model parameters: duration T (top frame), and start frequency f_0 (bottom frame).

Different detection schemes can be introduced by varying the number of polynomial coefficients f_m [see Eq. (5)]. We restricted our investigation here to the case of $M=3$. Such a class of detectors is nearly optimal for linear and quadratic FM signals. The histograms of the polynomial coefficients $\hat{W}(\boldsymbol{\lambda}|H)$ were computed using the ML estimates

$$\hat{\boldsymbol{\lambda}}_i = \arg\max_{p,t} |u(t,\boldsymbol{\lambda}_p)|, \quad t = t_i - 0.5N, \ldots, t_i, \ldots, t_i + 0.5N, \tag{22}$$

where t_i is the start index for ith data segment containing a NARW contact call ($i=1,\ldots,721$) and $\boldsymbol{\lambda}_p$ were taken from the search grid $\Lambda=[\boldsymbol{\lambda}|f_0=\{20,30,\ldots,170\},f_1=\{-150,-125,\ldots,300\},f_2=\{-200,-175,\ldots,350\}]$. The outputs of the MF bank, $u(t,\boldsymbol{\lambda}_p)$, were computed using the technique considered in Sec. III. The observation windows in Eq. (22) were 1.024 s in length and centered at the indexes t_i selected by an experienced human operator. Corresponding time intervals were chosen as follows: the length of data chunks used for computing noise PSD $N_R=32{,}768$ samples (16.384 s in length); the number of segments $N_{\text{seg}}=16$; the number of samples per segment $N=2048$

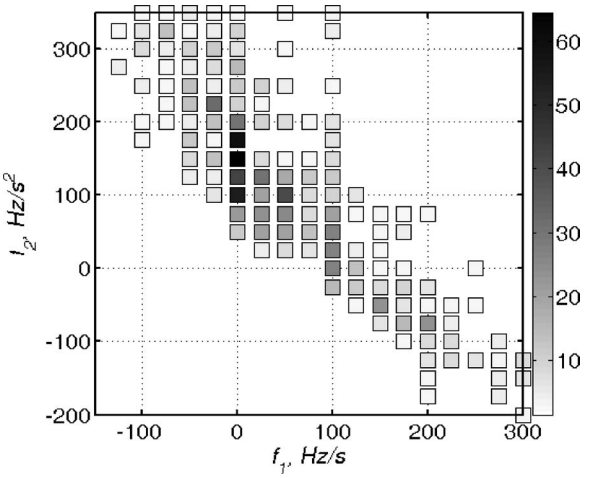


FIG. 3. The two-dimensional histogram of the model parameters f_1 (slope) and f_2 (curvature). The frequency of occurrence for any given parameter combination of f_1 and f_2 is represented by a gray-scale level.

(1.024 s in length). The spectrogram was calculated using a rectangular, short-time window $K=256$ samples in length and overlapped by $K_{ov}=128$ samples.

The one-dimensional (1D) histograms of the duration T (top frame) and start frequency f_0 (bottom frame) are shown in Fig. 2. The 2D histogram of the model parameters f_1, f_2 is depicted in Fig. 3.

In the second part of the analysis, the ROC curves were computed for the proposed statistic [see Eqs. (18)–(21)] as well as for some published detection techniques. For this purpose, the empirical statistical distributions for $\hat{W}(z|H_0)$ and $\hat{W}(z|H)$ under the null and alternative hypotheses were calculated. The distribution $\hat{W}(z|H)$ was computed using a test data set of 1283 contact calls selected by an experienced human operator. The test data set was different from the training set. About 20% of the calls used in the test set were hardly visible in the spectrogram.

To compute distributions of $\hat{W}(z|H_0)$, the two days (24 h) with the lowest and highest impulsive noise rates per day of observation were selected from the data [see Figs. 4 and 5 (top frame)]. These correspond to the data recordings collected on December 29 and 31, 2002, respectively. For each subset, a total of 84,375 values of the z_j statistic were used to compute $\hat{W}(z|H_0)$ and to build the ROC curve. Both the training and testing data subsets used in our experiments are available upon request.

For the GLRT detector, any $\boldsymbol{\lambda}_p$ specifies the frequency response of the corresponding linear filter [see Eq. (20)]. Since the number of filters P influences the computational complexity of the GLRT detector, the influence of the number of filters on the detection performance is of great practical interest. This issue was studied in the tests. For this purpose, the values of the histogram $\hat{W}(\boldsymbol{\lambda}|H)$ were sorted in a descending order so that $\hat{W}(\boldsymbol{\lambda}_1|H) \geq \hat{W}(\boldsymbol{\lambda}_2|H) \geq \cdots > 0$. The filter frequency responses were set up based on the first P values $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_P$ maximizing $\hat{W}(\boldsymbol{\lambda}|H)$. The exercise was completed for $P=1$, 12, 36, and 80 filters.

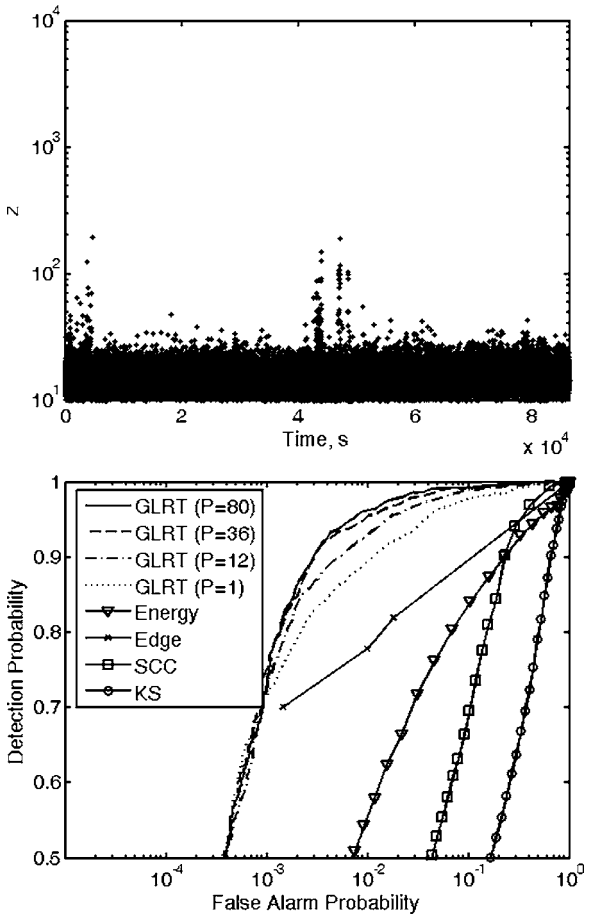


FIG. 4. The values of the GLRT-based statistic (top frame) and detection probability versus false alarm probability (bottom frame). Data collected on December 29, 2002 when the lowest impulsive noise rate was observed.

Along with the GLRT detector, several known NARW detection techniques were tested. The following four detectors and the corresponding statistics were considered in our experiments: a square-law "energy" detector[11]

$$r(t) = \boldsymbol{x}^T(t)\hat{\boldsymbol{R}}^{-1}\boldsymbol{x}(t), \tag{23}$$

a detector based on the Kolmogorov-Smirnov test (KS),[17]

$$k(t) = K[\boldsymbol{x}(t)], \tag{24}$$

a detector based on the spectrogram cross-correlation (SCC),[13–15]

$$y(t) = \sum_{t_0} \sum_{k} H(t_0, \omega_k) G(t - t_0, \omega_k), \tag{25}$$

and an "edge" detector[16]

$$d(t) = D[\boldsymbol{x}(t)]. \tag{26}$$

Here $K(\boldsymbol{x})$ and $D(\boldsymbol{x})$ are the values of the statistics calculated by the Kolmogorov-Smirnov test and the "edge" detector, respectively, and $H(t_0, \omega_k)$ is the kernel.[13,15] The statistics in Eqs. (23) and (24) were calculated using a sliding time window of 2048 samples with 254 overlapping samples. The data were prefiltered with a bandpass filter with cut-off frequencies of 40 and 250 Hz. The spectrogram in Eq. (25) was computed using a Hamming weighting function and a sliding time window of 512 samples with 384 overlapping

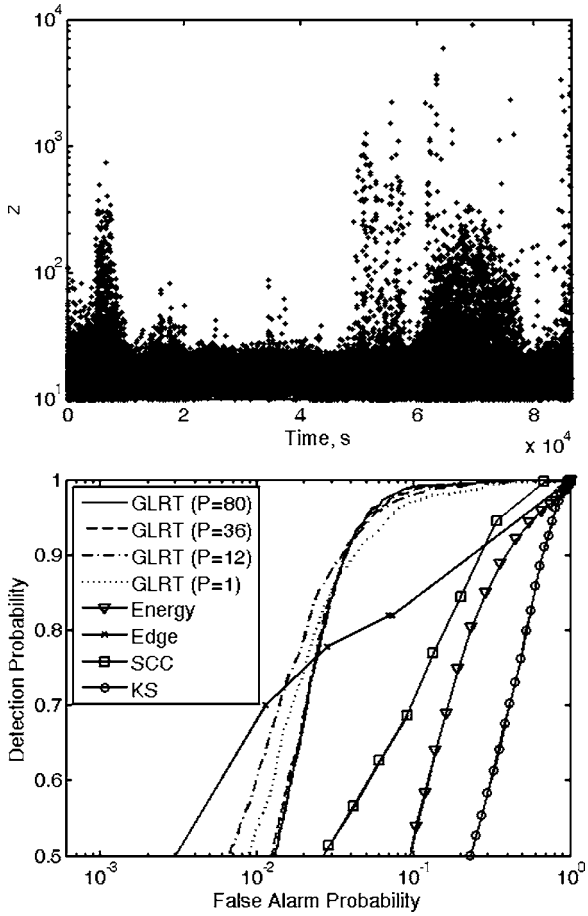


FIG. 5. The values of the GLRT-based statistic (top frame) and detection probability versus false alarm probability (bottom frame). Data collected on December 31, 2002 when the highest impulsive noise rate was observed.

samples.[14] Both the Mellinger[14] and Munger *et al.*[15] kernels were tested, and the Mellinger kernel (a 0.8 s duration, linear FM signal from 80 to 170 Hz) yielded a higher detection probability for a given probability of false alarm. Therefore, the SCC detector with the Mellinger kernel was used in the comparative test analysis. To calculate the statistic $d(t)$, MATLAB code, kindly provided by D. Gillespie,[16] was used.

Using the resultant distributions of $\hat{W}(z|H_0)$ and $\hat{W}(z|H)$, the detection probability $\beta(C)$, and false alarm probability $\alpha(C)$, were calculated over the range of thresholds $-\infty < C < \infty$. ROC curves were then plotted, as shown in Figs. 4 and 5 (bottom frames), with detection probability a function of false alarm probability.[11,12]

Computational cost is another important characteristic of the detection algorithms. Although the problem of optimization of computational procedures was not considered in this work, we compared the computational costs based on the run time required to process 24 h of recordings. The algorithms were tested using a standard PC with Pentium-IV 3 GHz processor and MATLAB 7.0 R14. The run times for computing the GLRT-based statistic [see Eqs. (18)–(21)] depending on the number of matched filters P were 1435 s ($P=1$), 3432 s ($P=12$), 7623 s ($P=36$), and 15,268 s ($P=80$). When testing the other detectors, the following values of run time were obtained: 153 s for the "energy" detector, 1570 s for the KS detector, 265 s for the SCC algorithm, and 299 s for the "Edge" detector.

V. DISCUSSION

The results of the analyses conducted here allowed us to compare the detection performances of various detectors operating on the same data. The influence of biological and other kinds of transient sounds on the performance of these detection algorithms could also be analyzed. Given that a human operator detected NARW calls based on visual and aural evaluation of spectrograms, only calls with relatively high SNR were used when calculating the distributions of $\hat{W}(z|H)$. As a result, the ROC curves do not reflect the true detection performance, especially for the calls with low SNR that were not detected by the human operator.

Figure 4 shows that under condition of low impulsive noise rate, the GLRT detector provides the highest detection probability under a given probability of false alarm. The performance of the GLRT detector improves as the number of filters increases. However, the improvement is negligible when the number of filters $P \geq 36$. This means that an acceptable tradeoff between performance and computational cost can be reached. In the case of high impulsive noise rate (see Fig. 5), in the region of the detection probability $0.75 < \beta < 0.9$, the minimum false alarm probability was reached by the GLRT detector with $P=12$ matched filters. This result can be explained by the fact that when there are many noise pulses, ambient noise has a nonuniform PSD. Therefore, enhancement of the detector's "passband" by using more matched filters increases the noise level on the detector output, leading to a decrease in detection probability. This phenomenon should be taken into account in future research.

The results of these analyses also demonstrate that the spectrogram cross-correlation detector[14] using a single kernel does not provide high detection probability for an acceptable probability of false alarm. Since there is natural variability in contact calls, a possible way of improving the performance of this type of detector would be to use a larger number of kernels calculated based on the empirical distribution of the model parameters.

The "edge" detector[16] performs both detection and recognition of NARW contact calls in the presence of background and impulsive noise. The discrete-valued statistic $d(t) \in \{0,1,\ldots,9\}$ can be interpreted as a "quality" of a NARW call where 0 and 9 represent the worst (noise) and the best (NARW call) quality, respectively. The "edge" detector makes it possible to reject a certain number of noise pulses with a frequency modulation different from that of NARW contact calls. As a result, in the case of a high impulsive noise rate, the "edge" detector provided a higher probability of detection under given false alarm probability at the region of $\beta < 0.75$. However, as Figs. 4 and 5 show, this algorithm yields unacceptably high false alarm values for a detection probability ≥ 0.8.

The nonparametric detector based on the statistic $k(t)$ [see Eq. (24)] provided the lowest detection performance.

This result can be explained by the fact that high values of the KS statistic were observed only when strong signals and noise pulses were present.

The results of these analyses have demonstrated that the presence of impulsive noise will affect the detection performances of all the tested algorithms. Increases in false alarm probability are not that likely if the impulsive noise rate is low [see Fig. 4]. However, a high impulsive noise rate results in a dramatic increase in false detections [see Fig. 5]. Since many kinds of impulsive sounds (e.g., humpback whale song units) can be similar to contact calls in terms of duration, frequency bandwidth, or FM rate, discriminating between impulsive noise and NARW calls is an important practical problem that needs further effort.

Comparison of detector run times shows that the GLRT-based technique is computationally more expensive than the nonparametric and heuristic techniques. Therefore, further work is needed to decrease the computational cost of calculating the GLRT-based detection statistic without negatively affecting detection performance.

VI. CONCLUSION

The problem of detecting North Atlantic right whale contact calls in the presence of ambient noise was considered. As a solution to this problem, a GLRT detector was developed. A statistical model based on a representative sample of empirical data was developed to derive a closed form representation of a minimal sufficient statistic. The model is based on the representation of contact calls as polynomial-phase signals and ambient noise as a locally stationary Gaussian random process. Empirical, *a priori* distributions of the model parameters were computed using training data sets consisting of manually selected contact calls.

The GLRT detector was implemented using a bank of matched filters. Short-term variability ($\leq$16 s) in ambient noise conditions was eliminated through an algorithm that updated an estimate of the noise PSD.

Evaluation of GLRT detection performance was based on a comparison of ROC curves from four other detectors using 31 days of empirical data. The results of these analyses demonstrate that for a given false alarm probability, the GLRT detector provides a higher detection probability than these other detection techniques. Thus, the proposed GLRT detector offers a successful and efficient algorithm for detecting NARW contact calls over a range of ambient noise conditions. These results suggest that this approach would also be successful for other cases in which detection of a relatively simple biological signal with a modest amount of natural variability is required.

ACKNOWLEDGMENTS

The authors wish to thank M. Fowler, D. Ponirakis, A. Warde, and E. Rowland for their assistance in marking right whale calls and to K. Fristrup, S. Vehrencamp, and T. Krein for useful discussions. Thanks also to D. Gillespie for providing Matlab code of the "edge" detector and for important comments on an earlier draft of this manuscript. The authors are grateful to P. M. Baggenstoss, Naval Undersea Warfare Center, Newport, RI for his comments. The authors thank the two anonymous reviewers for their helpful comments and suggestions, which led to improvements in the manuscript. Research funded by NOAA Grant No. NA03NMF4720493 to C.W.C.

[1]D. Mellinger and J. Barlow, "Future directions for acoustic marine mammal surveys: Stock assessment and habitat use," Report of a Workshop held in La Jolla, CA, November 20–22, 2002, NOAA OAR Special Report, NOAA/PMEL No. 2557.

[2]K. M. Stafford, S. L. Nieukirk, and C. G. Fox, "Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific," J. Acoust. Soc. Am. **106**, 3687–3698 (1999).

[3]R. A. Charif, K. A. Cortopassi, H. K. Figueroa, J. W. Fitzpatrick, K. M. Fristrup, M. Lammertink, M. D. Luneau, Jr., M. E. Powers, and K. V. Rosenberg, "Notes and Double Knocks from Arkansas," Science **309**, 1489 (2005).

[4]T. F. Norris, M. A. McDonald, and J. Barlow, "Acoustic detections of singing humpback whales (*Megaptera novaeangliae*) in the eastern North Pacific during their northbound migration,"J. Acoust. Soc. Am. **106**, 506–514 (1999).

[5]C. W. Clark and P. J. Clapham, "Acoustic monitoring on a humpback whale (*Megaptera novaeangliae*) feeding ground shows continual singing into late Spring," Proc. R. Soc. London, Ser. B **271**, 1051–1057 (2004).

[6]J. N. Matthews, S. Brown, D. Gillespie, M. Johnson, R. McLanaghan, A. Moscrop, D. Nowacek, R. Leaper, T. Lewis, and P. Tyack, "Vocalization rates of the North Atlantic right whale (*Eubalaena glacialis*)," J. Cetacean Res. Manage. **3**, 271–282 (2001).

[7]J. W. Bradbury and S. L. Vehrencamp, *Principles of Animal Communication* (Sinauer, Sunderland, MA, 1998).

[8]C. W. Clark and W. T. Ellison, "Potential use of low-frequency sounds by baleen whales for probing the environment: evidence from models and empirical measurements," in *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (The University of Chicago Press), pp. 564–582.

[9]S. E. Parks and C. W. Clark, "Acoustic Communication: Social Sounds and the Potential Impacts of Noise," in *The Urban Whale*, edited by S. Kraus and R. Rolland (Harvard University, Cambridge, MA), Chap. 10, to be published.

[10]E. L. Lehman, *Testing Statistical Hypotheses* (Wiley, New York, 1986).

[11]H. L. Van Trees, *Detection, Estimation and Modulation Theory*, Part I (Wiley, New York, 2001).

[12]A. Hero, "Signal detection and classification," in *Digital Signal Processing Handbook*, edited by E. Madisetti and D. Williams (CRC Press LLC, New York, 1999).

[13]D. K. Mellinger and C. W. Clark, "Recognizing transient low-frequency whale sounds by spectrogram correlation," J. Acoust. Soc. Am. **107**, 3518–3529 (2000).

[14]D. K. Mellinger, "A comparison of methods for detecting right whale calls," Can. Acoust. **32**, 55–65 (2004).

[15]L. M. Munger, D. K. Mellinger, S. M. Wiggins, S. E. Moore, and J. A. Hildebrand, "Performance of spectrogram cross-correlation in detecting right whale calls in long-term recordings from the Bering Sea," Can. Acoust. **33**, 25–34 (2005).

[16]D. Gillespie, "Detection and classification of right whale calls using an "edge" detector operating on a smoothed spectrogram," Can. Acoust. **32**, 39–47 (2004).

[17]B. R La Cour and M. A. Linford, "Detection and classification of North Atlantic right whales in the Bay of Fundy using independent component analysis," Can. Acoust. **32**, 48–54 (2004).

[18]B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal–Part 2: Algorithms and Applications," Proc.-IEEE Ultrason. Symp. **80**(4), 540–568 (1992).

[19]S. L. Marple, *Digital Spectral Analysis with Applications* (Prentice Hall, Englewood Cliffs, 1987).

[20]Cornell Lab of Ornithology, Bioacoustics Research Program. Autonomous Recording Units (ARUs). Pop-Up Ocean Bottom Recorders. Available: http://www.birds.cornell.edu/brp/ARUMarine.html

[21]S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction* (Wiley, Chichester, 2000).

[22]G. R. Arce, *Nonlinear Signal Processing: A Statistical Approach* (Wiley-Interscience, Hoboken, 2001).

Modeling of the influence of a prestress gradient on guided wave propagation in piezoelectric structures

Mickaël Lematre,[a] Guy Feuillard, Emmanuel Le Clézio, and Marc Lethiecq
Université François Rabelais de Tours, LUSSI-GIP Ultrasons, FRE CNRS 2448, E.I.V.L., Rue de la Chocolaterie, 41034 Blois Cedex France

(Received 11 December 2005; revised 5 April 2006; accepted 21 July 2006)

The objective of this study is to model the propagation of guided waves in piezoelectric structures subjected to a prestress gradient. The constitutive equations for a piezoelectric bulk material are first modified to take into account a uniform prestress on a given cross section. Then, these modified constitutive equations are used to derive a formalism for the propagation of guided waves in piezoelectric structures under a prestress gradient. In particular, we modify the recursive stiffness matrix method to introduce a gradient of stress in a piezoelectric structure. Numerical studies are then led for a bending and for an exponential stress profile. For a piezoelectric plate, the Lamb and shear horizontal modes are found to be sensitive to the prestress gradient. In particular, some key features of dispersion curves appearing in the presence of a gradient of properties are highlighted. In the last part, these results are extended to a piezoelectric film laid down on a substrate in order to model the importance of the stress gradient on the behavior of an integrated structure. Lithium niobate is used for the plate and film material, and a silicon crystal is used as the substrate.
 [DOI: 10.1121/1.2336989]

PACS number(s): 43.38.Fx, 43.25.Fe, 43.35.Ns [AJZ] Pages: 1964–1975

I. INTRODUCTION

The study of stress-induced effects in material is of practical importance in nondestructive testing but also in the application field of electroactive devices.[1–3] Indeed, applications such as high-precision quartz resonators, in frequency control, require having the smallest possible sensitivity to both temperature and stress. On the other hand, mechanical sensors require having the largest sensitivity to stress. Moreover, various environmental conditions can be encountered ranging from room temperature to high temperatures and high pressure uses. In many cases, these devices are based on piezoelectric ceramic thick or thin films that are most often submitted to a high internal prestress.[4,5] For instance, prestresses of several hundred MPa up to about 1.5 GPa have been reported for various PZT thin films.[6] Due to nonlinear effects, piezoelectric characteristics are significantly affected by the presence of high mechanical stresses inducing a shift of the material properties from their stress-free values.[6–10]

The prestress of a polycrystalline film originates from three main types of phenomena: the first one is the intrinsic prestress induced by the formation of grain boundaries. During heat treatment, the crystal grains grow and interact with neighboring ones, inducing a shrinkage due to water or solvent evaporation, and to decomposition and pyrolysis of nonvolatile organic species. This effect can be eventually associated with a phase transformation at the transition temperature. The second origin appears in the case where a film is laid down on a substrate. Prestress can appear during thermal treatment due to the mismatch of thermal expansion coefficients between the film and the substrate. Finally, a third extrinsic stress originates from the lattice parameter mismatch between the film and the substrate.

Research on the propagation of acoustic waves in layered piezoelectric media has long been of great interest.[11–13] Several authors have used different types of acoustic waves to characterize the influence of mechanical stress or initial electrical field in isotropic and anisotropic or piezoelectric materials. Sinha[14] studied the effects of temperature and stress on the propagation of bulk waves by a perturbation method. He later considered external extensional and flexural deformations and showed their significant effects on time delay values of surface acoustic waves.[15] Desmet *et al.*[16] examined the influence on the first symmetric and antisymmetric Lamb modes of a tensile stress applied on a polymer foil. Dowaikh[17] studied the propagation of Love waves in a prestressed layered transversely isotropic half-space. Si-Chaib[18] analyzed longitudinal waveforms both numerically and experimentally in the case of steel materials under bending forces. For piezoelectric ceramic plates, experimental studies of the influence of static stresses on Lamb wave propagation were performed.[19–21] The influence of a biasing electrical field on symmetric Lamb waves was theoretically and numerically studied by Liu *et al.*[22] Recently, another theoretical analysis of free vibrations of a piezoelectric body under a biasing electrical field was performed by Yang.[23]

Concerning the modeling of prestress gradient effects on wave propagation, Liu *et al.*[24] theoretically and numerically studied the Love wave propagation for a transversely isotropic piezoelectric film with an inhomogeneous initial stress. They decomposed the film into sublayers in order to discretize the initial stress gradient. Calculations expressed a transfer matrix for each sublayer mapping the prestress behavior. This method can be applied with a good accuracy for a high sampling of the stress profile, i.e., with a great number

[a]Electronic mail: mickael.lematre@univ-tours.fr

of sublayers. Indeed, no numerical instability appears in the transfer matrix method when it is applied to Love modes.

Recently, the authors have proposed a model of the influence of a uniform mechanical stress in piezoelectric structures especially on piezoelectric coupling coefficients and guided waves propagation.[25] Here, we propose to study the influence of an inhomogeneous initial mechanical stress on the propagation of guided waves in piezoelectric plates, i.e., Lamb and shear horizontal (SH) waves, and surface acoustic waves (SAW) in layered structures. The general approach consists in discretizing the plate and in assuming that the initial stress remains uniform in each layer. However, in order to obtain accurate results for sharp prestress profiles, the number of stress samples, i.e., the number of sublayers, has to be high enough. As far as reflection and transmission coefficients and guided waves are concerned, the classical transfer matrix method leads to numerical instabilities for layer thicknesses of several wavelengths. This method uses generalized state vectors including displacements and stress components in each layer.[26,27] Then, the transfer matrix of a layer that relates the displacement-stress vector of the top surface to the similar vector at the bottom is defined as a function of the initial stress, the material properties, and the boundary conditions. The procedure is repeated for each layer leading to the definition of the global transfer matrix of the multilayer. The numerical instabilities come from the fact that this global transfer matrix tends toward a singular matrix for high frequencies and/or thicknesses. Hence, among the different approaches to push numerical instabilities toward higher frequency-thickness products,[28–31] we have chosen the recursive stiffness matrix method proposed by Rokhlin and Wang. The formalism was first presented for general anisotropic layered media,[32,33] and later for a general piezoelectric medium.[34,35] The principle of the recursive stiffness matrix method consists in defining two vectors: a stress vector (resp. a displacement vector) containing the stress components (resp. the displacement components) at both surfaces of the layer. These vectors are related by the stiffness matrix of the layer. Then, the global stiffness matrix is found by using a recursive relation between the stiffness matrices of all the layers. The main interest of the method is the fact that the stiffness matrix never tends toward a singular matrix as the thickness of the layer or the frequency tends to infinity. Hence, its determinant does not approach zero, unlike that of the transfer matrix method. This appropriate conditioning of the stiffness matrix method leads then to an unconditional numerical stability and allows us to obtain an equal efficiency in terms of calculation time compared to the transfer matrix method.

In the first part of the paper, the main steps of the formalism allowing us to obtain the modified constitutive equations of a prestressed piezoelectric material are presented. Then, these equations are used to describe how the recursive stiffness matrix method is modified. This will allow us to derive a model of Lamb waves and SAW propagation in a piezoelectric structure submitted to a prestress gradient. In the second part, the efficiency of the model and the effects of a stress gradient on Lamb, shear horizontal, and surface acoustic waves are studied. Numerical results are produced first for a $LiNbO_3$ plate and then for a $LiNbO_3$ film laid down on a Si substrate.

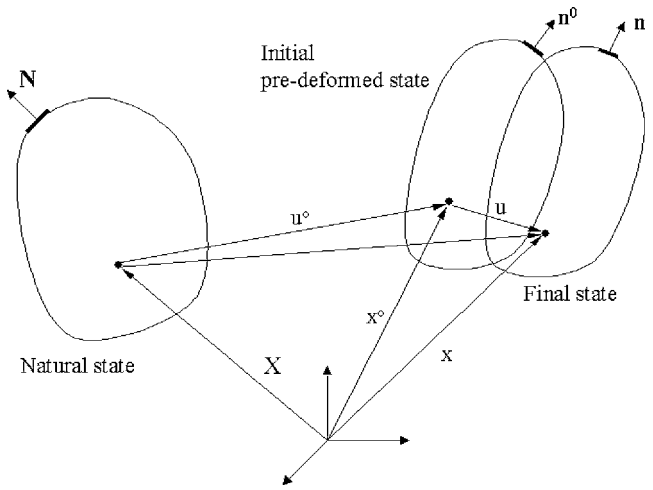


FIG. 1. Coordinate system for an elastic body under an initial deformation.

II. THEORETICAL MODEL

In order to develop a modified Christoffel equation, the constitutive relations of a piezoelectric material under an external stress are derived. Then, the modified equation of motion and Poisson equation are obtained from those of the non-deformed state.

A. Constitutive equations for a general prestressed piezoelectric medium

For a continuous medium under stress, the deformations and motions of a material point can be described through the coordinates x_α^0 and x_k:[36–38]

$$x_\alpha^0 = x_\alpha^0(X_K), \tag{1}$$

$$x_k = \hat{x}_k(x_\alpha^0, t) = x_k(X_K, t), \tag{2}$$

where $k=1, 2, 3$ and $K=\mathrm{I}, \mathrm{II}, \mathrm{III}$.

Here, X denotes a particle coordinate in the un-deformed configuration, called the natural state, and x^0 denotes the coordinate in the pre-deformed state. Thus, the coordinate x (resp. $\hat{x}$) corresponds to the particle position at the final state and described in the un-deformed configuration (resp. in the pre-deformed state), as described in Fig. 1. The coordinate x^0 results of the superimposition of the finite initial displacement, $\mathbf{u}^0$, due to the applied mechanical stress, on the natural state. The coordinate x results in the superimposition of the dynamic infinitesimal displacement of the wave motion, $\mathbf{u}$, on the pre-deformed state. The subscripts K, α, and k denote the components in the Lagrangian coordinate system in the natural state, and the Eulerian coordinate system in the pre-deformed one and in the final configuration, respectively. Finally, t denotes time.

Hence, the partial derivatives are defined as

$$x_{a,K}^0 = \frac{\partial x_\alpha^0}{\partial X_K}, \quad X_{K,\alpha} = \frac{\partial X_K}{\partial x_\alpha^0}, \quad x_{k,\alpha} = \frac{\partial \hat{x}_k}{\partial x_\alpha^0}$$

$$x^0_{\alpha,k} = \frac{\partial x^0_\alpha}{\partial x_k}, \quad X_{K,k} = \frac{\partial X_K}{\partial x_k}, \quad x_{k,K} = \frac{\partial x_k}{\partial X_K}. \tag{3}$$

The motion and Poisson equations for a dielectric material in the natural state are given by

$$(\boldsymbol{\sigma}_{KL} x_{i,L})_{.K} + \rho_0 f_i = \rho_0 \ddot{u}_i, \tag{4}$$

$$D_{I,I} = 0, \tag{5}$$

while boundary conditions are expressed as

$$\boldsymbol{\sigma}_{KL} x_{i,L} N_K = \Delta^\sigma_i \text{ on } S^\sigma, \tag{6}$$

$$D_K N_K = \Delta^q \text{ on } S^q, \tag{7}$$

where $\boldsymbol{\sigma}$ stands for the second Piola-Kirchhoff stress tensor, $\mathbf{D}$ is the electrical displacement vector, ρ_0 is the mass density and $\mathbf{f}$ is the body force per unit mass. Δ^σ_i is the external traction surface density applied on surface S^σ, and Δ^q is the external electrical charge surface density applied on surface S^q. N_K is the component along X_K of the vectors normal to surfaces S^σ or S^q in the natural state (Fig. 1).

A mechanical biasing state produced by an initial stress leads to a new equilibrium state. All the physical variables in the pre-deformed state are designated by a superscript label "0." The equilibrium equations of the pre-deformed body are

$$(\boldsymbol{\sigma}^0_{KL} \delta_{iL} + \boldsymbol{\sigma}^0_{KL} u^0_{i,L})_{.K} + \rho_0 f^0_i = 0, \tag{8}$$

$$D^0_{I,I} = 0, \tag{9}$$

where $\boldsymbol{\sigma}^0$ is the initial stress tensor and δ_{ij} is the Kronecker symbol.

A small external dynamic mechanical and electrical load superimposed on the initial prestressed state creates a wave motion of small amplitude leading to the following relations for the final state:

$$x^t_{i,L} = x^0_{i,L} + x_{i,L}, \quad u^t_i = u^0_i + u_i$$

$$\sigma^t_{KL} = \sigma^0_{KL} + \sigma_{KL}, \quad D^t_1 = D^0_1 + D_1 \tag{10}$$

where σ'_{KL} and D^t_1 are the total Kirchhoff stress and total electrical displacement referring to the natural state, u^t_i is the total particle displacement component in the Euler coordinate system and x^t_i is the final coordinate. The corresponding incremental values due to the dynamic disturbance are σ_{KL}, D_1, and u_i.

Inserting Eqs. (10) in Eqs. (4), (6), and (8) and subtracting (4) and (8) allows us to obtain the motion equation and the mechanical boundary conditions for the current configuration in the form of a second Piola-Kirchhoff stress:

$$(\sigma_{KL}\delta_{iL} + \sigma_{KL} u^0_{i,L} + u_{i,L}\sigma^0_{KL})_{,K} + \rho_0 f_i = \rho_0 \ddot{u}_i, \tag{11}$$

$$(\sigma_{KL} + \sigma_{KL} u^0_{i,K} + u_{i,K}\sigma^0_{JK}) N_J = \Delta^\sigma_i. \tag{12}$$

In practice, the Eulerian coordinate system is assumed to coincide with the Lagrangian coordinate system. Hence, by expressing and subtracting the Taylor developments of the constitutive equations for (σ^T_{ij}, D^T_m) in the current state and for (σ^0_{ij}, D^0_m) in the pre-deformed one, and expressing the strain tensor and the electrical field in terms of the displacement and electrical potential gradients, the constitutive equations of piezoelectricity for the dynamic stress tensor σ_{ij} are

$$\sigma_{ij} = \bar{C}_{ijkl} u_{k,l} + \bar{e}_{mij} \Phi_{,m} \tag{13a}$$

where Φ is the scalar electrical potential and

$$\bar{C}_{ijkl} = C_{ijkl} + (C_{ijnl}\delta_{km} + C_{ijklmn}) u^0_{m,n} + e_{mijkl} \Phi^0_{,m}, \tag{13b}$$

$$\bar{e}_{mij} = e_{mij} + e_{mijkl} u^0_{k,l} - l_{mnij} \Phi^0_{,n}, \tag{13c}$$

and the electrical displacement D_m is

$$D_m = \tilde{e}_{mij} u_{i,j} - \tilde{\varepsilon}_{mn} \Phi_{,n}, \tag{14a}$$

where

$$\tilde{e}_{mij} = e_{mij} + (e_{mjl}\delta_{ik} + e_{mijkl}) u^0_{k,l} - l_{mnij} \Phi^0_{,n}, \tag{14b}$$

$$\tilde{\varepsilon}_{mn} = \varepsilon_{mn} + l_{mnij} u^0_{i,j} - \varepsilon_{mnp} \Phi^0_{,p}. \tag{14c}$$

The parameters $\bar{C}_{ijkl}$, $\bar{e}_{mij}$, $\tilde{e}_{mij}$, and $\tilde{\varepsilon}_{mn}$ depend on the second- and third-order tensors of the piezoelectric material in its natural state, and on the initial displacement and electrical potential gradients. The parameters C_{ijkl}, e_{mij}, and ε_{mn} stand for the second-order elastic, piezoelectric, and dielectric tensors, respectively; C_{ijklmn}, e_{mijkl}, and ε_{mnp} stand for the third-order ones, and l_{mnij} are the electrostrictive constants. All tensors are defined for the piezoelectric material in its natural state. Elastic parameters C_{ijkl} and C_{ijklmn} are defined at constant electrical field, and dielectric tensors, ε_{mn} and ε_{mnp}, are defined at constant strain.

Thus, substituting Eqs. (13a)–(13c) and (14a) into Eqs. (11) and (12) and neglecting the terms of order higher than two allows us to define the fundamental governing equations for motion, electrical displacement, and boundary conditions in the pre-stressed state.

First defining

$$\tilde{\boldsymbol{\sigma}}_{ij} = \tilde{C}_{ijkl} u_{k,l} + \tilde{e}_{mij} \Phi_{,m} = \sigma_{ij} + C_{njkl} u^0_{i,n} u_{k,l} + e_{mjl} u^0_{i,l} \Phi_{,m}, \tag{15a}$$

where

$$\tilde{C}_{ijkl} = \bar{C}_{ijkl} + C_{njkl}\delta_{im} u^0_{m,n}, \tag{15b}$$

the motion equation in a piezoelectric material can be written as

$$(\tilde{\boldsymbol{\sigma}}_{ij} + u_{i,k}\boldsymbol{\sigma}^0_{jk})_{,j} + \rho_0 f_i = \rho_0 \ddot{u}_i, \tag{16}$$

$$D_{i,i} = 0, \tag{17}$$

$$(\tilde{\boldsymbol{\sigma}}_{ij} + u_{i,k}\boldsymbol{\sigma}^0_{jk}) N_j = \Delta^\sigma_i, \tag{18}$$

$$D_i N_i = \Delta^q. \tag{19}$$

Here, due to initial stresses, the tensors $\tilde{C}_{ijkl}$, $\tilde{e}_{mij}$, and $\tilde{\varepsilon}_{mn}$ no longer possess the same symmetry as their equivalents in the natural state, C_{ijkl}, e_{mij}, and ε_{mn}, respectively. Nevertheless, from the definitions of $\tilde{C}_{ijkl}$ and $\tilde{\varepsilon}_{mn}$, the following properties are verified: $\tilde{C}_{ijkl} = \tilde{C}_{klij}$ in all cases, and $\tilde{\varepsilon}_{mn} = \tilde{\varepsilon}_{nm}$ if Φ^0_p is equal to zero.

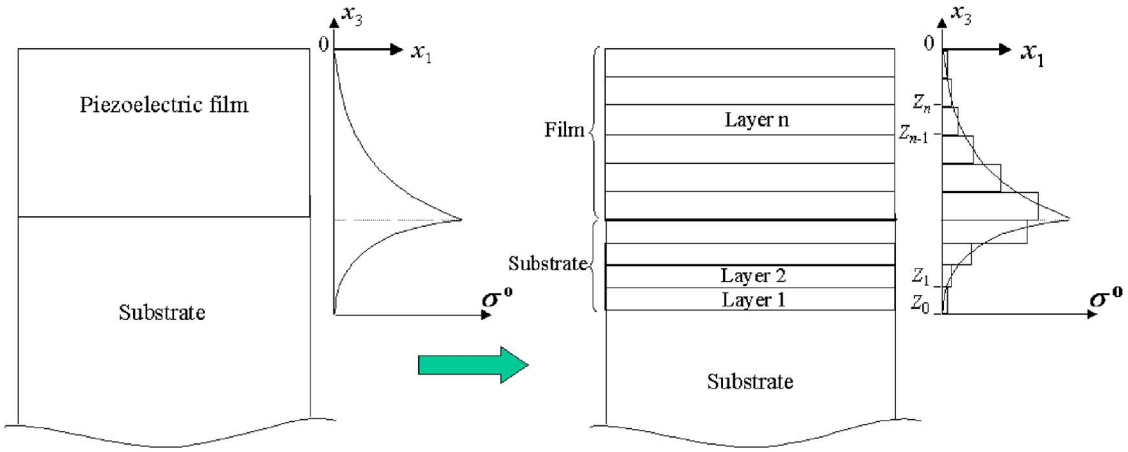


FIG. 2. Configuration used to model the influence of a gradient of prestress in a piezoelectric structure: (a) physical configuration and (b) configuration for computation.

B. Propagation of guided waves in a prestressed piezoelectric layered structure

In the previous section, the constitutive equations and the motion equation for a prestressed piezoelectric material has been derived. The first part of this section describes the method that was chosen to model a prestress gradient in a piezoelectric plate. Then, the associated formalisms for a piezoelectric plate and film laid down on a substrate are described. In the following, the body force is no longer considered and will be taken as equal to zero.

1. The prestress gradient model and initial configuration

As a general configuration, let us consider a piezoelectric structure, consisting in a film laid down on a substrate, and oriented according to the axes defined in Fig. 2(a). Both film and substrate are assumed to be under a gradient of initial stress. The piezoelectric film and the part of the substrate possessing an initial stress are discretized in a multilayered structure in which each layer is submitted to a uniform initial stress. An illustration is given in Fig. 2(a) where a uniaxial initial mechanical stress $\boldsymbol{\sigma}^0$ along x_1 is represented. In order to optimize the sampling of the stress profile, the initial stress value in each layer is obtained by considering the average of the stress in the corresponding layer thickness as shown in Fig. 2(b).

2. Stiffness matrix method for a prestressed layered structure

A prestressed piezoelectric structure consisting of N piezoelectric layers without a substrate and oriented according to the axes defined in Fig. 2(b) is considered. Piezoelectric Lamb modes are polarized in the (Ox_1, Ox_3) plane and shear horizontal (SH) modes in the (Ox_1, Ox_2) plane. In general, SH modes and Lamb modes can be coupled in piezoelectric structures. In the following, only guided modes with a direction of propagation along the Ox_1 axis will be considered.

For a homogeneous stress in each layer, the motion equation (16) for the layer n expresses as

$$\tilde{C}_{ijkl}u_{k,jl} + \tilde{e}_{mij}\Phi_{,jm} + \boldsymbol{\sigma}^0_{jk}\, u_{i,jk}|_{(n)} = \rho_0 \ddot{u}_i|_{(n)}. \tag{20}$$

This reduces to

$$M_{ijkl}u_{k,jl} + \tilde{e}_{mij}\,\Phi_{,jm}|_{(n)} = \rho_0 \ddot{u}_i|_{(n)}, \tag{21a}$$

where

$$M_{ijkl} = \tilde{C}_{ijkl} + \delta_{ik}\boldsymbol{\sigma}^0_{jl}|_{(n)}. \tag{21b}$$

The electrical displacement $D_i|_{(n)}$ satisfies Poisson's equation (17), which gives

$$\tilde{e}_{ikl}u_{k,li} - \tilde{\varepsilon}_{im}\,\Phi_{,mi}|_{(n)} = 0. \tag{22}$$

If plane wave propagation is assumed, solutions are of the form

$$u_i|_{(n)} = A_i \exp[jk_1(x_1 + \alpha x_3 - Vt)], \quad i = 1,2,3, \tag{23}$$

$$\Phi|_{(n)} = A_4 \exp[jk_1(x_1 + \alpha x_3 - Vt)], \tag{24}$$

where A is the amplitude of the mechanical of electrical potential, V is the phase velocity, and α is the ratio of the wave number component k_3 along x_3 to the wave number component k_1 along x_1.

Substituting Eqs. (23) and (24) into Eqs. (21a) and (22) gives

$$[\Psi_{ik}][A_i] = 0, \quad i,k = 1, \ldots, 4. \tag{25a}$$

The components Ψ_{ik} are given after few manipulations by

$$\Psi_{ik} = M_{i3k3}\alpha^2 + (M_{i1k3} + M_{i3k1})\alpha + M_{i1k1} - \rho V^2 \delta_{ik}, \tag{25b}$$

$$i,k = 1, \ldots, 3,$$

$$\Psi_{i4} = \tilde{e}_{3i3}\alpha^2 + (\tilde{e}_{1i3} + \tilde{e}_{3i1})\alpha + \tilde{e}_{1i1}, \quad i = 1, \ldots, 3, \tag{25c}$$

$$\Psi_{44} = -(\tilde{\varepsilon}_{33}\alpha^2 + (\tilde{\varepsilon}_{13} + \tilde{\varepsilon}_{31})\alpha + \tilde{\varepsilon}_{11}). \tag{25d}$$

Nontrivial solutions are obtained when the determinant of Eq. (25a) is equal to zero. For each value of (k_1, V), it leads to eight solutions for $\alpha(\alpha_p)$ and eight four-component eigenvectors $A^p = (A_1^P, A_2^p, A_3^P, A_4^P)$, $p = 1, \ldots, 8$. However, the eight-degree polynomial equation coefficients in α cannot be

obtained directly as for non-prestressed materials. Indeed, the coefficients M_{ijkl} no longer possess the same symmetry as C_{ijkl}, because of the presence of the initial stress and the unusual symmetry of coefficients $\tilde{C}_{ijkl}$. In particular, it is impossible to use contracted subscripts for M_{ijkl}. However, these coefficients, although extremely long, can be obtained from the determinant of (25a) using a symbolic computation tool and replacing M by its expression in (21b).

The stiffness matrix formalism for a piezoelectric structure requires us to define the general displacement vector $\mathbf{U}=[\mathbf{u},\Phi]^T$ and the general stress vector $\mathbf{T}=[\tilde{\boldsymbol{\Sigma}},D_3]^T$. The elements $[\tilde{\Sigma}_{13},\tilde{\Sigma}_{23},\tilde{\Sigma}_{33}]$ of the generalized stress tensor $\tilde{\boldsymbol{\Sigma}}$ are defined by $\tilde{\Sigma}_{ij}=\tilde{\sigma}_{ij}+u_{i,k}\sigma^0_{jk}$ from the mechanical boundary condition (18). Symbol $(\ ^T)$ denotes the transposition operator.

Thus, Eqs. (14a), (15a), and (25a) allow us to define the solutions for the general displacement vector and the general stress vector for the nth layer. These vectors can be written as linear combinations of the eight partial waves:

$$\begin{aligned}[u_1,u_2,u_3,\Phi]^T_{(n)} = \sum_{p=1}^{4} \{&(1,R_1^{+p},R_2^{+p},R_3^{+p})A_1^{+p}\\ &\times\exp(jk_1\alpha_p^{+}(x_3-Z_{n-1}))\\ &+(1,R_1^{-p},R_2^{-p},R_3^{-p})A_1^{-p}\\ &\times\exp(jk_1\alpha_p^{-}(x_3-Z_n))\}_{(n)}\\ &\times\exp(jk_1(x_1-Vt)),\end{aligned} \tag{26}$$

$$\begin{aligned}[\tilde{\Sigma}_{33},\tilde{\Sigma}_{13},\tilde{\Sigma}_{23},D_3]^T_{(n)} = \sum_{p=1}^{4} jk_1\{&(T_1^{+p},T_2^{+p},T_3^{+p},T_4^{+p})A_1^{+p}\\ &\times\exp(jk_1\alpha_p^{+}(x_3-Z_{n-1}))\\ &+(T_1^{-p},T_2^{-p},T_3^{-p},T_4^{-p})A_1^{-p}\\ &\times\exp(jk_1\alpha_p^{-}(x_3-Z_n))\}_{(n)}\\ &\times\exp(jk_1(x_1-Vt)).\end{aligned} \tag{27}$$

(R_1^p,R_2^p,R_3^p) are the amplitude ratios defined by $R_i^p=A_{i+1}^p/A_1^p$, which depend on the matrix elements Ψ_{ij}, and coefficients T_i^p are explicit functions of R_i^p, α_p, $\boldsymbol{\sigma}^0$ and of the modified initial elastic and piezoelectric parameters $\tilde{C}_{ijkl}$ and $\tilde{e}_{ijk}$. These parameters are given in Appendix A. The parameters with positive and negative superscripts correspond to waves that propagate in the $+x_3$ and $-x_3$ directions, respectively. Eigenvalues associated with forward and backward traveling waves can be sorted on the basis of power flow considerations using the relation given by Havlice,[39] or numerically by computing the vector normal to the slowness curves for each value of phase velocity V.

Considering a layer n, the generalized stress vector on both sides of the layer can be related to the generalized displacement vector:

$$\begin{bmatrix}\mathbf{T}(Z_n)\\ \mathbf{T}(Z_{n-1})\end{bmatrix} = \tilde{\mathbf{K}}^n \begin{bmatrix}\mathbf{U}(Z_n)\\ \mathbf{U}(Z_{n-1})\end{bmatrix}, \tag{28a}$$

where $\tilde{\mathbf{K}}^n$ is the (8×8) stiffness matrix of layer (n) defined by

$$\tilde{\mathbf{K}}^n = \begin{bmatrix}\mathbf{G}^- & \mathbf{G}^+\mathbf{H}^+\\ \mathbf{G}^-\mathbf{H}^- & \mathbf{G}^+\end{bmatrix}\begin{bmatrix}\mathbf{P}^- & \mathbf{P}^+\mathbf{H}^+\\ \mathbf{P}^-\mathbf{H}^- & \mathbf{P}^+\end{bmatrix}^{-1}, \tag{28b}$$

where $\mathbf{P}^{\pm}(4\times4)=[\mathbf{P}_1^{\pm},\mathbf{P}_2^{\pm},\mathbf{P}_3^{\pm},\mathbf{P}_4^{\pm}]$ and $\mathbf{G}^{\pm}(4\times4)=[\mathbf{G}_1^{\pm},\mathbf{G}_2^{\pm},\mathbf{G}_3^{\pm},\mathbf{G}_4^{\pm}]$ are the matrices that contain the generalized polarization vectors and generalized stress amplitude vectors, respectively.

Hence, from Eqs. (26) and (27), the vector components of $\mathbf{P}_p^{\pm}$ and $\mathbf{G}_p^{\pm}$ are given by

$$\mathbf{P}_p^{\pm}=[1,R_1^{\pm p},R_2^{\pm p},R_3^{\pm p}]^T, \quad p=1,\ldots,4 \tag{29}$$

$$\mathbf{G}_p^{\pm}=[T_1^{\pm p},T_2^{\pm p},T_3^{\pm p},T_4^{\pm p}]^T. \tag{30}$$

$\mathbf{H}^{\pm}$ is a diagonal matrix containing the exponential propagation terms of Eqs. (26) and (27) along x_3:

$$\begin{aligned}\mathbf{H} = \mathbf{I}\ \mathrm{diag}[&\exp(jk_1\alpha_1^{\pm}h),\exp(jk_1\alpha_2^{\pm}h),\\ &\times\exp(jk_1\alpha_3^{\pm}h),\exp(jk_1\alpha_4^{\pm}h)],\end{aligned} \tag{31}$$

where h is the thickness of the nth layer and $\mathbf{I}$ is the (4×4) unit matrix.

The total stiffness matrix for a prestressed multilayered piezoelectric structure can now be obtained from the layer stiffness matrices using the recursive relation[34]

$$\tilde{\mathbf{K}}^M = \begin{bmatrix}\tilde{\mathbf{K}}_{11}^m+\tilde{\mathbf{K}}_{12}^m(\tilde{\mathbf{K}}_{11}^{M-1}-\tilde{\mathbf{K}}_{22}^m)^{-1}\tilde{\mathbf{K}}_{21}^m & -\tilde{\mathbf{K}}_{12}^m(\tilde{\mathbf{K}}_{11}^{M-1}-\tilde{\mathbf{K}}_{22}^m)^{-1}\tilde{\mathbf{K}}_{12}^{M-1}\\ \tilde{\mathbf{K}}_{21}^{M-1}(\tilde{\mathbf{K}}_{11}^{M-1}-\tilde{\mathbf{K}}_{22}^m)^{-1}\tilde{\mathbf{K}}_{21}^m & \tilde{\mathbf{K}}_{22}^{M-1}-\tilde{\mathbf{K}}_{21}^{M-1}(\tilde{\mathbf{K}}_{11}^{M-1}-\tilde{\mathbf{K}}_{22}^m)^{-1}\tilde{\mathbf{K}}_{12}^{M-1}\end{bmatrix}, \tag{32}$$

where $\tilde{\mathbf{K}}_{ij}^m$ are the (4×4) stiffness submatrices of layer (m) and $\tilde{\mathbf{K}}_{ij}^{M-1}$ are the (4×4) stiffness submatrices of the $(m-1)$ layers lying below the mth layer.

Using the recursive equation (32) from the bottom layer to the top one, the total stiffness matrix $\tilde{\mathbf{K}}$ of the multilayer relating the generalized displacement vector to the general-

ized stress vector on top and bottom surfaces is expressed as

$$\begin{bmatrix} \mathbf{T}(Z_N) \\ \mathbf{T}(0) \end{bmatrix} = \tilde{\mathbf{K}} \begin{bmatrix} \mathbf{U}(Z_N) \\ \mathbf{U}(0) \end{bmatrix}. \tag{33}$$

Equation (33) allows us to derive the mechanical and electrical boundary conditions. For a multilayer structure in vacuum, the top and bottom surfaces are mechanically free. Hence, from the mechanical boundary conditions (18), the mechanical stress parts of (33) are zero, which implies the general displacement vector to be a function only of the electrical part (electrical displacement) of the general stress vector.[34] Hence, it is possible to define the effective permittivity as

$$\tilde{\varepsilon}_{eff} = \frac{D_3(Z_N) - D_3(0)}{|k_x|\Phi}, \tag{34}$$

where $D_3(Z_N)$ and $D_3(0)$ are the electrical displacements at the top and bottom surface of the multilayer, respectively.

Specifying the electrical boundary conditions at the bottom surface, one can obtain the effective permittivity at the top surface.[34]

$$\tilde{\varepsilon}_{eff} = \varepsilon_0 - \frac{1}{|k_x|(\tilde{S}^{el}_{11} - \tilde{S}^{el}_{12}\tilde{S}^{el}_{21}/\tilde{S}^{el}_{22})} \tag{35}$$

for a metallized bottom surface (short circuit case), or

$$\tilde{\varepsilon}_{eff} = \varepsilon_0 - \frac{1}{|k_x|(\tilde{S}^{el}_{11} - \tilde{S}^{el}_{12}\tilde{S}^{el}_{21}\,/[\,(|k_x|\varepsilon_0)^{-1} + \tilde{S}^{el}_{22})]} \tag{36}$$

for an electrically free bottom surface (open circuit case).[34]

In Eqs. (35) and (36), the (1×1) $\tilde{S}^{el}_{ij}$ coefficients represent the electrical part of the (4×4) compliance submatrices $\tilde{\mathbf{S}}_{ij}$ defined by

$$\tilde{\mathbf{S}}_{ij} = (\tilde{\mathbf{K}}^{-1})_{ij} = \begin{bmatrix} \tilde{\mathbf{S}}^{m}_{ij} & \tilde{\mathbf{S}}^{mel}_{ij} \\ \tilde{\mathbf{S}}^{elm}_{ij} & \tilde{S}^{el}_{ij} \end{bmatrix} \tag{37}$$

The terms $\tilde{\mathbf{S}}^{m}_{ij}$, $\tilde{\mathbf{S}}^{mel}_{ij}$ and $\tilde{\mathbf{S}}^{elm}_{ij}$ are (3×3) mechanical, (3×1) mechanical-electrical coupling, and (1×3) electrical-mechanical coupling submatrices, respectively.

According to the definition of the boundary conditions at the bottom surface of the multilayer, the Lamb and SH modes for short and free electrical boundary conditions at the top surface are obtained from the poles and zeros of the surface permittivity in Eqs. (35) and (36), respectively. It can be noticed that the mixed free-shorted case can be obtained indifferently from the zeros of Eq. (35) or the poles of Eq. (36).

3. Stiffness matrix method for a prestressed piezoelectric multilayer structure on a piezoelectric half space

In order to correspond with practical integrated devices, the case of a multilayered piezoelectric structure laid down on a substrate is now discussed. In practice, the substrate has a thickness considerably larger than that of the film and can be considered as a semi-infinite medium.

For waves propagating in the $(-x_3)$ direction of a bottom half-space, Eq. (28a) implies that the corresponding stiffness matrix must be equal to[34]

$$\tilde{\mathbf{K}}^{-}_{S} = (\mathbf{G}^{-}\mathbf{P}^{-})^{-1} \tag{38}$$

where $\mathbf{G}^{-}$ and $\mathbf{P}^{-}$ are obtained from Eqs. (29) and (30) with the parameters of the substrate.

Following a similar procedure as in the previous section, the total stiffness matrix for a multilayer system on a substrate is defined. The presence of a half-space implies that only the first (4×4) submatrice $\tilde{\mathbf{K}}^{M}_{11}$ of the recursive equation (32) is different from zero.[34] Thus, the recursive procedure to obtain the total stiffness matrix $\tilde{\mathbf{K}}^{M}_{S}$ of the multilayer structure on a substrate reduces to

$$\tilde{\mathbf{K}}^{M}_{S} = \tilde{\mathbf{K}}^{M}_{11} = \tilde{\mathbf{K}}^{m}_{11} + \tilde{\mathbf{K}}^{m}_{12}(\tilde{\mathbf{K}}^{M-1}_{S} - \tilde{\mathbf{K}}^{m}_{22})^{-1}\tilde{\mathbf{K}}^{m}_{21}. \tag{39}$$

Hence, the recursive procedure is the same as in the previous section, starting with the stiffness matrix of the bottom substrate: $\tilde{\mathbf{K}}^{0}_{S} = \tilde{\mathbf{K}}^{-}_{S}$.

If the mechanical stress is zero at the top surface of the structure, Eq. (33) implies that the surface electrical potential must be a function only of the surface electrical displacement. Thus, the surface effective permittivity is given by[34]

$$\tilde{\varepsilon}_{eff} = \varepsilon_0 - \frac{\tilde{K}^{e}_{S} - \tilde{\mathbf{K}}^{elm}_{S}(\tilde{\mathbf{K}}^{m}_{S})^{-1}\tilde{\mathbf{K}}^{mel}_{S}}{|k_x|} \tag{40}$$

where the different components of the total stiffness matrix $\tilde{\mathbf{K}}^{M}_{S}$ are defined as in Eq. (37). The SAW modes for short and free electrical boundary conditions at the top surface are obtained from the poles and zeros of the surface permittivity, respectively.

Finally, with the definitions of the matrices $\tilde{K}^{m}$ and $\tilde{K}^{M}$ given previously, it is also possible to compute the displacement and stress fields inside the plate or inside the film on the substrate by using the relations (30)–(33) given by Wang *et al.*[34] An example will be given in the following.

III. NUMERICAL RESULTS

In this section, effects of an applied stress gradient on Lamb and shear horizontal wave propagation in a lithium niobate plate are analyzed. Then, the behavior of surface acoustic waves in a $LiNbO_3$ film deposited on a Si substrate is discussed. In all calculations, second- and third-order material properties of $LiNbO_3$ and Si were taken from Refs. 40 and 41, respectively. The stress values are given in percentage of the elastic constant C_{11} and extend up to 0.75% which represents a maximum stress of 1.5 GPa. This large value of prestress has already been measured for PzT films[6] and is used in our calculations in a way to obtain significant effects of the prestress gradient on dispersion curves. Two types of stress profiles are studied. The first one is linear and corresponds to a bending stress. The second one is an exponential profile corresponding to a stress decaying in thickness. Indeed, during the fabrication process, surface tractions can appear into the plate. Their effects are then localized in the vicinity of the surfaces. The calculation procedure of the initial strain components due to the applied stress is given in

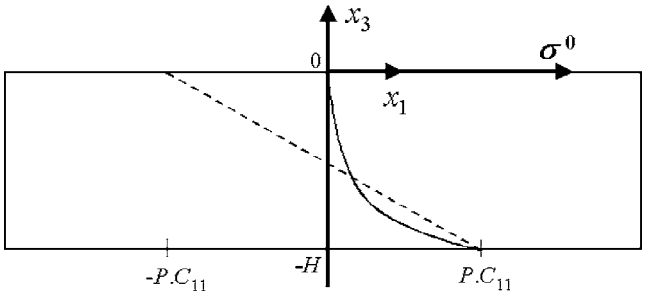


FIG. 3. Geometry of the model for the two studied gradients. Solid lines are for the exponential profile, dashed lines are for the bending profile.

Appendix B. Here, the initial stress has only one component along the x_1 direction, and a zero initial electrical potential is assumed.

The two initial stress gradient profiles are shown Fig. 3 and are given by the following equations:

$$\boldsymbol{\sigma}^0(x_3) = -(2PC_{11}/H)x_3 - PC_{11}, \tag{41}$$

$$\boldsymbol{\sigma}^0(x_3) = \frac{PC_{11}}{\exp(-\beta H) - 1}[1 - \exp(-\beta(x_3 + H))] + PC_{11}, \tag{42}$$

where P is the maximum stress expressed in percentage of C_{11} and H is the thickness of the plate. For the numerical results, P will be equal to 0.75% and β was chosen such that $\beta H = 4$ in order to have a relatively sharp exponential stress profile.

A. Dispersion curves of a prestressed lithium niobate plate

In this section, we will only consider waves propagating in the same direction as the applied stress, i.e., x_1.

In order to obtain a good accuracy on dispersion curves, the number of layers used to discretize the stress gradient has to be high enough. This number depends on several factors: the desired accuracy, the considered guided mode, the frequency, the stress amplitude, and its profile. For both stress profiles, a numerical study on the convergence of the cutoff Lamb and SH wave velocities with the number of layers showed that eight layers are enough to obtain a good accuracy on dispersion curves. Indeed, in this configuration, only variations of a few meters per second were observed for velocities up to 30 000 m/s. Thus, in the following, numerical studies are performed with a number of layers equal to 8 for both stress profiles. The computation time for this configuration will be a few up to tens of hours, depending on the velocity and frequency steps.

In Fig. 4 are plotted the dispersion slowness curves of guided modes propagating in a 1-mm-thick $LiNbO_3$ plate in the case of electrical short-circuit boundary conditions. A slowness representation was chosen in order to better observe the variations of the dispersion curves and the mode cutoff frequencies with the different stresses. These curves were computed for frequencies up to 10 MHz and minima of slownesses of 33×10^{-6} s/m, i.e., corresponding velocities up to 30 000 m/s. The variations of slowness curves in the

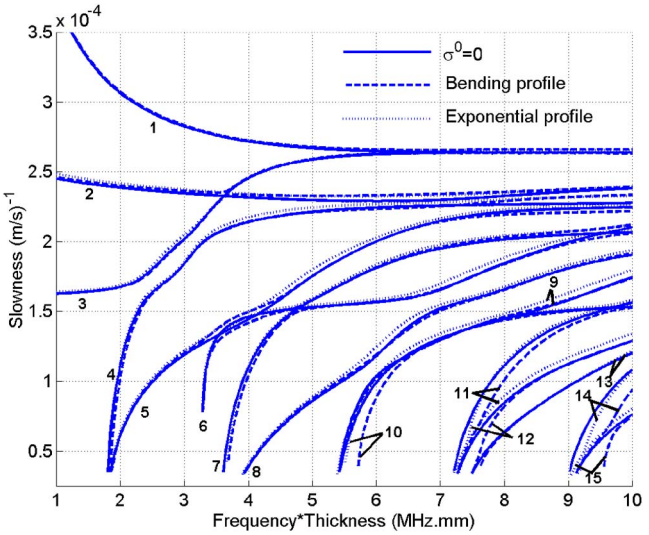


FIG. 4. Dispersion slowness curves for the 1-mm-thick lithium niobate plate.

nonstressed case are compared to the dispersion curves obtained in the case of an applied stress gradient corresponding to the two previous profiles. To facilitate the following discussion, modes are indexed. When no prestress exists in the plate, the classical labeling is used. However, when a stress gradient is applied, modes are no longer symmetric or antisymmetric and the classical labeling of the modes is then no longer valid. Our notations for the modes propagating in a prestressed plate are the following: Fundamental branches are ordered from 1 to 3 corresponding to decreasing slownesses at zero frequency. The upper modes of the prestressed plate are then indexed regarding their cutoff frequency, starting from 4.

The slowness dispersion curves of all the modes exhibit a complex behavior with stress since both increases and decreases of the slownesses of the modes are observed, depending on the mode, the frequency, and the stress profile. For example, it can be seen that the first cutoff mode (labeled 4) has slownesses higher in the case of exponential stress profile than in the prestress free case, whatever the frequency. However, the slownesses for the bending profile are lower (resp. higher) than those in the prestress-free case, for frequencies lower (resp. higher) than 7.2 MHz. Few modes, like the eighth mode at 6 and 8.5 MHz, even exhibit multiple crossing between prestress free and bending profile dispersion curves. The cutoff frequencies of modes 4–8 are not significantly modified, whereas higher modes possess cutoff frequencies that are strongly shifted from their initial value.

Compared to other modes, the first antisymmetric A0 mode is only slightly modified by the prestress, whatever the frequency. S0 and SH0 modes intersect in the stress free case. However, it will be seen in the following paragraph that they fork in the presence of a prestress gradient. Other modes exhibit stronger slowness variations. For example, mode 8 presents a slowness variation of +0.01 s/m, corresponding to a velocity variation of +100 m/s, for the bending profile at 6.6 MHz, whereas the variation is of 22×10^{-4} s/m, corresponding to a velocity variation of −460 m/s, for the exponential profile at 5.8 MHz.

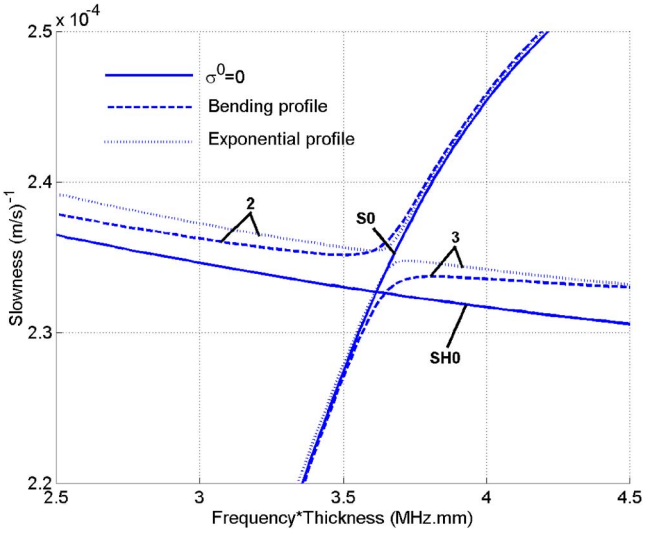


FIG. 5. Local behavior of S0 and SH0 modes in the presence of an initial stress gradient.

A fundamental point is the fact that some modes whose curves cross each other for the prestress free case, do not cross when a gradient of stress is applied. This is particularly the case for modes SH0 and S0 as pointed out in Fig. 5. Around 3.6 MHz, these modes become repulsive: the second (resp. third) mode with stress, behaving like the SH0 (resp. S0) mode before 3.6 MHz, forks toward the slowness of the S0 (resp. SH0) mode after this frequency. The same type of behavior applies to several cutoff modes, for example modes indexed 6 and 7 around 4.8 MHz.

An other key feature of the dispersion curves with a prestress gradient is that the fundamental branches of A0 and S0 possess two different asymptotics (Fig. 6). When the bending or the exponential stress is applied, the first mode corresponding to A0 is shifted toward higher slownesses. Conversely, the second mode, corresponding to S0 at high frequencies, is shifted toward lower slownesses for the bending profile and remains almost unchanged for the exponential one. These results were already observed in literature in the

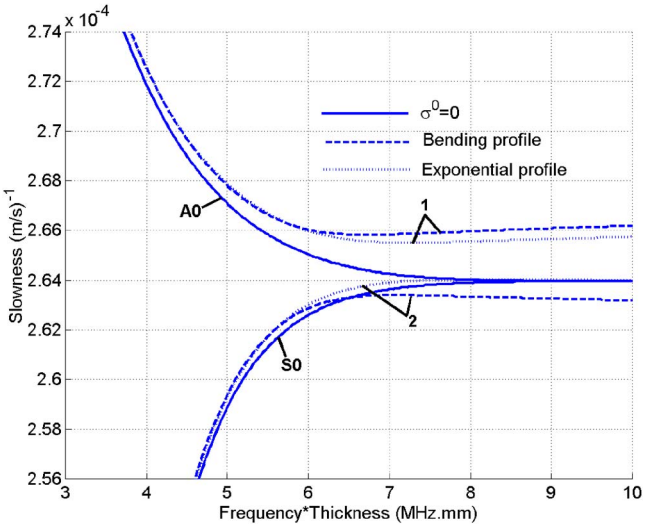


FIG. 6. Evolution of A0 and S0 and corresponding mode slownesses at high frequency.

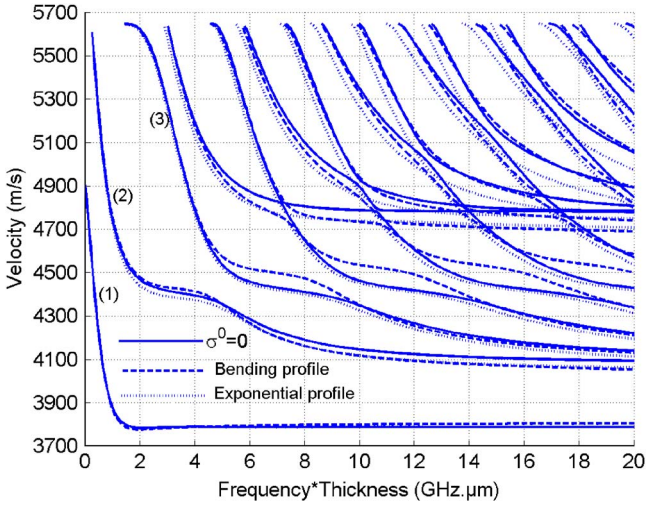


FIG. 7. Dispersion curves of SAW modes for a $LiNbO_3$ film laid down on a Si substrate.

case of isotropic and anisotropic plates with gradients of longitudinal and shear wave velocities[42,43] but they are, to our knowledge, first numerically observed in the case of piezoelectric materials.

Numerical results show that, for frequencies higher than 7 MHz, the slowness of the first mode increases to reach a constant value around 20 MHz. The final variations with the free prestress case are −323 and −259 m/s, respectively, for the bending profile and for the exponential stress profile. The same study for S0 mode shows that it exhibits a maximum variation of about +118 m/s for the bending stress profile. Note that for the exponential stress profile, the asymptotic limit of the second mode is similar to that of the S0 mode.

B. Dispersion curves of a $LiNbO_3$ film on a Si substrate

The influence of a prestress gradient on surface acoustic waves (SAWs) generated in a $LiNbO_3$ film laid down on an oriented Si substrate is studied. The symmetry axes of the film and the substrate are chosen so they coincide. The direction of wave propagation is the same as the direction of the prestress, i.e., x_1. The gradient stress profiles in the film are the same as those in the case of the plate and are given by Eqs. (40) and (41) and Fig. 3. Since the velocity variation domain of the dispersion curves (3700–5700 m/s) is relatively small, a dispersion velocity representation is used. The corresponding SAW velocities are then plotted in Fig. 7 for frequency-thickness products up to 20 GHz.μm in the case of short-circuit boundary conditions (metallized surfaces) of the film. For further commodities, the first three modes are indexed.

As in the case of Lamb and SH waves, the behavior of the modes with the applied stress depends on the stress profile, the mode, and the considered frequency domain. The first SAW mode is almost insensitive to the exponential stress profile whatever the frequency. The first two cutoff frequencies are almost not modified by the presence of the stress whatever the stress profile. For higher order modes, all the cutoff frequencies are shifted toward lower frequencies

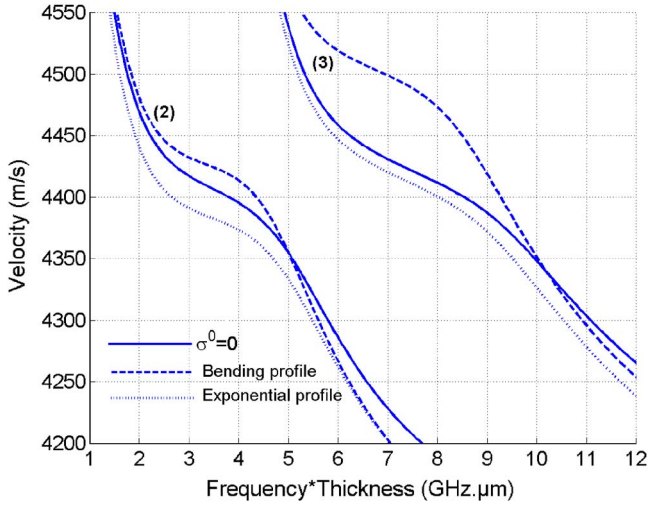


FIG. 8. Local behavior of the first two cutoff SAW modes for a $LiNBO_3$ film on Si substrate.

for the exponential profile and either toward lower or higher frequencies for the bending profile, depending on the mode. All the modes are more sensitive to the bending profile than to the exponential one. This may be due to the fact that the energy of the SAW modes is mainly localized at the top surface of the film. To corroborate this hypothesis, we studied the influence, on SAW modes, of a nonphysical exponential prestress profile similar to the previous one, but having a maximum stress at the free surface of the film. In this configuration, numerical computations showed that the SAW modes exhibit the same sensitivity level than for the bending profile.

The highest sensitivity of the SAW modes for the bending profile allows us to identify it for velocities between approximately 4400 and 4600 m/s. Indeed, in this interval, SAW modes velocities exhibit a strong variation up to about 90 m/s compared to velocities for the free prestress case. An analogous study for free boundary conditions (nonmetallized surfaces) led to a corresponding variations up to about 110 m/s.

Moreover, as the stress is localized in the film, its influence on velocities increases with frequency. Since the penetration depth of the modes become smaller, the properties of the film tend to dominate. When the frequency is high enough, the velocity difference compared to the prestress free case becomes constant for both stress profiles and for a given mode.

For the exponential stress profile, all the modes exhibit velocities that are lower than the corresponding ones in the free prestress case. Conversely, for the bending stress profile, the curves have a more complex behavior, as illustrated in Fig. 8 for the first two cutoff modes indexed 2 and 3. It can be observed that these modes exhibit curves with velocity variations either positive or negative since they cross their corresponding free prestress curves as the frequency-thickness product reaches 5.0 and 10.2 GHz.μm, respectively. As the first cutoff mode is concerned, it can also be noticed that the two stress gradient profiles can be well discriminated only for frequency-thickness products between

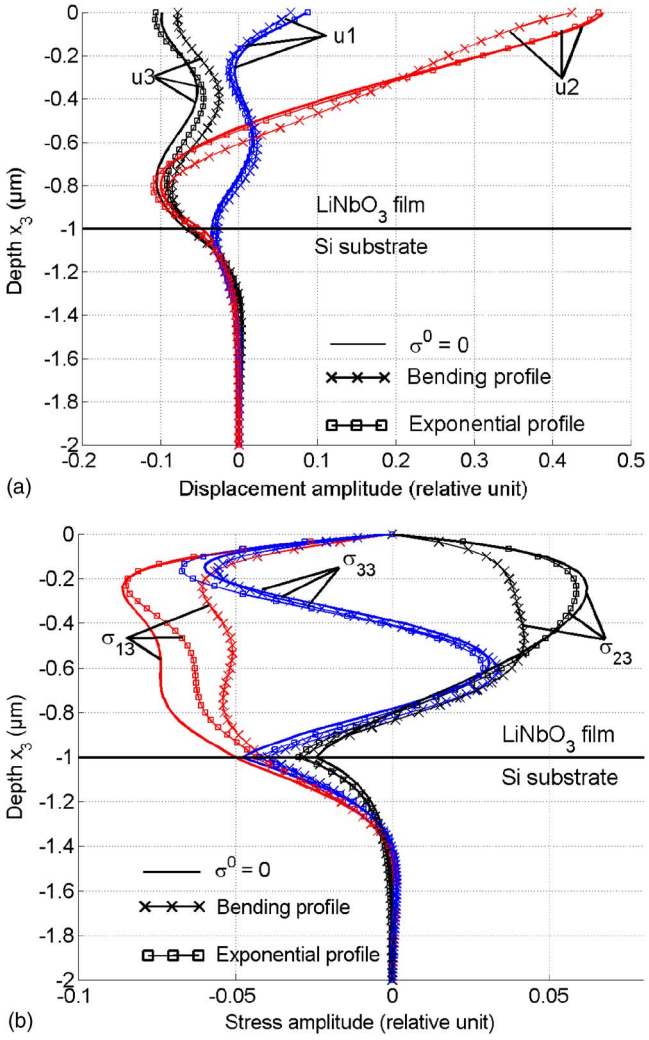


FIG. 9. Displacement (a) and stress (b) fields for a 1-μm-thick $LiNbO_3$ film on a Si substrate at a frequency of 7 GHz for mode 3.

about 1.5 to 6 GHz.μm, since the two corresponding dispersion curves tend to be superimposed outside this interval.

Computations of SAW modes dispersion curves have also been performed taking into account an exponential decaying stress gradient in the Si substrate. Compared to the case where the prestress localized only in the film, the phase velocity variations are only a few meters per second. This is due to the fact that the SAW modes are essentially localized into the film. As an example, Fig. 9(a) shows a strong decrease in the substrate of the three components of the displacement field of mode 3 propagating, at 7 GHz, in a 1 -μm-thick lithium niobate film on silicon substrate. Figure 9(b) shows a similar evolution of the three components of the stress field involved in the dispersion equation: $\boldsymbol{\sigma}_{33}$, $\boldsymbol{\sigma}_{13}$, $\boldsymbol{\sigma}_{23}$. Note that, because the sagittal plane (x_1,x_3) is not a plane of symmetry, the mode has three significant displacement components. However, at this specific frequency, mode 3 is close to a SH mode because of its predominant transverse, u_2, component. As expected, the influence of the exponential stress profile on the displacement and stress fields of mode 3 is predominant in the vicinity of the interface between the film and the substrate.

Studies were carried out to highlight the influence of the

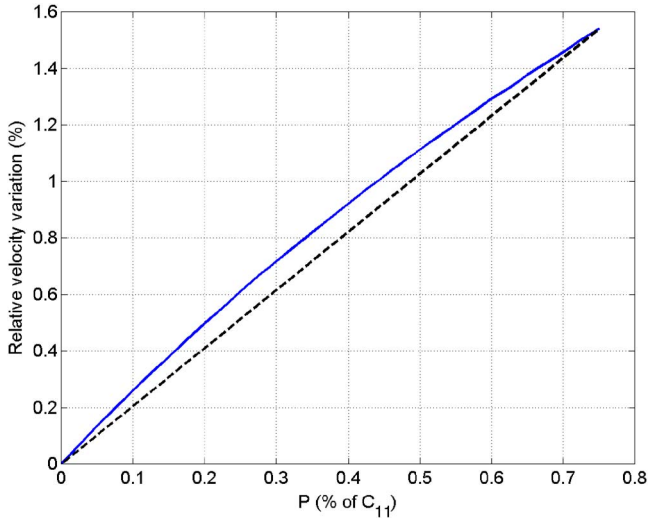


FIG. 10. Relative velocity variation with the stress amplitude level of the bending profile for mode 3 at a frequency-thickness product equal to 7 GHz.μm (solid line). Comparison with a linear variation (dashed line).

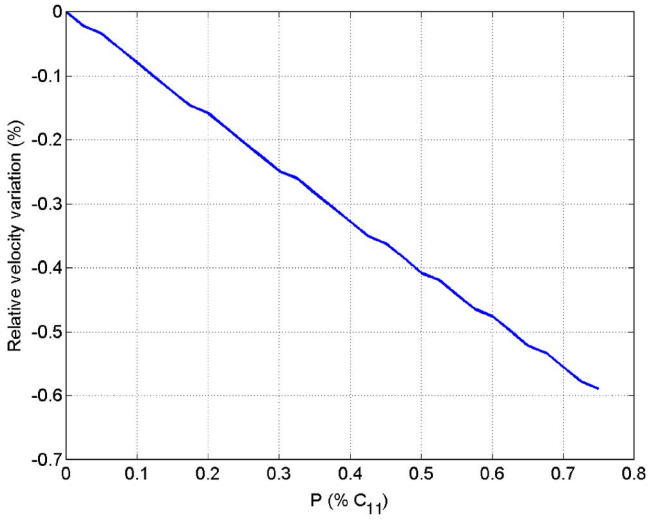


FIG. 11. Relative velocity variation with the stress amplitude level of the exponential profile for mode 2 at a frequency-thickness product equal to 3 GHz.μm.

level of stress amplitude (parameter P) of both stress profiles and of the degree of curvature of the exponential profile (parameter β). They revealed that the behavior of the dispersion curves of SAW with these parameters depend slightly on the considered mode and on the frequency-thickness product. Illustrations of their effect are given in regions where each prestress profile had the biggest influence.

Figure 10 shows the effects of P parameter on SAW mode 3, at a frequency-thickness product equal to 7 GHz.μm, in the case of a bending prestress profile. The relative variation of SAW velocity with the level of stress amplitude, P, is slightly nonlinear compared to the linear variation represented by the dashed line. Second-order polynomial fit was used to characterize the degree of nonlinearity. It led to $y=-0.71x^2+2.57x$. The same study was carried out for several modes at different frequencies, revealing the similar behavior regardless of the domain of variation corresponding to the mode sensitivity.

Considering the exponential profile, the influence of P at constant $\beta H=4$ on SAW mode 2 was studied, at a frequency-thickness product equal to 3 GHz.μm (Fig. 11). The relative variation of SAW velocity with the level of stress amplitude P for the exponential profile is almost linear.

Figure 12 shows that the relative variation of SAW velocity with the degree of curvature β of the exponential profile is nonlinear. In this case, the value of the parameter P was kept constant and equal to 0.75% of C_{11}. Second-order polynomial fit was used to characterize its degree of nonlinearity. It led to $y=-0.029x^2+0.39x$.

Since the velocity values for the exponential profile are always lower than for the non-prestressed case, the sign of the relative velocity variation cannot change whatever the mode and the frequency.

IV. CONCLUSIONS

In this paper, acoustical effects of an applied external stress gradient on the propagation of guided waves in a piezoelectric structure were analyzed. First, the constitutive equations of a prestressed piezoelectric material were derived to take into account the initial displacement field and initial applied stress. These modified constitutive equations were then used to derive a formalism of Lamb, SH mode, and SAW propagation in a piezoelectric structure submitted to an external stress gradient. In particular, it was shown how the recursive stiffness matrix method has to be modified to take into account a gradient of stress in a piezoelectric structure.

In the first part, numerical results were carried out on lithium niobate piezoelectric single crystal for a bending and an exponential stress profile. They highlight the sensitivity of Lamb and SH modes to an external stress gradient. All the modes exhibit a complex variation of slowness dispersion curves with stress gradient since both increases and decreases of the slownesses of the modes are observed, depending on the mode, the frequency, and the stress profile. Few

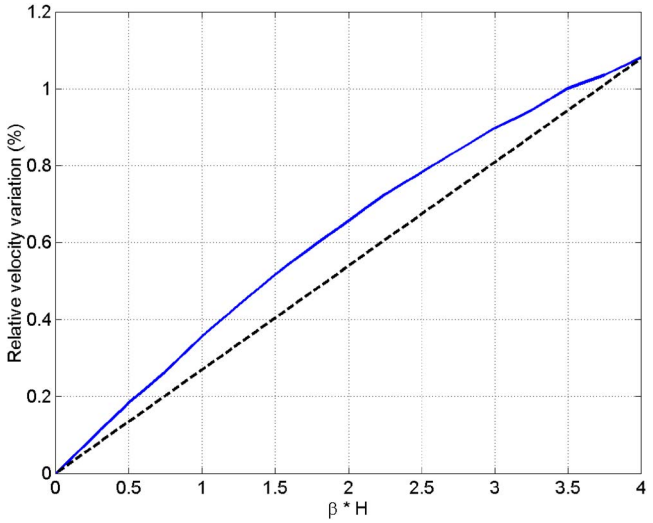


FIG. 12. Relative velocity variation with the curvature level of the exponential profile for mode 2 at a frequency-thickness product equal to 3 GHz.μm (solid line). Comparison with a linear variation (dashed line).

modes have a positive or negative variation of phase velocity compared to the free prestress case. We have shown that the asymptotic limits of the first and second fundamental modes are different in the presence of an initial stress gradient than those of A0 and S0. In presence of stress, the first mode is shifted toward higher slownesses, whereas the second mode is shifted toward lower values. Furthermore, some modes whose curves cross each other for the prestress free case become repulsive and do not cross when a stress gradient is applied. This is particularly the case for the second and third modes whose curves fork at low frequencies.

In the last part, a lithium niobate film submitted to a gradient of stress and laid down on a silicon substrate was studied. As in the case of Lamb and SH waves, the behavior of the SAW modes with the applied stress is complex and depends on the stress profile, the mode, and the considered frequency domain. Depending on frequency, several SAW modes have positive or negative phase velocity variations when compared to their corresponding prestress free case, while others exhibit only positive or only negative velocity variations. SAW modes are more sensitive to the bending stress profile than to the exponential stress profile. Numerical studies also reveal that SAW modes are insensitive to a gradient of stress localized in the substrate, since the velocity variations are only a few meters per second.

The numerical results showed that it is possible to discriminate between the two profiles by studying the velocity variations or the shifts of the cutoff frequencies of selected modes. Moreover, since the number of sampling layers required to obtain a good accuracy of dispersion velocities is relatively low, the computation time remains reasonable.

Considering future works, experimental characterization of prestressed structure should contribute to explain the physical behavior of SAW. Results are given for a lithium niobate plate or a lithium niobate film laid down on a silicon substrate, but further studies can be carried out to evaluate the sensitivity, to a prestress gradient, of guided modes propagating in other material structures. Finally, the formalism presented here takes into account a tensor of applied stress that allows us to characterize the influence of an applied stress gradient with spatial components. This method can be also simplified in the case of non-piezoelectric materials, in order to study stress gradients in metallic laminates.

ACKNOWLEDGMENT

This work was funded by the EC Minuet Project, Contract No. NMP2-CT-2004-505657.

APPENDIX A: COMPONENTS OF THE MODIFIED GENERAL DISPLACEMENT AND STRESS VECTORS

With the use of the definitions (25b)–(25d) of Ψ_{ik} calculated for each α_p, the corresponding expressions of R_i^p of Eq. (26) are

$$R_1^p = \frac{\Psi_{11}(\Psi_{33}\Psi_{24} - \Psi_{23}\Psi_{34}) + \Psi_{12}(\Psi_{13}\Psi_{34} - \Psi_{14}\Psi_{33}) + \Psi_{13}(\Psi_{14}\Psi_{23} - \Psi_{13}\Psi_{24})}{\Psi_{12}(\Psi_{23}\Psi_{34} - \Psi_{24}\Psi_{33}) + \Psi_{13}(\Psi_{23}\Psi_{24} - \Psi_{22}\Psi_{34}) + \Psi_{14}(\Psi_{22}\Psi_{33} - \Psi_{23}\Psi_{23})}, \tag{A1}$$

$$R_2^p = \frac{\Psi_{11}(\Psi_{22}\Psi_{34} - \Psi_{24}\Psi_{23}) + \Psi_{12}(\Psi_{13}\Psi_{24} - \Psi_{12}\Psi_{34}) + \Psi_{14}(\Psi_{12}\Psi_{23} - \Psi_{13}\Psi_{22})}{\Psi_{12}(\Psi_{23}\Psi_{34} - \Psi_{24}\Psi_{33}) + \Psi_{13}(\Psi_{23}\Psi_{24} - \Psi_{22}\Psi_{34}) + \Psi_{14}(\Psi_{22}\Psi_{33} - \Psi_{23}\Psi_{23})}, \tag{A2}$$

$$R_3^p = \frac{\Psi_{11}(\Psi_{23}\Psi_{23} - \Psi_{22}\Psi_{33}) + \Psi_{12}(\Psi_{12}\Psi_{33} - \Psi_{13}\Psi_{23}) + \Psi_{13}(\Psi_{13}\Psi_{22} - \Psi_{12}\Psi_{23})}{\Psi_{12}(\Psi_{23}\Psi_{34} - \Psi_{24}\Psi_{33}) + \Psi_{13}(\Psi_{23}\Psi_{24} - \Psi_{22}\Psi_{34}) + \Psi_{14}(\Psi_{22}\Psi_{33} - \Psi_{23}\Psi_{23})}. \tag{A3}$$

With the use of Eqs. (14a) and (18) for the dynamic electrical displacement and dynamic stress tensor for the prestressed state, the expressions of T_i^p of Eq. (27) are

$$T_1^p = \tilde{C}_{3311} + \tilde{C}_{3313}\alpha_p + (\tilde{C}_{3321} + \tilde{C}_{3323}\alpha_p)R_1^p + [\tilde{C}_{3331} + \sigma_{31}^0 + (\tilde{C}_{3333} + \sigma_{33}^0)\alpha_p]R_2^p + (\tilde{e}_{133} + \tilde{e}_{333}\alpha_p)R_3^p \tag{A4}$$

$$T_2^p = \tilde{C}_{1311} + \boldsymbol{\sigma}_{31}^0 + (\tilde{C}_{1313} + \boldsymbol{\sigma}_{33}^0)\alpha_p + (\tilde{C}_{1321} + \tilde{C}_{1323}\alpha_p)R_1^p + (\tilde{C}_{1331} + \tilde{C}_{1333}\alpha)R_2^p + (\tilde{e}_{113} + \tilde{e}_{313}\alpha_p)R_3^p, \tag{A5}$$

$$T_3^p = \tilde{C}_{2311} + \tilde{C}_{2313}\alpha_p + [\tilde{C}_{2321} - \boldsymbol{\sigma}_{31}^0 + (\tilde{C}_{2323} + \boldsymbol{\sigma}_{33}^0)\alpha_p]R_1^p + (\tilde{C}_{2331} + \tilde{C}_{2333}\alpha_p)R_2^p + (\tilde{e}_{123} + \tilde{e}_{323}\alpha_p\alpha_p)R_3^p, \tag{A6}$$

$$T_4^p = \tilde{e}_{311} + \tilde{e}_{313}\alpha_p + (\tilde{e}_{321} + \tilde{e}_{323}\alpha_p)R_1^p + (\tilde{e}_{331} + \tilde{e}_{333}\alpha_p)R_2^p + (\tilde{\varepsilon}_{31} + \tilde{\varepsilon}_{33}\alpha_p)R_3^p, \tag{A7}$$

APPENDIX B: CALCULATION PROCEDURE OF THE INITIAL STRAIN COMPONENTS

All calculations carried out in the paper assume that the initial stresses are constant along the thickness of each layer and that no initial external electrical field is applied between electrodes. Since the initial strains represent finite deformations compared to the dynamic ones, the generalized Hooke law defined up to the second-order strains has to be used:

$$\sigma_{ij}^0 = C_{ijkl}S_{kl}^0 + \tfrac{1}{2}C_{ijklmn}S_{kl}^0S_{mn}^0. \tag{B1}$$

An optimization tool is required to find the initial strain components. Here, the initial guess of strain components required

for the optimization algorithm is found by first solving only the linear part of (B1).

As a specific case, the initial displacement is assumed to be independent of x_2: $u_i^0 = u_i^0(x_1, x_3)$, $i = 1, 2, 3$. Thus, the strain components $u_{i,2}^0$ are equal to zero and the other nonvanishing terms are given by

$$u_{1,1}^0 = S_{11}^0,$$

$$u_{3,3}^0 = S_{33}^0,$$

$$u_{2,1}^0 = 2S_{12}^0,$$

$$u_{2,3}^0 = 2S_{23}^0,$$

$$u_{1,3}^0 + u_{3,1}^0 = 2S_{13}^0. \tag{B2}$$

Since no added restriction exists for the last equation, it can be considered that $u_{1,3}^0$ and $u_{3,1}^0$ are of the same order: $u_{1,3}^0 \cong u_{3,1}^0 \cong S_{13}^0$.

[1] J. M. Herbert, *Ferroelectric Transducers and Sensors* (Gordon and Breach, New York, 1982).
[2] K. A. Snook, J. Z. Zhao, C. H. F. Alves, J. M. Cannata, W. H. Chen, R. J. Meyer, T. A. Ritter, and K. K. Shung, "High frequency transducers for medical ultrasonic imaging," Proc. SPIE **3982**, 92–99 (2000).
[3] J. M. Cannata, T. A. Ritter, W.-H. Chen, and K. K. Shung, "Design of focused single element (50–100 MHz) tranducers using lithium niobate," Proc. IEEE Ultrason. Symp. (2000).
[4] L. Lian and N. R. Sottos, "Stress effects in sol-gel derived ferroelectric thin films," J. Appl. Phys. **95**, 629–634 (2004).
[5] A. Cimpoiasu, N. M. Van Der Pers, T. H. de Keyser, A. Venema, and M. J. Vellekoop, "Stress control of piezoelectric ZnO films on silicon substrates," Smart Mater. Struct. **5**, 744–750 (1996).
[6] N. R. Sottos, T. Berfield, R. Ong, and D. A. Payne, "Residual stress effects on ferroelectric thin film patterning, properties and performance," XXI ICTAM, Poland (2004).
[7] G. A. Maugin, J. Pouget, R. Drouot, and B. Collet, *Nonlinear Electromechanical Couplings* (Wiley, Chichester, 1992).
[8] Q. M. Zhang and J. Zhao, "Electromechanical properties of lead zirconate titanate piezoceramics under the influence of mechanical stresses," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 1518–1526 (1999).
[9] G. Yang, S.-F. Liu, W. Ren, and B. K. Mukherjee, "Effects of uniaxial stress on the piezoelectric, dielectric, and mechanical properties of lead zirconate titanate piezoceramics," 3rd Asian Meet. Ferr. (AMF3), Hong Kong (2000).
[10] G. Yang, S.-F. Liu, W. Ren, and B. K. Mukherjee, "Uniaxial stress dependence of the piezoelectric properties of lead zirconate titanate ceramics," Proc. SPIE **3992**, 103–113 (2000).
[11] A. H. Nayfeh, "The general problem of elastic wave propagation in multilayered anisotropic media," J. Acoust. Soc. Am. **89**, 1521–1531 (1991).
[12] A. H. Nayfeh and H.-T. Chien, "The influence of piezoelectricity on free and reflected waves from fluid-loaded anisotropic plates," J. Acoust. Soc. Am. **91**, 1250–1261 (1992).
[13] P. Heyliger and D. A. Saravanos, "Exact free-vibration analysis of laminated plates with embedded piezoelectric layers," J. Acoust. Soc. Am. **98**, 1547–1557 (1995).
[14] B. K. Sinha, "Elastic waves on crystals under a bias," Ferroelectrics **41**, 61–73 (1982).
[15] B. K. Sinha, W. J. Tanski, T. Lukaszek, and A. Ballato, "Influence of biasing stresses on the propagation of surface waves," J. Appl. Phys. **57**, 767–776 (1985).
[16] C. Desmet, U. Kawald, A. Mourad, W. Lauriks, and J. Thoen, "The behaviour of Lamb waves in stressed polymer foils," J. Acoust. Soc. Am. **100**, 1509–1513 (1996).
[17] M. A. Dowaikh, "On SH waves in a pre-stressed layered half space for an incompressible elastic material," Mech. Res. Commun. **26**, 665–672 (1999).
[18] M. O. Si-Chaib, H. Djelouah, and T. Boutkedjirt, "Propagation of ultrasonic waves in materials under bending forces," NDT & E Int. **38**, 283–289 (2005).
[19] L. Palmieri, G. Socino, and E. Verona, "Electroelastic effect in layer acoustic mode propagation along ZnO films on Si substrates," Appl. Phys. Lett. **49**, 1581–1583 (1986).
[20] A. Palma, L. Palmieri, G. Socino, and E. Verona, "Acoustic Lamb wave-electric field nonlinear interaction in YZ $LiNbO_3$ plates," Appl. Phys. Lett. **46**, 25–27 (1985).
[21] S. G. Joshi, "Electronically variable time delay in ultrasonic Lamb wave delay lines," Proc.-IEEE Ultrason. Symp., 893–896 (1996).
[22] H. Liu, T. J. Wang, Z. K. Wang, and Z. B. Kuang, "Effect of a biasing electric field on the propagation of symmetric Lamb waves in piezoelectric plates," Int. J. Solids Struct. **39**, 2031–2049 (2002).
[23] J. Yang, "Free vibrations of an electroelastic body under biasing fields," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 358–364 (2005).
[24] H. Liu, Z. B. Kuang, and Z. M. Cai, "Application of transfer matrix method in analysing the inhomogeneous initial stress problem in pre-stressed layered piezoelectric media," IUTAM Symp. Dyn. Advanc. Mat. & Smart Struct. (2003), pp. 263–272.
[25] M. Lematre, G. Feuillard, T. Delaunay, and M. Lethiecq, "Modelling of ultrasonic wave propagation in integrated piezoelectric structures under pre-stress," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 685–696 (2006).
[26] W. T. Thomson, "Transmission of elastic waves through a stratified solid medium," J. Appl. Phys. **21**, 89–93 (1950).
[27] N. A. Haskell, "The dispersion of surface waves in multilayered media," Bull. Seismol. Soc. Am. **43**, 17–34 (1953).
[28] D. Levesque and L. Piche, "A robust transfer matrix formulation for the ultrasonic response of multilayered absorbing media," J. Acoust. Soc. Am. **92**, 452–467 (1992).
[29] M. Castaings and B. Hosten, "Delta operator technique to improve the Thomson-Haskel method stability for propagation in multilayered anisotropic absorbing plates," J. Acoust. Soc. Am. **95**, 1931–1941 (1993).
[30] T. Pastureaud, V. Laude, and S. Ballandras, "Stable scattering matrix method for surface acoustic waves in piezoelectric multilayers," Appl. Phys. Lett. **80**, 2544–2547 (2002).
[31] B. Honein, A. M. B. Braga, P. Barbone, and G. Hermann, "Wave propagation in piezoelectric layered media with some applications," J. Intell. Mater. Syst. Struct. **2**, 542–557 (1991).
[32] S. I. Rokhlin and L. Wang, "Stable recursive algorithm for elastic wave propagation in layered anisotropic media: Stiffness matrix method," J. Acoust. Soc. Am. **112**, 822–834 (2002).
[33] L. Wang and S. I. Rokhlin, "Stable reformulation of transfer matrix method for wave propagation in layered anisotropic media," Ultrasonics **39**, 413–424 (2001).
[34] L. Wang and S. I. Rokhlin, "A compliance/stiffness matrix formulation of general Green's function and effective permittivity for piezoelectric multilayers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 453–463 (2004).
[35] L. Wang and S. I. Rokhlin, "Recursive geometric integrators for wave propagation in a functionally graded multilayered elastic medium," J. Mech. Phys. Solids **52**, 2473–2506 (2004).
[36] H. F. Tiersten, "Electroelastic interactions and the piezoelectric equations," J. Acoust. Soc. Am. **70**, 1567–1576 (1981).
[37] H. F. Tiersten, "Electroelastic interactions, biasing states, and precision crystal resonators," J. Acoust. Soc. Am. **85**, S7–S8 (1989).
[38] Z. B. Kuang, "Propagation of Bleustein-Gulyaev waves in a pre-stressed layered piezoelectric structure," Ultrasonics **41**, 397–405 (2003).
[39] J. F. Havlice, W. L. Bond, and L. B. Wighton, "Elastic Poynting vector in piezoelectric medium," IEEE Trans. Sonics Ultrason. **SU-17** (1970).
[40] Y. Cho and K. Yamanouchi, "Nonlinear, elastic, piezoelectric, electrostrictive, and dielectric constants of lithium niobate," J. Appl. Phys. **61**, 875–887 (1987).
[41] P. N. Keating, "Theory of the Third-Order Elastic Constants of Diamond-Like Crystals," Phys. Rev. **149**, 674–678 (1966).
[42] P. J. Shull, D. E. Chimenti, and S. K. Datta, "Elastic guided waves and the Floquet concept in periodically layered plates," J. Acoust. Soc. Am. **95**, 99–108 (1994).
[43] C. Baron, O. Poncelet, and A. Shuvalov, "Calculation of the velocity spectrum of the vertically inhomogeneous plates," WCU Symposium (2003).

Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction

Mingsian R. Bai[a] and Chih-Chung Lee
Department of Mechanical Engineering, National Chiao-Tung University, 1001 Ta-Hsueh Road, Hsin-Chu 300, Taiwan

(Received 2 December 2005; revised 27 June 2006; accepted 29 June 2006)

A comprehensive study was conducted to explore the effects of listening angle on crosstalk cancellation in spatial sound reproduction using two-channel stereo systems. The intention is to establish a sustainable configuration of crosstalk cancellation system (CCS) that best reconciles the separation performance and the robustness against lateral head movement, not only in theory but also in practice. Although crosstalk can in principle be suppressed using multichannel inverse filters, the CCS does not lend itself very well to practical application owing to the fact that the sweet spot is being so small. Among the parameters of loudspeaker deployment, span angle is a crucial factor that has a profound impact on the separation performance and sweet spot robustness achievable by the CCS. This paper seeks to pinpoint, from a more comprehensive perspective, the optimal listening angle that best reconciles the robustness and performance of the CCS. Two kinds of definitions of sweet spot are employed for assessment of robustness. In addition to the point source model, head related transfer functions (HRTF) are employed as the plant models in the simulation to emulate more practical localization scenarios such as the high-frequency head shadowing effect. Three span angles including 10, 60, and 120 deg are then compared via objective and subjective experiments. The Friedman test is applied to analyze the data of subjective experiments. The results indicate that not only the CCS performance but also the panning effect and head shadowing will dictate the overall performance and robustness. The 120-deg arrangement performs comparably well as the standard 60-deg arrangement, but is much better than the 10-deg arrangement. [DOI: 10.1121/1.2257986]

PACS number(s): 43.38.Md, 43.60.Tj, 43.60.Pt [AJZ] Pages: 1976–1989

I. INTRODUCTION

The central idea of spatial audio reproduction is to synthesize a virtual sound image. The listener perceives as if the signals reproduced at the listener's ears would have been produced by a specific source located at an intended position.[1,2] This attractive feature of spatial audio lends itself to an emerging audio technology with promising application in mobile phone, personal computer multimedia, video games, home theater, etc.

The rendering of spatial audio is either by headphones or loudspeakers. Headphones reproduction is straightforward, but suffers from several shortcomings such as in-head localization, front-back reversal, and discomfort to wear. While loudspeakers do not have the same problems as the headphones, another issue adversely affects the performance of spatial audio rendering using loudspeakers. The issue frequently encountered in loudspeaker reproduction is the crosstalk in the contralateral paths from the loudspeakers to the listener's ears, which may obscure source localization. To overcome the problem, crosstalk cancellation systems (CCS) that seek to minimize, if not totally eliminate, the crosstalks have been studied extensively by researchers.[3–8] Various inverse filtering approaches were suggested for designing multichannel prefilters for CCS.

Notwithstanding the preliminary success of CCS in an academic community, a problem seriously hampers the use of CCS in practical applications. The problem stems from the limited size of the so-called "sweet spot" in which CCS remains effective. The sweet spots are generally so small especially at the lateral side that a head movement of a few centimeters would completely destroy the cancellation performance. Two kinds of approaches can be used to address this problem—the adaptive design and the robust design. An example of adaptive CCS with head-tracker was presented in the work of Kyriakakis *et al.*[9,10] This approach dynamically adjusts the CCS filters by tracking the head position of the listener using optical or acoustical sensors. However, the approach has not been widely used because of the increased hardware and software complexity of the head tracker. On the other hand, instead of dynamically tracking the listener's head, an alternative CCS design using fixed filters can be taken to create a "widen" sweet spot that accommodates larger head movement. Ward and Elko in Bell Labs have conducted a series of insightful analysis of the robustness issue of CCS. In their paper[11] on this topic in 1998, robustness of a two-channel stereo loudspeaker (2×2) CCS was investigated using weighted cancellation performance measure at the pass zone and stop zone, respectively. In the other paper[12] by the same authors in 1999, robustness issue of a 2×2 CCS was revisited using a different measure, the condition number, which focuses more on numerical stability during matrix inversion, in the presence of noise in data

[a]Elecrtonic mail: msbai@mail.nctu.edu.tw

and/or perturbations to system properties. Yet, in another paper[13] by Ward, a joint least squares optimization method is employed to obtain a CCS that is robust to head misalignment. The above-mentioned research winds up with a simple but important conclusion that the optimal loudspeaker spacing should be inversely proportional to the operating frequency. Along the line of robust CCS design, a celebrated "stereo dipole", configuration was suggested by Kirkeby, Nelson, and Hamada[14] and Takeuchi and Nelson.[15] In their arrangement, two loudspeakers are closely spaced with only a 10° span. Their analysis of robustness of CCS also focused primarily on numerical stability in relation to the errors in matrix inversion. The consistent finding of these studies was that the optimal loudspeaker spacing is inversely proportional to the operating frequency. Since the optimal spacing is frequency dependent, a multidrive configuration of the optimal source distribution (OSD) system,[16] comprising pairs of loudspeakers with various spacings, was suggested to deal with crosstalks for different frequency bands. Another multidrive CCS design was also developed by Bai *et al.*[18] based on the genetic algorithm and array signal processing. Their approach requires no crossover circuits as in the OSD system.

According to Gardner,[19] loudspeakers spaced apart tend to yield a smaller equalization zone than loudspeakers spaced closely. However, the improvement is predominantly along the front-back axis and the equalization zone widens only slightly when the speakers are positioned closely together. One disadvantage of close spacing is the lack of natural high frequency separation due to head shadowing. Another problem is that small head rotation will cause both speakers to fall on the same side so that the panning mechanism fails.

Thus far, there have been pros and cons in the closely spaced CCS. The question of which kind of loudspeaker arrangement is the best has been puzzling people for quite some time. It is worth exploring further the underlying physical insights from all possible angles. This motivates the current research to undertake a comprehensive study in a hope to resolve this optimal CCS problem more conclusively. In Gardner's work,[19] the head-related transfer functions (HRTF) were measured in the MIT Media Lab[20,21] and subjective listening tests were conducted. However, only the crosstalk below 6 kHz was considered to result in a band-limited CCS design. Furthermore, the robustness of CCS to head misalignment were discussed in depth by Takeuchi and Nelson.[15] In both works, only two listening spans including 10- and 60-deg spans were investigated. On the other hand, the emphasis of this paper is placed on the analysis of the effects of listening angle on CCS in terms of not only robustness but also performance. There are several special features in this paper. First, not only the robustness but also the performance of CCS is examined with the aid of a more comprehensive set of indices. Second, two kinds of definitions of sweet spot are employed for assessment of robustness. Third, the present work considers the entire audible 20 kHz band in which the listener's head may provide natural separation for certain loudspeaker arrangements. Fourth, apart from the objective physical tests, subjective listening tests are conducted to practically assess the CCS arrangements with different listening angles. The results of subjective tests will be validated by using the Friedman test. Although the last three points have been investigated in Refs. 15 and 19, this study examines the design issues in further detail and in some cases reaches different conclusions than the previous research. The intention is to establish a sustainable configuration of CCS that best reconciles the separation performance and the robustness against lateral head movement, not only in theory but also in practice.

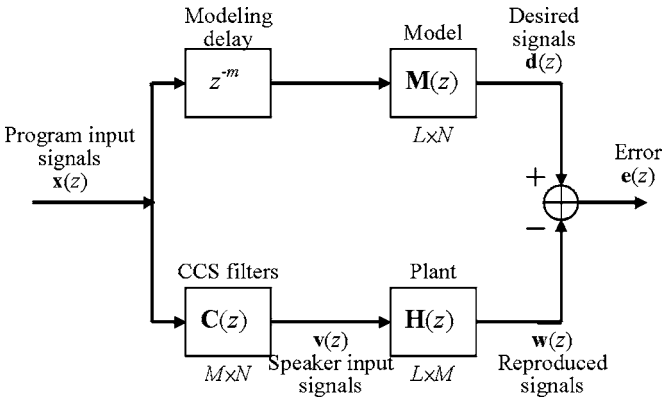


FIG. 1. The block diagram of a multichannel model-matching problem in the CCS design.

II. MULTICHANNEL INVERSE FILTERING FOR CCS FROM A MODEL-MATCHING PERSPECTIVE

The CCS aims to cancel the crosstalk in the contralateral paths from the multichannel loudspeakers to the listener's ears so that the binaural signals are reproduced at two ears like those reproduced using headphones. This problem can be viewed from a model-matching perspective, as shown in Fig. 1. In the block diagram, $\mathbf{x}(z)$ is a vector of N program input signals, $\mathbf{v}(z)$ is a vector of M loudspeaker input signals, and $\mathbf{e}(z)$ is a vector of L error signals. $\mathbf{M}(z)$ is a $L\times N$ matrix of matching model, $\mathbf{H}(z)$ is a $L\times M$ plant transfer matrix, and $\mathbf{C}(z)$ is a $M\times N$ matrix of the CCS filters. The z^{-m} term accounts for the modeling delay to ensure causality of the CCS filters. Let us neglect the modeling delay for the moment, it is straightforward to write down the input-output relationship

$$\mathbf{e}(z)=[\mathbf{M}(z)-\mathbf{H}(z)\mathbf{C}(z)]\mathbf{u}(z). \tag{1}$$

For arbitrary inputs, minimization of the error output is tantamount to the following optimization problem:

$$\min_{\mathbf{C}}\|\mathbf{M}-\mathbf{HC}\|_F^2, \tag{2}$$

where F symbolizes the Frobenius norm.[22] For a $L\times N$ matrix $\mathbf{A}$, Frobenius norm is defined as

$$\|\mathbf{A}\|_F^2=\sum_{n=1}^{N}\sum_{l=1}^{L}|a_{ln}|^2$$

$$=\sum_{n=1}^{N}\|\mathbf{a}_n\|_2^2,\ \mathbf{a}_n \text{ begin the } n\text{th column of } \mathbf{A}. \tag{3}$$

Hence, the minimization problem of Frobenius norm can be

converted to the minimization problem of two norm by partitioning the matrices into columns. Assume that **H** is of full column rank and there is no coupling between the columns of the resulting matrix **C** which approximates the inverse of **H**, the minimization of the square of the Frobenius norm of the entire matrix **H** is tantamount to minimizing the square of each column independently. Therefore, Eq. (2) can be equal to the following equation:

$$\min_{\mathbf{c}_n, n=1,2,\ldots,N} \sum_{n=1}^{N} \|\mathbf{H}\mathbf{c}_n - \mathbf{m}_n\|_2^2, \tag{4}$$

where $\mathbf{c}_n$ and $\mathbf{m}_n$ are the nth column of the matrices **C** and **M**, respectively. The optimal solution of $\mathbf{c}_n$ can be obtained by applying the method of least squares to each column

$$\mathbf{c}_n = \mathbf{H}^+\mathbf{m}_n, \quad n = 1,2,\ldots,N, \tag{5}$$

where $\mathbf{H}^+$ is the pseudoinverse of **H**.[22] This optimal solution in the least-squares sense can be assembled a more compact matrix form

$$[\mathbf{c}_1 \quad \mathbf{c}_2 \quad \cdots \quad \mathbf{c}_N] = \mathbf{H}^+[\mathbf{m}_1 \quad \mathbf{m}_2 \quad \cdots \quad \mathbf{m}_N] \tag{6a}$$

or

$$\mathbf{C} = \mathbf{H}^+\mathbf{M}. \tag{6b}$$

For a matrix **H** with full-column rank ($L \geq M$), $\mathbf{H}^+$ can be calculated according to

$$\mathbf{H}^+ = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H. \tag{7}$$

Here, $\mathbf{H}^+$ is also referred to as the left pseudoinverse of **H** in that $\mathbf{H}^+\mathbf{H} = \mathbf{I}$.

In practice, the number of loudspeakers is usually greater than the number of ears, i.e., $L \leq M$. Regularization can be used to prevent the singularity of $\mathbf{H}^H\mathbf{H}$ from saturating the filter gains.[23,24]

$$\mathbf{H}^+ = (\mathbf{H}^H\mathbf{H} + \beta\mathbf{I})^{-1}\mathbf{H}^H. \tag{8}$$

The regularization parameter β can either be constant or frequency dependent.[25] It is noted that the procedure to obtain the filter **C** in Eq. (6) is essentially a frequency-domain formulation, inverse Fourier transform along with circular shift (hence the modeling delay) are needed to obtain causal FIR filters.

III. NUMERICAL SIMULATIONS

In this section, numerical simulations are conducted to examine the effects that listening angle has on CCS. The free-field point source model and HRTFs are employed as the plant models in the simulations. Only lateral misalignment is considered because it has been concluded by the previous research that the lateral misalignment has more pronounced effect on CCS than the other types of head movements.[15]

A. Free-field point source model

For the free-field point source model illustrated in Fig. 2, the plant transfer matrix can be shown to be

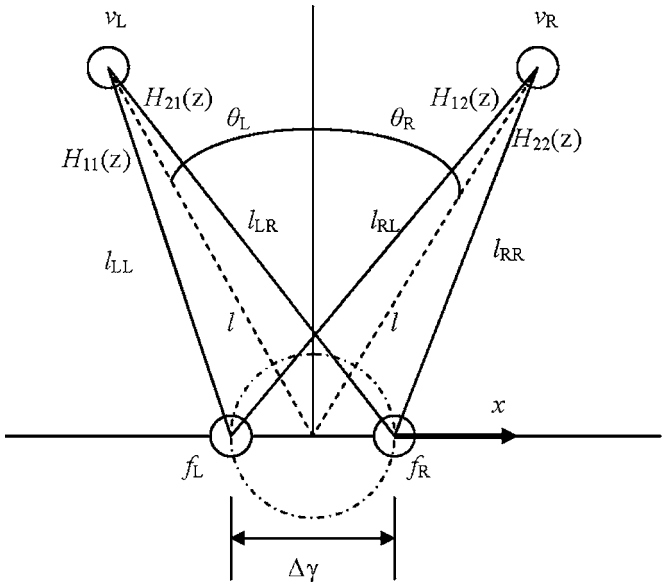


FIG. 2. The geometry of the free-field point source model.

$$\mathbf{H} = \frac{\rho_0}{4\pi}\begin{bmatrix} e^{-jk_a l_{LL}/l_{LL}} & e^{jk_a l_{RL}/l_{RL}} \\ e^{-jk_a l_{LR}/l_{LR}} & e^{-jk_a l_{RR}/l_{RR}} \end{bmatrix} k_a = \omega/c_0, \tag{9}$$

where k_a, ρ_0, and c_0 represent the wave number, the density, and sound speed, respectively. In the simulation, we assume that $c_0 = 343$ m/s, $\rho_0 = 1.21-$ kg/m^3, $l = 1.4$ m, and the spacing between ears. $\Delta\gamma = 0.1449$ m.[26] In Eq. (9), the lengths are calculated as

$$l_{LL} = \left[(l\cos\theta)^2 + \left(l\sin\theta - \frac{\Delta\gamma}{2} + x \right)^2 \right]^{1/2}, \tag{10a}$$

$$l_{LR} = \left[(l\cos\theta)^2 + \left(l\sin\theta + \frac{\Delta\gamma}{2} + x \right)^2 \right]^{1/2}, \tag{10b}$$

$$l_{RL} = \left[(l\cos\theta)^2 + \left(l\sin\theta + \frac{\Delta\gamma}{2} - x \right)^2 \right]^{1/2}, \tag{10c}$$

$$l_{RR} = \left[(l\cos\theta)^2 + \left(l\sin\theta - \frac{\Delta\gamma}{2} - x \right)^2 \right]^{1/2}. \tag{10d}$$

The CCS filters are obtained by using the aforementioned inverse filtering procedure with constant regularization parameters. Overall, 256 frequencies equally spaced from 20 to 20 kHz on a logarithmic frequency scale are selected. The kth selected frequency can be represented as

$$f(k) = 10^{\log_{10}^{20} + (\log_{10}^{20\,000} - \log_{10}^{20})k/256}, \quad k = 0,1,\ldots,255, \tag{11}$$

where $\log_{10}^{20}$ and $\log_{10}^{20\,000}$ symbolize the logarithm with base 10 for 20 Hz and 20 kHz, respectively. In the simulation, the power of each CCS filter at different span angles is constrained to be equal, which can be achieved by using different regularization values. The 2×2 transfer function matrix is assumed to be symmetric. The power of CCS filters is defined as

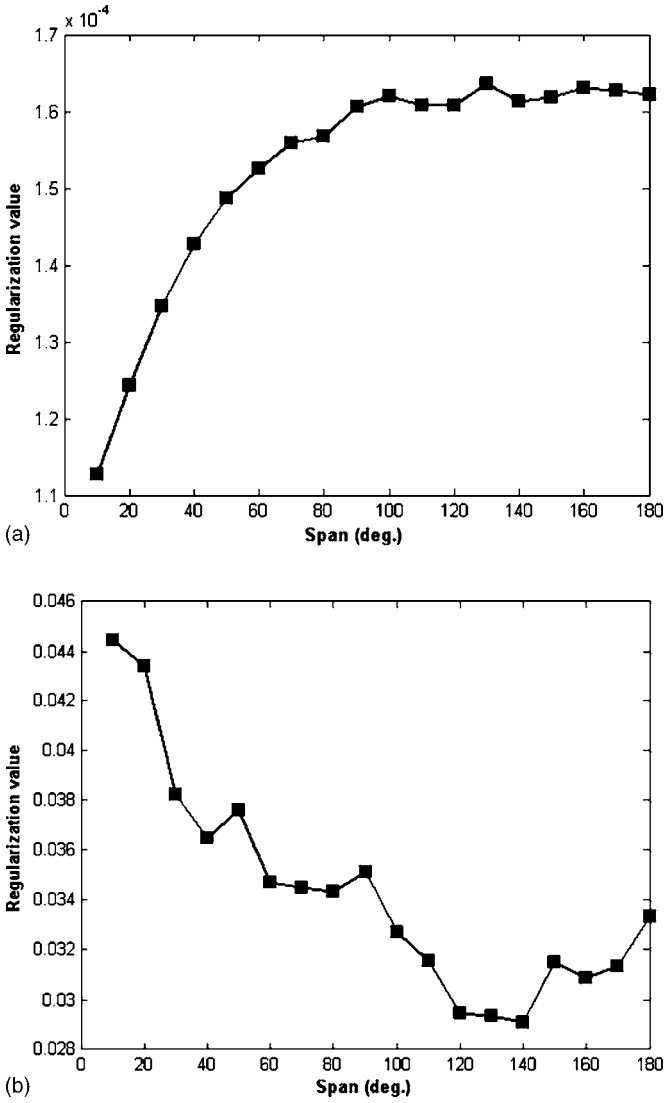


FIG. 3. The values of regularization in (a) the free-field point source model and (b) the HRTF model.

$$\frac{1}{P}\sum_{k=0}^{P-1}[|C_{11}(k)|^2+|C_{12}(k)|^2], \tag{12}$$

where C_{11} and C_{12} are diagonal and off-diagonal component of the CCS filter, P is the number of frequency samples and k represents the frequency index. The regularization values in each span angle are shown in Fig. 3(a).

Let the overall response of the CCS filters cascaded with the acoustic plant be

$$\mathbf{G}=\begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}=\mathbf{HC}. \tag{13a}$$

Channel separation, defined as the ratio of the contralateral response and the ipsilateral response compensated by CCS, is employed as a performance index

$$CHSP_{\mathrm{L}}(k)=G_{12}(k)/G_{11}(k) \quad \text{or } CHSP_{\mathrm{R}}(k) = G_{21}(k)/G_{22}(k). \tag{13b}$$

Figure 4(a) shows the contour plot of the condition number

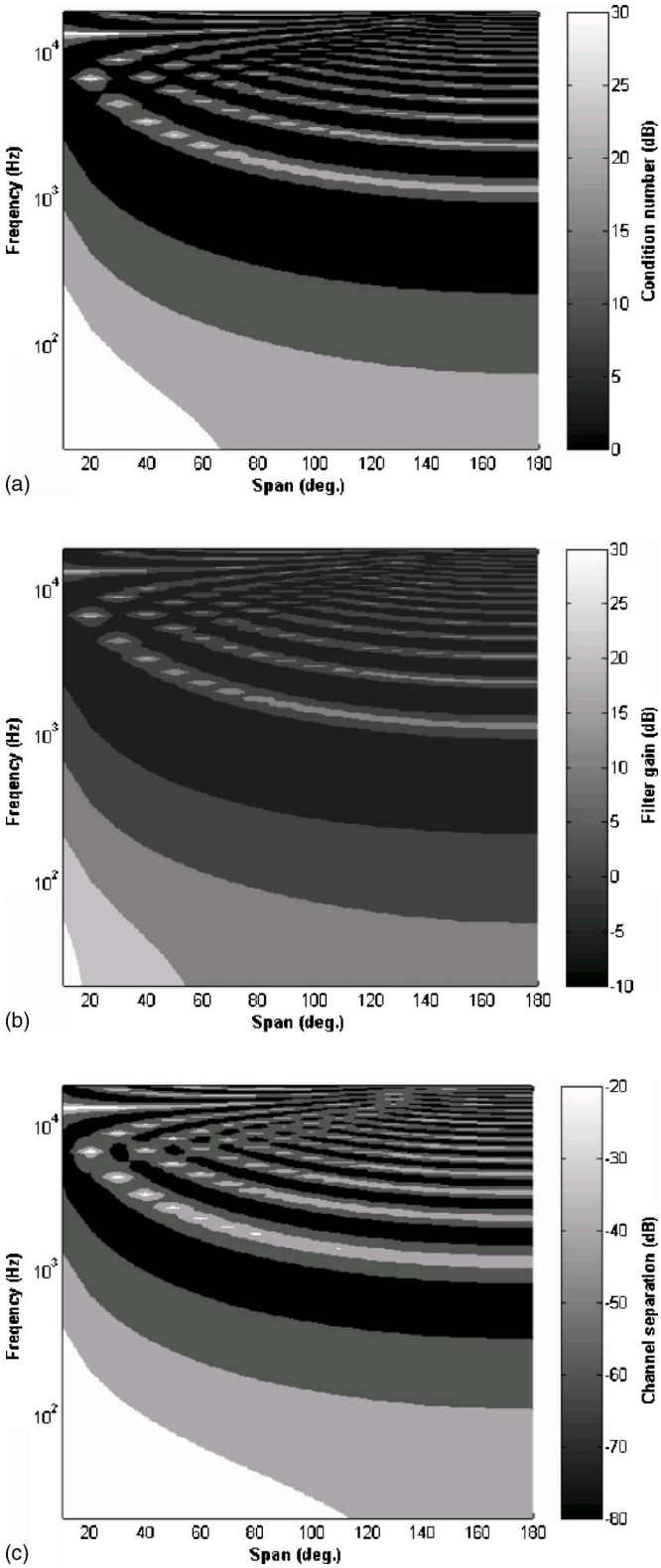


FIG. 4. The contour plots calculated using the point source model of (a) the condition number of acoustical plant matrix **H**, (b) the filter gain, and (c) the channel separation.

of the plant matrix **H** in the nominal center position ($x=0$). The x axis is the listening angles in degrees and the y axis is logarithmic frequency in hertz. Condition number in decibels is represented by gray levels. In addition, the contour plots of the filter gain and the channel separation shown in Figs. 4(b) and 4(c) are plotted versus the same coordinates as in Fig.

4(a). From the plots, the condition number follows a similar trend to the filter gain. This reveals that there is indeed a tradeoff between numerical stability and separation performance. Specifically, a large condition number leads to high filter gain. This in turn calls for regularization to restrain the filter gain at the compromise of some performance.

Another issue of CCS is concerned with the *ringing frequency* given by[15–17]

$$f_n = \frac{nc}{2\Delta r \sin\theta}, \quad n = 0,1,2,\ldots,. \tag{14}$$

Ringing frequencies appear at high frequency particularly for small span arrangement. Suppose the frequency range of our interest is from 100 to 6 kHz. Although the 10-deg span arrangement is well conditioned at frequencies below the intersection of the 6 kHz line and the first ringing, it suffers from the "corner problem," where poor conditioning and high gain arise at low frequencies and small spans. This is to be expected because the acoustic plants are almost identical in magnitude and phase when the listening angle becomes exceedingly small.

Figures 5(a)–5(c) show the contour plots of channel separation at the right ear for three span angles (2θ), 10, 60, and 120 deg, respectively. The span of 10 and 60 deg are selected because they correspond to stereo dipole and International Telecommunications Union (ITU) standard.[27] The x axis is the lateral head displacements in centimeters and the y axis is logarithmic frequency in hertz. Channel separation in decibel is represented by gray levels. The darker the gray level, the better the separation performance. From the contour plot, it can be seen that the pattern becomes progressively complicated as span angle increases. In the nominal center position, the region of good separation performance (the dark stripe) extends toward lower frequency limit (near 100 Hz) for the 120-deg span than the frequency limit (above 1 kHz) for the 10-deg span. On the other hand, the region of ringing frequencies (the white stripes for positive head displacements) occurs at lower frequency (600 Hz) for the 120-deg span versus 6 kHz for the 10-deg span. Thus, stereo dipole indeed has the advantage of having a much higher usable frequency limit before hitting the first ringing frequency which could lead to high gain inverse filters. However, it is argued by the authors that stereo dipole also suffers performance problems at low frequencies. These facts also suggest that large span arrangement should be used at low frequency, while small span arrangement should be used at high frequency, as suggested by many previous researchers.[11–16]

In order to explore further the effect of listening angle on the separation performance of CCS, an index, average channel separation, is defined as follows:

$$\frac{1}{M_2 - M_1 + 1} \sum_{k=M_1}^{M_2} 20 \times \log_{10}(|\mathrm{CHSP}_y(k)|) \text{ (dB)} \tag{15}$$

where M_1 and M_2 are the frequency indices of the lower and upper limits, and the subscript y denotes either L or R. In the simulation, the lower frequency limit was selected to be 100 Hz (M_1=60) below which the sound is known to be ineffective for localization. The average channel separation in relation to the listening angle and the lateral head displacement is shown with a contour plot in Fig. 6. Figures 6(a)–6(c) correspond to the average channel separations for three different frequency upper limits, 1 kHz (M_2=145), 6 kHz (M_2=211), and 20 kHz (M_2=255), re-

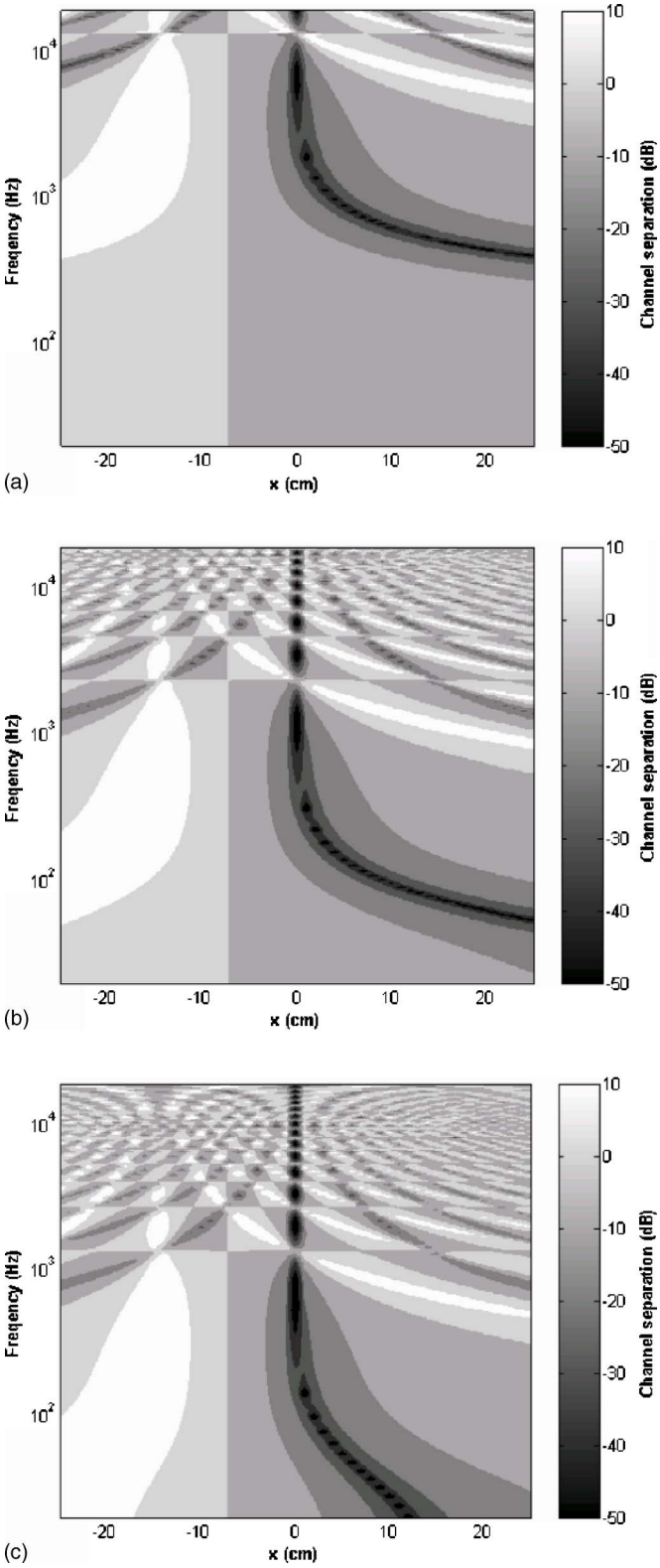


FIG. 5. The contour plots of channel separation at the right ear calculated using the point source model. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span.

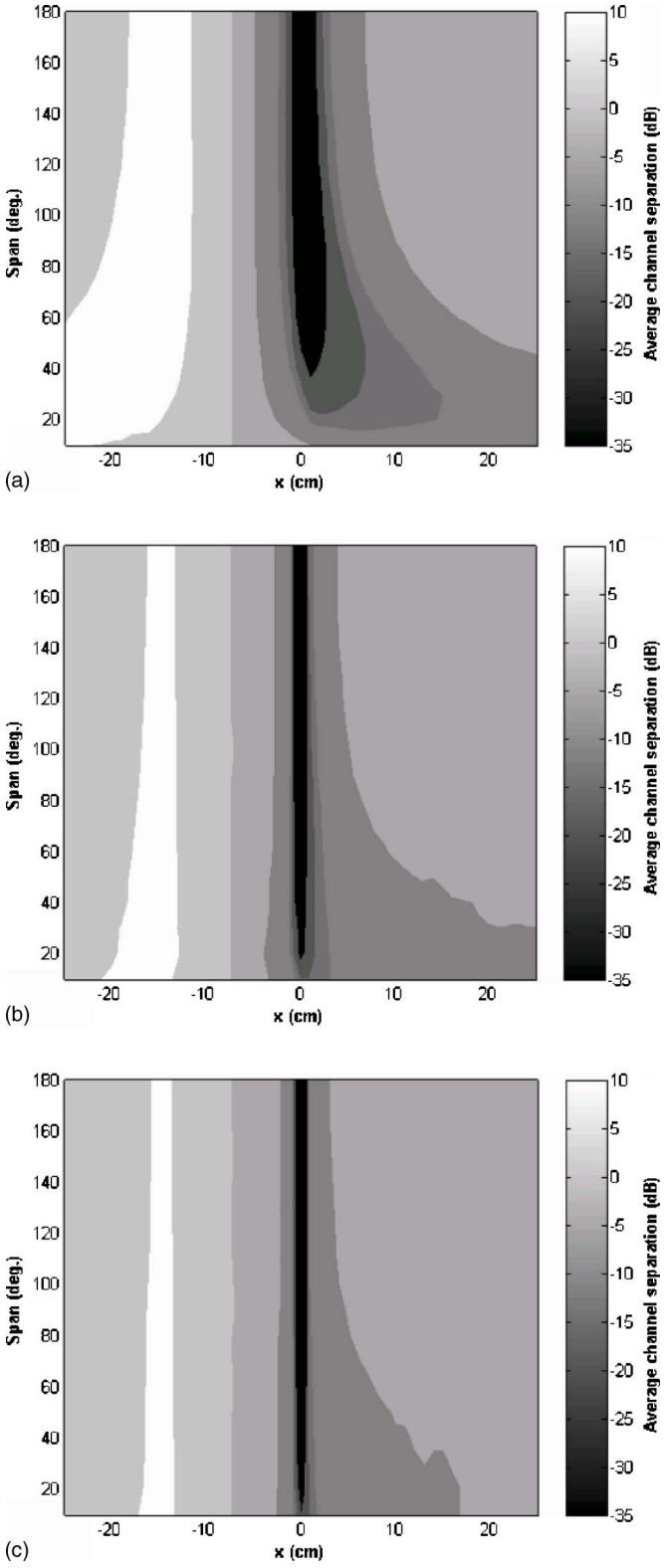


FIG. 6. The contour plots of average channel separation at the right ear calculated using the point source model. (a) Bandwidth to 1 kHz. (b) Bandwidth to 6 kHz. (c) Bandwidth to 20 kHz.

spectively. Using small span angle, a wider region of good separation performance (the second darkest stripe) can be attained at the expense of poor performance, especially for extremely small span. For example, Fig. 6(a) shows the 1-kHz-upper-limit average separation, where the lower tip of the second darkest region barely touches the 20-deg span.

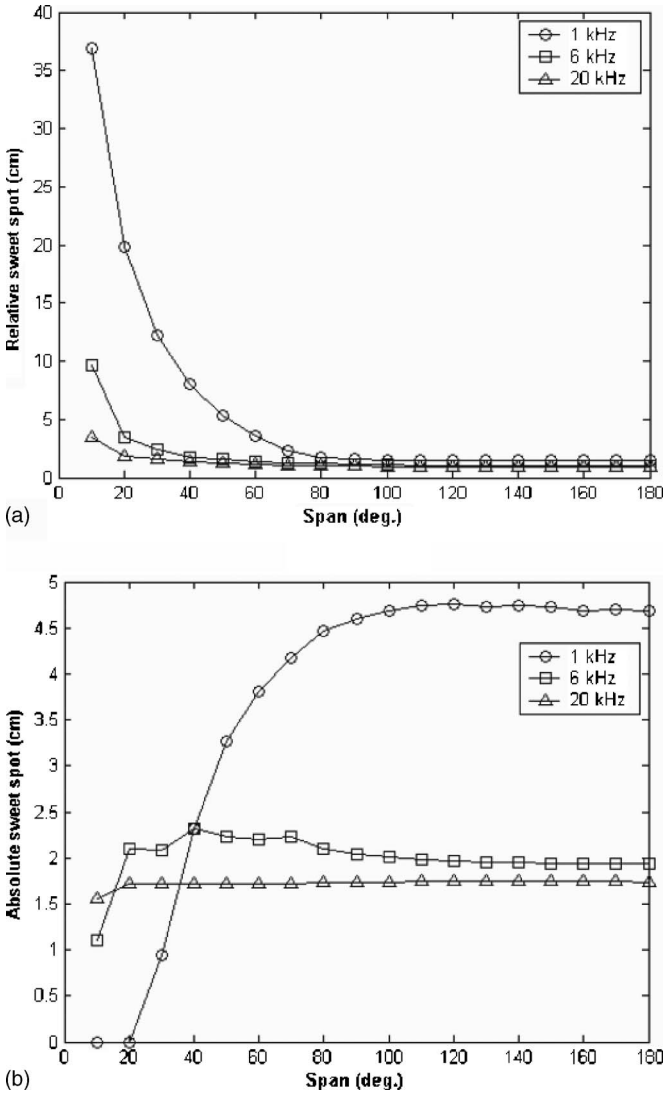


FIG. 7. Two sweet spot definitions calculated using the point source model for 1, 6, and 10 kHz bandwidths. (a) Relative sweet spot. (b) Absolute sweet spot.

The performance of CCS can also be characterized by sweet spot which refers to the region in which the CCS is effective. To be able to better assess the sweet spot quantitatively, two kinds of sweet spot are defined in the paper: the absolute sweet spot and the relative sweet spot. The size of absolute sweet spot is defined as two times the maximum leftward displacement that makes the average channel separation go below −12 dB. The size of relative sweet spot is defined with reference to Fig. 6 as two times the maximum leftward displacement for which the average channel separation is degraded by 12 dB as compared to that of the nominal center position ($x=0$). A value of −12 dB, or 25%, is an empirical value suggested by experience. For the absolute sweet spot, this value is the minimal requirement for CCS. For the relative sweet spot, this value corresponds to the point when the performance drops by 75% from the nominal position. The relative and absolute sweet spots calculated for the point source model are plotted versus span angle in Figs. 7(a) and 7(b), respectively. Three curves plotted in each figure correspond to three different bandwidths, 1, 6, and 20 kHz. As seen in the Fig. 7(a), the relative sweet spot is

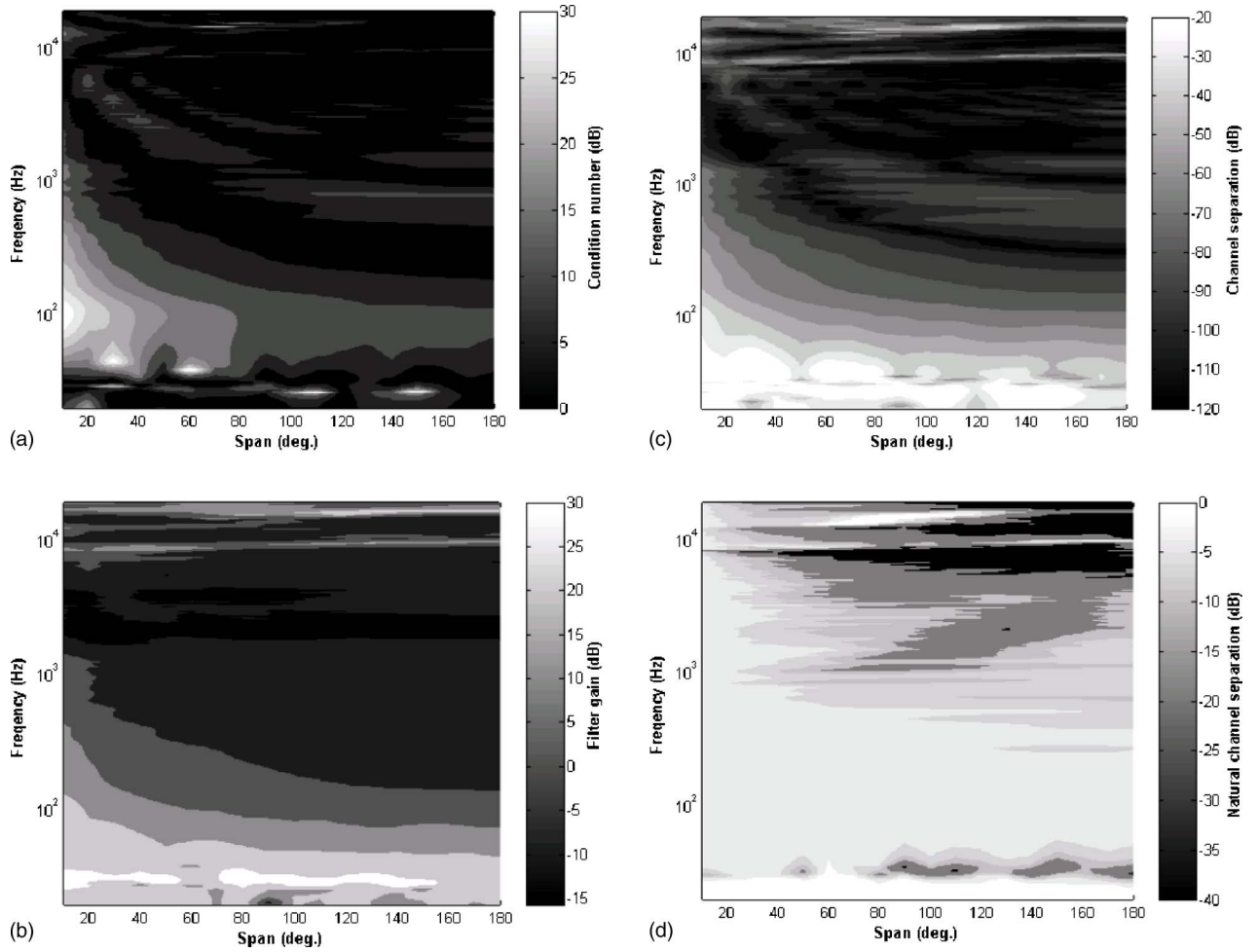


FIG. 8. The contour plots calculated using the HRTF model of (a) the condition number of acoustical plant matrix **H**, (b) the filter gain, (c) the channel separation, and (d) the uncompensated natural channel separation.

increased monotonically as the span is decreased, as predicted by previous researchers. This suggests that small span arrangement is more robust against head misalignment notwithstanding the poor separation performance at the nominal position. However, if the absolute sweet spot is taken as the robustness index, the conclusion is quite different. If this definition of sweet spot is used, the simulation result suggests that the optimal span angle ranges from 80 to 180 deg.

B. HRTF model

In addition to the point source model, a more sophisticated model based on HRTF is employed in the simulation to better account for the diffraction and shadowing effects due to the head, ears, and torso. The HRTF database measured by MIT Media Lab was employed. In the nominal position, the plant transfer function matrix is written as

$$\mathbf{H} = \begin{bmatrix} H_\theta^i & H_\theta^c \\ H_\theta^c & H_\theta^i \end{bmatrix}, \tag{16}$$

where θ is the span angle and the superscript i and c refer to ipsilateral and contralateral side, respectively. As the head moves to the right by x centimeters, the plant matrix is no longer symmetric and should be modified. The azimuth angle should be modified according to

$$\theta_{\mathrm{L}} = \tan^{-1}\frac{l\sin\theta_{\mathrm{L}_0} + x}{l\cos\theta_{\mathrm{L}_0}}, \tag{17a}$$

$$\theta_{\mathrm{R}} = \tan^{-1}\frac{l\sin\theta_{\mathrm{R}_0} - x}{l\cos\theta_{\mathrm{R}_0}}, \tag{17b}$$

where θ_{L_0} and θ_{R_0} are the angles in the nominal position, i.e., $x=0$. Linear interpolation is called for when the angle is not a multiple of a five-degree interval as the database was originally organized.[19] In addition to angles, the magnitudes and phases are also adjusted to account for attenuation and delay due to distance change. Thus,

$$\mathbf{H} = \begin{bmatrix} H_{\theta_{\mathrm{L}}}^i & H_{\theta_{\mathrm{R}}}^c \\ H_{\theta_{\mathrm{L}}}^c & H_{\theta_{\mathrm{R}}}^i \end{bmatrix} \times \begin{bmatrix} \frac{l_{\mathrm{LL}_0}}{l_{\mathrm{LL}}} e^{\frac{-j\omega(l_{\mathrm{LL}}-l_{\mathrm{LL}_0})}{c}} & \frac{l_{\mathrm{RL}_0}}{l_{\mathrm{RL}}} e^{\frac{-j\omega(l_{\mathrm{RL}}-l_{\mathrm{RL}_0})}{c}} \\ \frac{l_{\mathrm{LR}_0}}{l_{\mathrm{LR}}} e^{\frac{-j\omega(l_{\mathrm{LR}}-l_{\mathrm{LR}_0})}{c}} & \frac{l_{\mathrm{RR}_0}}{l_{\mathrm{RR}}} e^{\frac{-j\omega(l_{\mathrm{RR}}-l_{\mathrm{RR}_0})}{c}} \end{bmatrix}, \tag{18}$$

where the subscript "0" refers to the nominal position.

The contour plots of the condition number, filter gain, and channel separation are shown in Figs. 8(a)–8(c). The

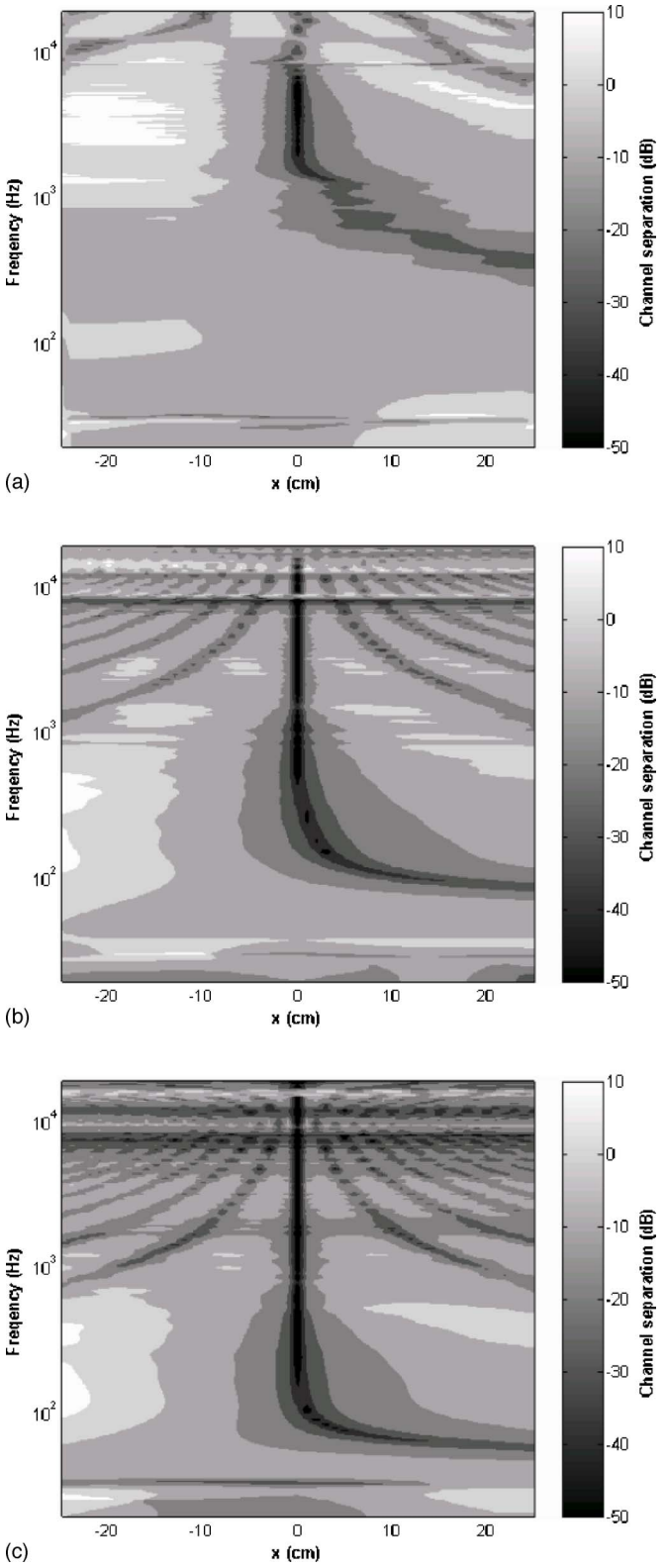


FIG. 9. The contour plots of channel separation measured at the right ear of the acoustic manikin. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span.

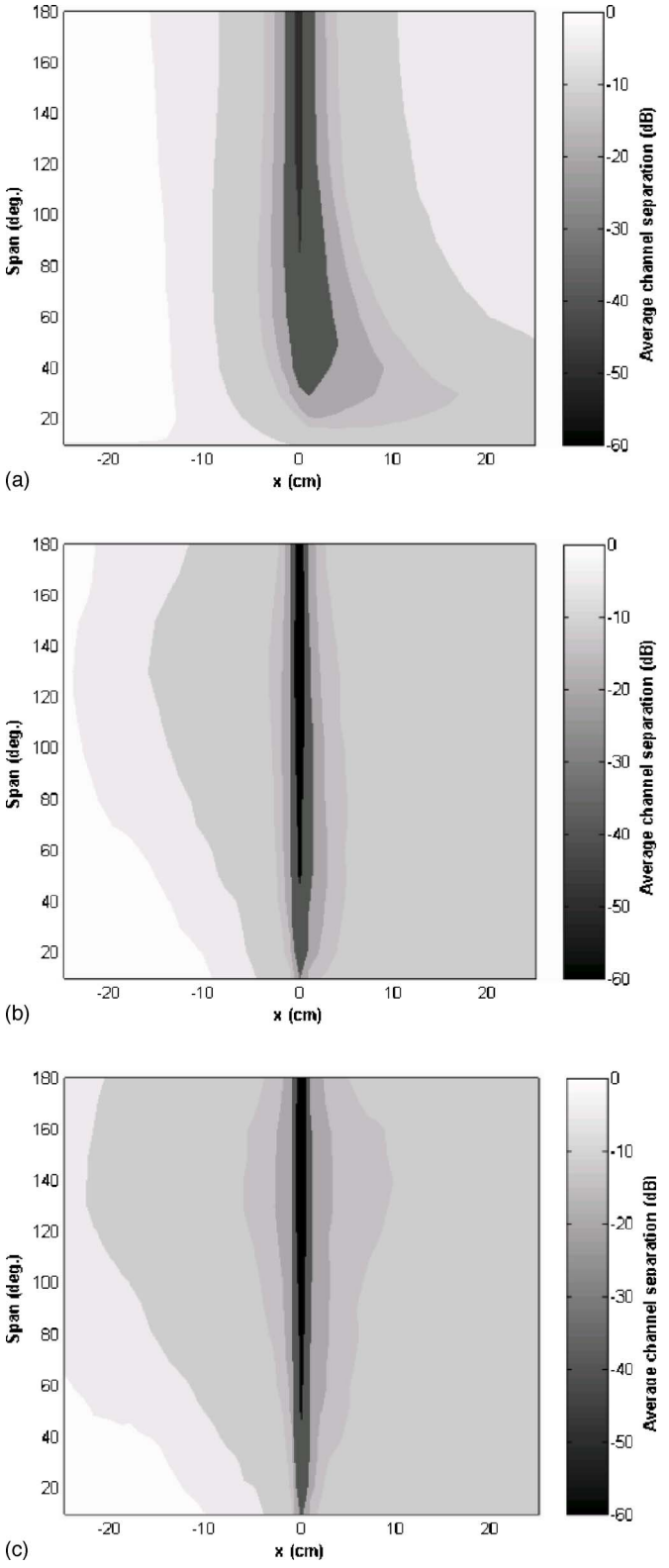


FIG. 10. The contour plots of band-average channel separation measured at the right ear of the acoustic manikin. (a) 1 kHz bandwidth. (b) 6 kHz bandwidth. (c) 20 kHz bandwidth.

uncompensated natural channel separation is also shown in Fig. 8(d) for reference, where the effect of head shadowing is clearly visible. By and large, the results of point source and HRTF follow a similar trend except one important distinction. Because of head shadowing at high frequencies, ringing does not show up in the HRTF results as pronouncedly as in the point source model except a constant ringing around 8–10 kHz due to the concha dip which is almost independent of span. The operation zone of HRTF is thus bounded from above by the concha dip, in contrast to the point source case that is bounded from above by the first ringing. This suggests that a large span arrangement seems to provide bet-

ter numerical stability with a larger useful frequency range than the small span arrangement. The separation performance at high frequencies for large spans is also better (reflected by more dark areas) than that of the small span owing to natural separation provided by head shadowing.

The contour plots of channel separation versus displacement and frequency are shown in Figs. 9(a)–9(c), corresponding to span angles 10, 60, and 120 deg, respectively. The trends of this result are largely the same as that of the point source model. The separation performance at low frequencies is still not good for the 10-deg span [Fig. 9(a)]. Figures 10(a)–10(c) show the contour plots of average channel separation versus displacement and span angle for frequency bandwidth, 1, 6, and 20 kHz, respectively. The trend of the HRTF result is similar to that of the point source result if only a narrow bandwidth, e.g., 1 kHz, is considered [Fig. 6(a) versus Fig. 10(a)]. However, if average separation performance is calculated for a larger bandwidth, e.g., 20 kHz, the results turn out to be quite different. The average performance is poor for extremely small spans. The region of good performance (the darkest strip) is mainly located around the median span area, say, from 100 to 160 deg. This difference of conclusion with the previous point source model is again due to the fact that the head shadowing effect will come into play at high frequencies.

The relative and absolute sweet spots, as defined previously in the point source simulation, are calculated for the HRTF model in three different bandwidths, 1, 6, and 20 kHz, as shown in Figs. 11(a) and 11(b). Similar to the point source results, the relative sweet spot is increased monotonically as the span is decreased, which suggests that small span arrangement is relatively robust against head misalignment notwithstanding the poor separation performance at the nominal position. On the other hand, the results of the absolute sweet spot suggest that arrangements with listening angles ranging from 120 to 150 deg [the intersection of bandwidth of 6 and 20 kHz in Fig 11(b)] seem to be good choices.

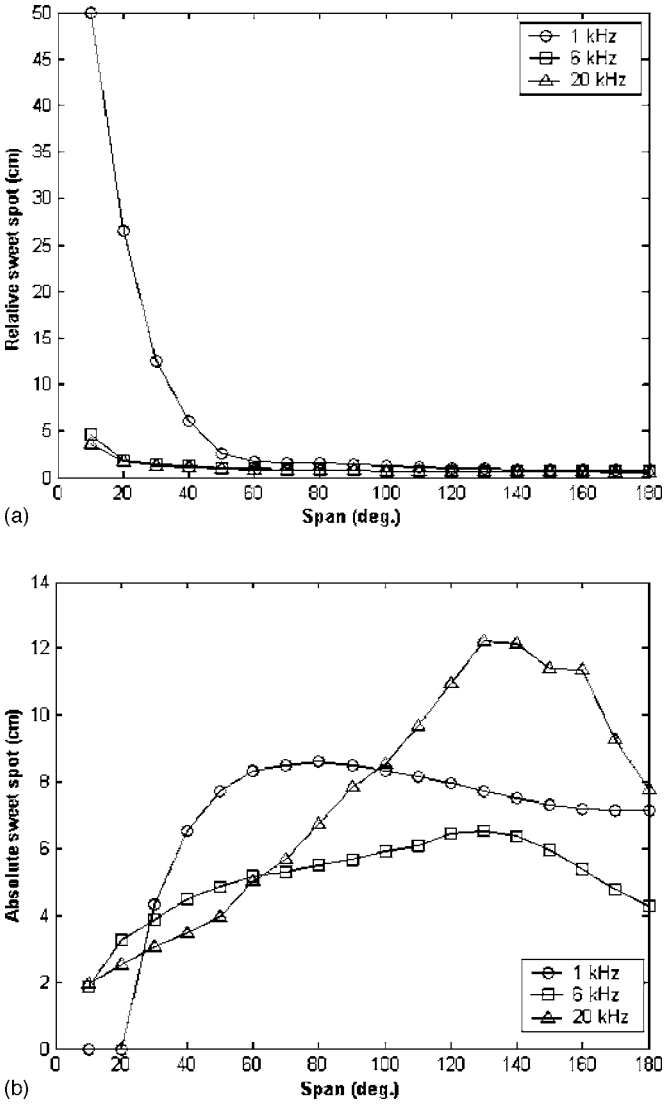


FIG. 11. Two sweet spot definitions calculated using the HRTF model for 1, 6, and 10 kHz bandwidths. (a) Relative sweet spot. (b) Absolute sweet spot.

IV. OBJECTIVE AND SUBJECTIVE EXPERIMENTS

The forgoing simulation results suggest that the optimal listening angle ranges from 120 to 150 deg. This observation is further examined in a series of objective and subjective experiments. Three loudspeaker arrangements with 10-, 60-, and 120-deg spans were compared in the experiments. The 10-deg span represents stereo dipole. The 60-deg span is suggested in the ITU standard of a multichannel stereophonic system.[27] The 120-deg span represents the optimal span previously found in the simulation. All experiments were carried out in an anechoic room, as shown in Fig. 12.

A. Objective experiment

This experiment employed a 5.1-channel loudspeaker system, Inspire 5.1 5300 of Creative, and a digital signal processor (DSP), Blackfin-533, of Analog Device. The microphones and the preamplifier used are GRAS 40AC and GRAS 26AM. The plant transfer function matrixes were measured on an acoustical manikin, KEMAR (Knowles Electronics Manikin for Acoustic Research) along with the ear model, DB-065.

The designed CCS filters were implemented on the DSP using 512-tapped Finite Impulse Response (FlR) filters. The performance of CCS was evaluated in terms of channel sepa-

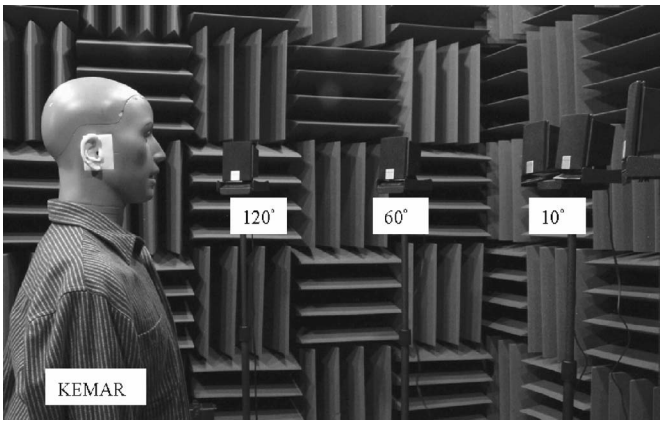


FIG. 12. Photo of the experimental arrangement.

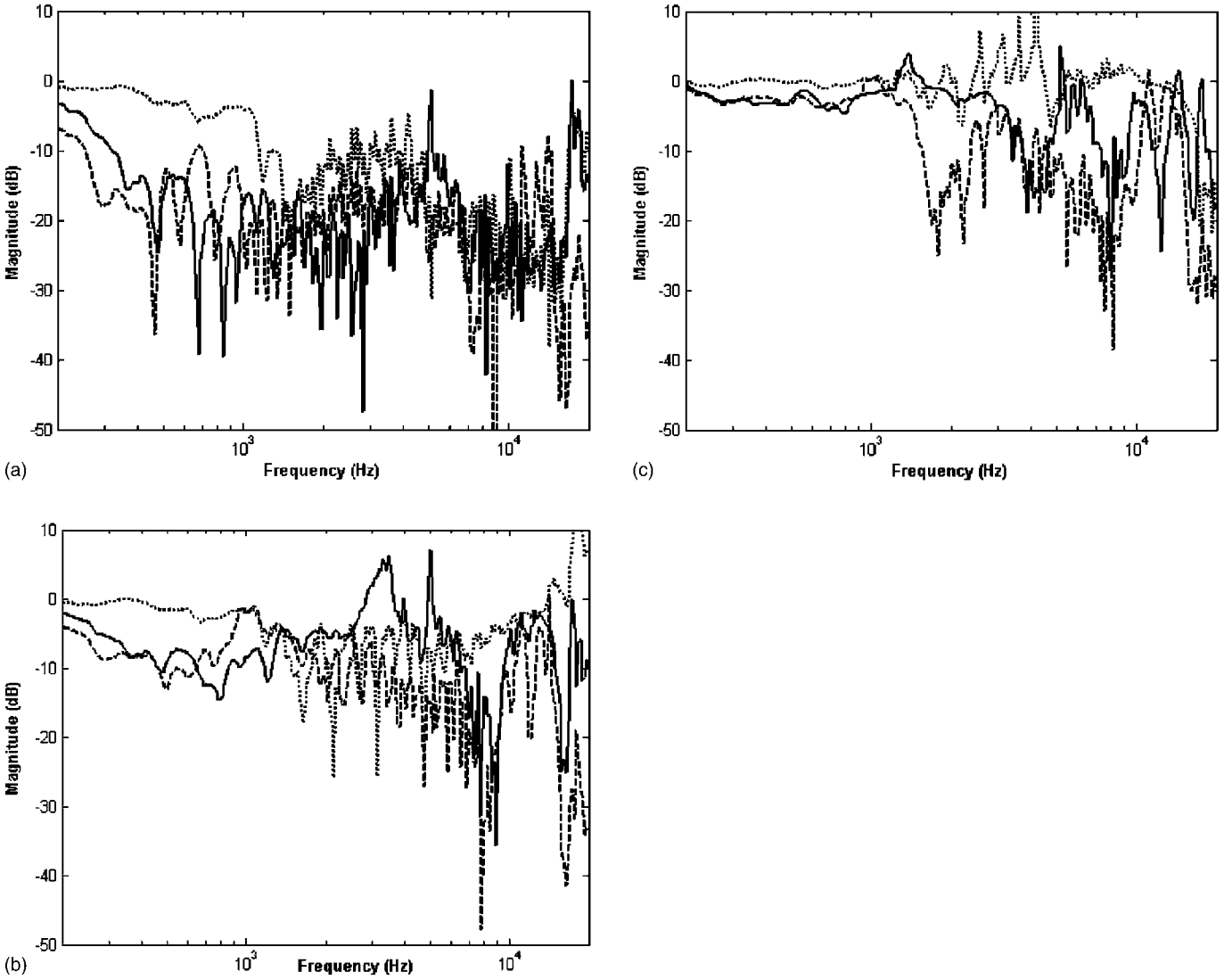


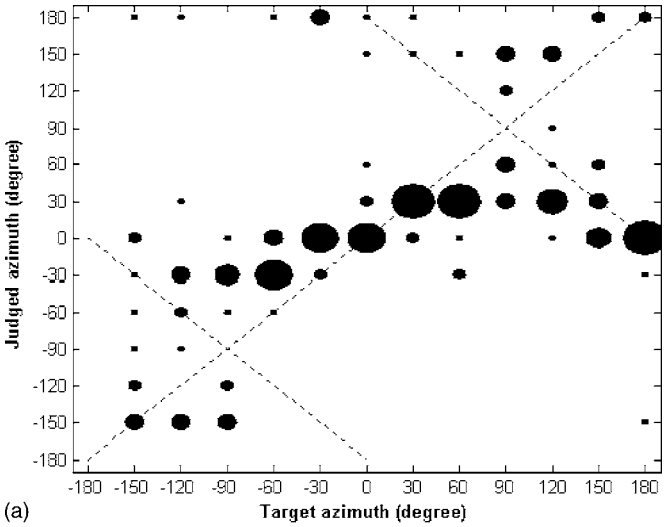
FIG. 13. Channel separations measured at the right ear of the acoustic manikin. The dotted lines, solid lines, and dashed lines represent 10-, 60-, and 120-deg spans, respectively. (a) In the nominal position (x=0 cm). (b) Rightward 5 cm displacement. (c) Rightward 10 cm displacement.

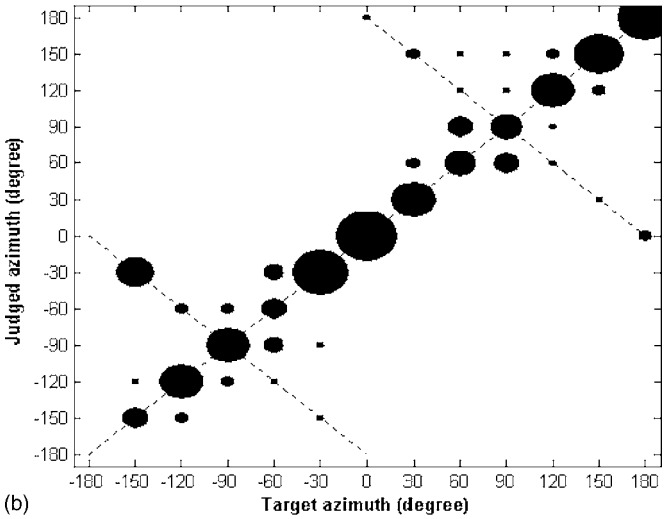
ration. Figure 13(a) shows the right-ear channel separation at the nominal position with three span angles. The x axis and y axis represent frequency in hertz and magnitude in decibels, respectively. The dotted line, the solid line, and the dashed line signify 10°, 60°, and 120° span angles, respectively. The results of Figs. 13(b) and 13(c) were obtained for the cases when the manikin was moved to the right by 5 and 10 cm. Notable of these results is that the 10-deg span performed badly at the frequencies below 1 kHz. The separation performance significantly degraded by as much as 15 dB as the head moved to the right by 5 cm irrespective of which span was used. As the head was displaced by 10 cm, CCS failed almost completely, except at high frequencies, when the large 120-deg span arrangement still maintained natural separation because of head shadowing.

B. Subjective experiment

For the purpose of comparing the CCS with different span angles, a subjective listening experiment of source localization was undertaken in the anechoic room. Eleven subjects participated in the test. The listeners were instructed to sit at three positions: the nominal position, 5-cm displacement to the right, and 10-cm displacement to the right. In order to ensure that each listener sat at the same designated position, the test subjects were asked to rest their chins on a steel frame. The height of the listener's ear was 120 cm which is the same height as the loudspeaker. A pink noise was used as the test stimulus whose bandwidth ranges from 20 Hz to 20 kHz and the reproduction level was 95 dB. Each stimulus was played five times in 25-ms duration with 50-ms silent interval. Virtual sound images at 12 prespecified directions on the horizontal plane with increment 30° azimuth are rendered by using HRTFs. Listeners were well trained by playing the stimuli of all angles prior to the test. The listeners were asked to report the perceived direction of source in the range (−180,180] with a 30-deg interval. Experiments were divided into two groups: 10 deg versus 120 deg and 60 deg versus 120 deg. The experiments were blind tests in that stimuli were played randomly without informing the subjects the source direction. One session of test lasts 15–20 min.

The results of the localization test are shown in terms of target angles versus judged angles in Figs. 14–16, corre-





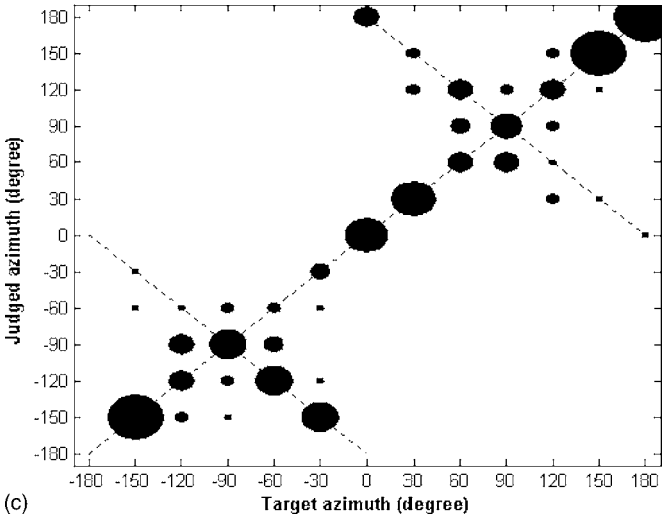
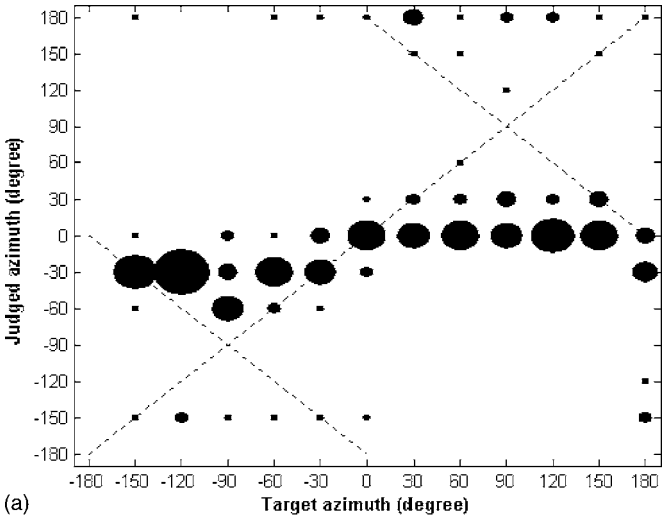
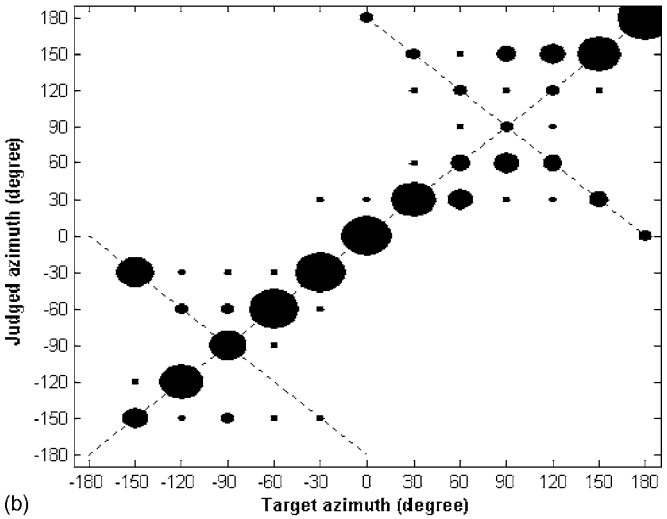


FIG. 14. Results of the subjective localization test of azimuth angles with no head displacement. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span.





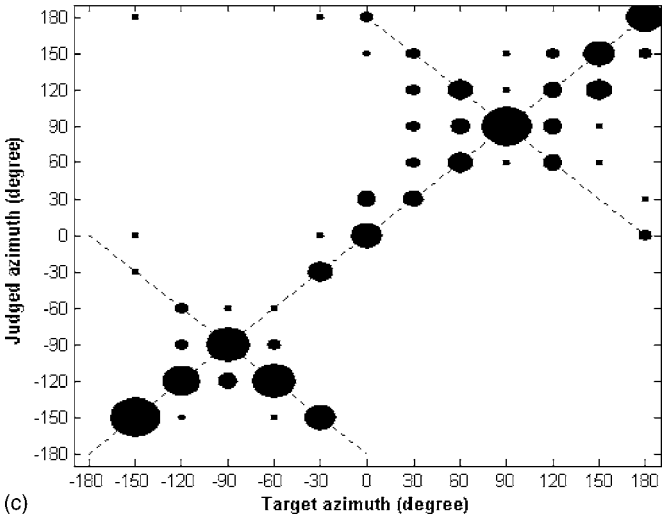


FIG. 15. Results of the subjective localization test of azimuth angles with 5-cm head displacement to the right. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span.

sponding to the cases of nominal position, 5-cm displacement to the right, and 10-cm displacement to the right. In each figure, subplot (a) to (c) refer to the 10°, 60°, and 120° spans, respectively. The size of each circle is proportional to the number of the listeners who localized the same perceived angle. The 45-deg line indicates the perfect localization. It is observed from the results that the subjects tend to localize the sources within ±30 deg about the center line using the 10-deg span arrangement, especially when there is head dis-

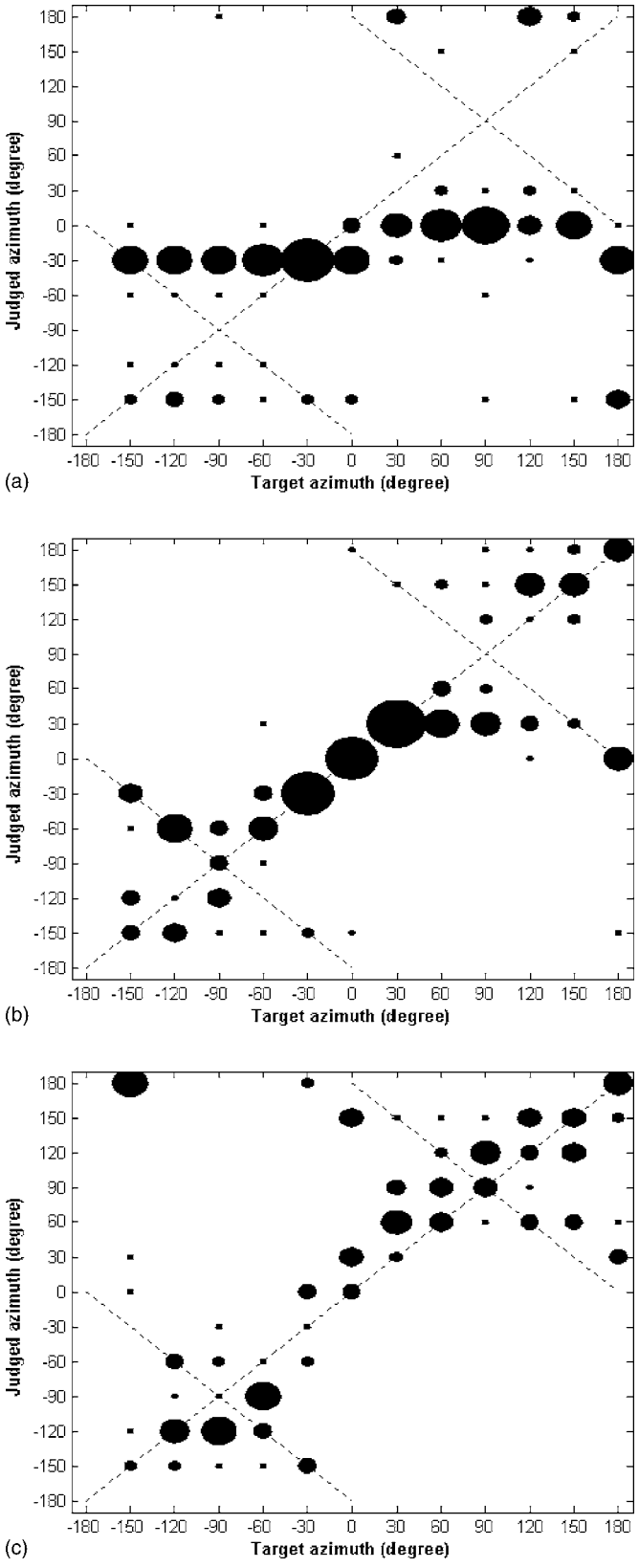


FIG. 16. Results of the subjective localization test of azimuth angles with 10-cm head displacement to the right. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span.

placement. On the other hand, the 60-deg span and the 120-deg span were found to be effective in localizing good frontal images and rear images albeit some front-back reversals. Localization error increases with head displacement irrespective of which span arrangement was used. The 10-deg span seemed to have difficulty localizing sources outside the subtending angle because the separation performance in low frequencies is too poor in small span arrangement to maintain proper spatial cues such as interaural time difference (ITD) which works only under 1 kHz. In contrast, the arrangement with large span appears to be more robust than the small span because head shadowing and panning effect help to provide localization effect to certain degree even if CCS breaks down.

TABLE I. The description of five levels of grade for the subjective localization test.

Grade	Description
5	The judged angle is the same as the target angle
4	30° difference between the judged angle and the target angle
3	Front-back reversal of the judged angle identical to the target angle
2	30° difference between front-back reversal of the judged angle and the target angle
1	Otherwise

To justify the finding, a Friedman test on the subjective localization results in relation to span was conducted. These results were preprocessed into five levels of grade, as described in Table I. The results of the Friedman test are summarized in Table II for the first and second groups. Figure 17(a) shows the medians (small square), quartiles (box), and ranges (whiskers) of the 10-deg span and the 120-deg span. Friedman test output of the first group in Table II reveals that the span effect is statistically significant ($p<0.001$). This indicted that the 120-deg span outperformed the 10-deg span. Figure 17(b) shows the medians, quartiles, and ranges of the 60-deg span and the 120-deg span. The Friedman test of the second group in Table II reveals that the difference of performance of two listening angles is found statistically insignificant ($p<0.3458$). This does not seem to agree with the prediction of the previous simulation that the 120-deg span should perform slightly better than the 60-deg span. It is suspected that the enormous span of 120-deg arrangement is actually quite detrimental to localizing sources at the center position, especially when CCS beaks down. Experience shows overly large angle arrangements seem to have difficulties in positioning images at the center region. In fact, some of the test subjects reported that it sounded like there was an opening of sound field in the front. This offsets somewhat the expected performance gain of CCS using large span arrangement.

TABLE II. The Friedman test result of the subjective experiments.

	First group (10 vs 120)	Second group (60 vs 120)
Chi-Squares ($N=396$, $df=1$)	47.4568	0.8889
Significant p value	<0.001	<0.3458

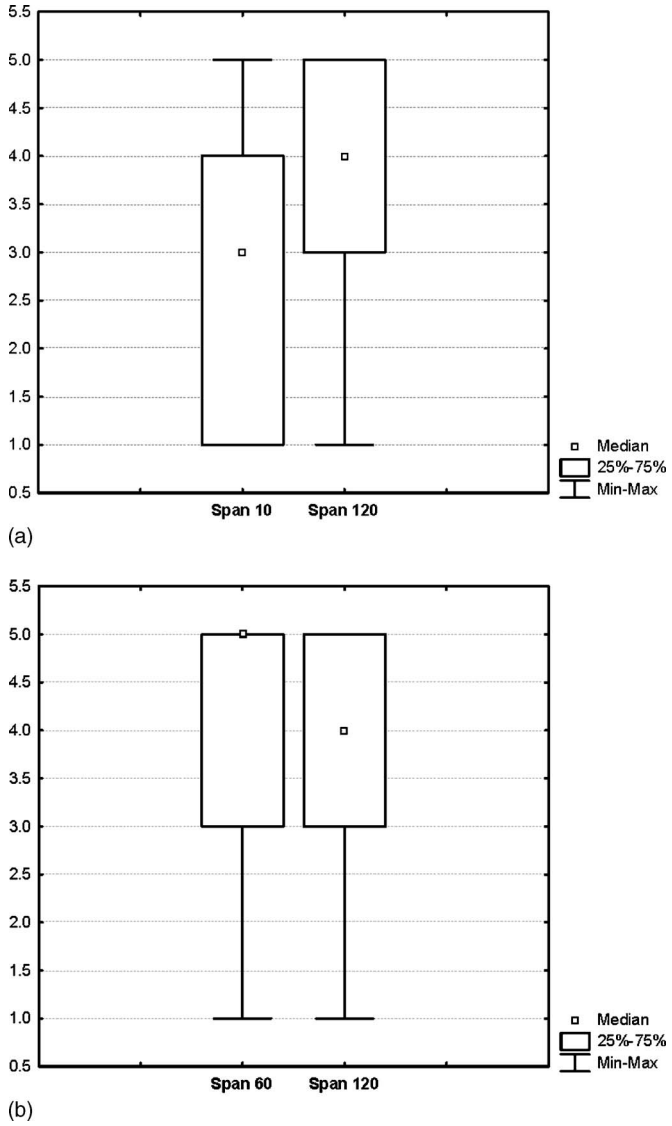


FIG. 17. The box and whisker plots, (a) 10-deg arrangement vs 120-deg arrangement. (b) 60-deg arrangement vs 120-deg arrangement.

V. CONCLUSIONS

A comprehensive study has been conducted to explore the effects of listening angle on crosstalk cancellation in spatial sound reproduction using two-channel stereo systems. The intention is to establish a sustainable configuration of CCS that best reconciles the separation performance and the robustness against lateral head movement, not only in theory but also in practice. Similar to the previous research which focuses mainly on numerical stability, the present work arrives at the conclusion that inversion of ill-conditioned systems results in high gain filters, loss of dynamic range, and hence separation performance. Regularization is required to compromise between numerical stability and separation performance. However, findings different from the previous study had also been reached because this work employed a comprehensive approach. First, it is found from the HRTF results that the problem of high frequency ringing is not as critical as in the point source model owing to head shadowing. In addition, poor conditioning, high gain, and low performance problems at low frequencies may arise for extremely small span arrangements, whereas there is broader useful frequency range with performance and numerical stability if wide span arrangement can be used. The effects of listening angle were also examined in the context of the sweet spot. Two kinds of sweet spot definitions are employed in the simulation. The relative sweet spot suggests that robustness is excellent with the use of small span arrangement notwithstanding the poor performance in the nominal position, which is in agreement with the previous research. However, it is not very useful in practical application if the average channel separation in the sweet spot is very poor even though it is relatively robust. Therefore, in addition to the conventional relative definition, we suggest another definition, the *absolute* sweet spot, to make the evaluation more complete. In an absolute sweet spot, the performance is guaranteed in complement to the relative robustness, which is desirable in practical use of the CCS. The results of absolute sweet spot reveal that arrangements with a listening angle ranging from 120 to 150 deg are optimal choices.

To justify the conjectures above, objective and subjective experiments were undertaken in an anechoic room for three loudspeaker arrangements, including the stereo dipole (10 deg), standard span (60 deg), and proposed span (120 deg). The results postprocessed by the Friedman test indicate that the 120-deg configuration performs comparably well as the standard 60-deg configuration, but is better than the 10-deg configuration. Small span arrangement produces a large relative sweet spot because head displacement would cause minimal change of time-of-arrival differences between two loudspeakers using closely spaced loudspeakers. This configuration is well suited to applications that must be spatially compact, e.g., mobile phones and other portable devices. Nevertheless, the benefit of small span arrangement comes at the price of poor conditioning, high gain, and limited performance problems at low frequencies. Apart from this, due to the lack of natural high frequency separation provided by head shadowing, the small span arrangement is not able to position "out-of-range" source when CCS breaks down at high frequencies, where the phantom source is incorrectly panned within a narrow span. The arrangement with large span appears to be more effective than the small span because head shadowing and panning effect help to provide a localization effect to a certain degree even if CCS breaks down. While it may seem from this report that large-span configuration is predominantly favored, problems inherent to large span prevent the span to grow indefinitely, e.g., sound image stability will become an issue for wide apart loudspeakers. A practical recommendation is perhaps the conventional 60-deg configuration which is a reasonable compromise between the two extremes (10 and 120 deg) to achieve both robustness and performance. It was also found that the 120-deg arrangement did not perform as well as the 60-deg arrangement in positioning frontal images. If an additional center loudspeaker is available, the 3/0 format with 120-deg span would be an ideal choice.

ACKNOWLEDGMENTS

The work was supported by the National Science Council in Taiwan, Republic of China, under the Project No. NSC94-2212-E009-019.

[1]J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1997).
[2]D. R. Begault, *3-D Sound for Virtual Reality and Multimedia* (AP Professional, Cambridge, MA, 1994).
[3]R. Schroeder and B. S. Atal, "Computer simulation of sound transmission in rooms," IEEE Int. Convention Record **7**, 150–155 (1963).
[4]P. Damaske and V. Mellert, "A procedure for generating directionally accurate sound images in the upper-half space using two loudspeakers," Acoustics **22**, 154–162 (1969).
[5]D. H. Cooper, "Calculator program for head-related transfer functions," J. Audio Eng. Soc. **30**, 34–38 (1982).
[6]W. G. Gardner, "Transaural 3D audio," MIT Media Laboratory Tech. Report 342 (1995).
[7]D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," J. Audio Eng. Soc. **37**, 3–19 (1989).
[8]J. L. Bauck and D. H. Cooper, "Generalized transaural stereo and applications," J. Audio Eng. Soc. **44**, 683–705 (1996).
[9]C. Kyriakakis, T. Holman, J. S. Lim, H. Homg, and H. Neven, "Signal processing, acoustics, and psychoacoustics for high-quality desktop audio," J. Visual Commun. Image Represent **9**, 51–61 (1997).
[10]C. Kyriakakis, "Fundamental and technological limitations of immersive audio systems," IEEE Signal Process. Mag. **86**, 941–951 (1998).
[11]D. B. Ward and G. W. Elko, "Optimal loudspeaker spacing for robust crosstalk cancellation," Proc. ICASSP 98 (IEEE, Seattle, WA, 1998), pp. 3541–3544.
[12]D. B. Ward and G. W. Elko, "Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation," IEEE Signal Process. Lett. **6**(5), 106–108 (1999).
[13]D. B. Ward, "Joint squares optimization for robust acoustic crosstalk cancellation," IEEE Trans. Speech Audio Process. **8**(2), 211–215 (2000).
[14]O. Kirkeby, P. A. Nelson, and H. Hamada, "The "stereo dipole" a virtual source imaging system using two closely spaced loudspeakers," J. Audio Eng. Soc. **46**, 387–395 (1998).
[15]T. Takeuchi and P. A. Nelson, "Robustness to head misalignment of virtual sound imaging systems," J. Audio Eng. Soc. **109**, 958–971 (2001).
[16]T. Takeuchi and P. A. Nelson, "Optimal source distribution for binaural synthesis over loudspeakers," J. Acoust. Soc. Am. **112**, 2786–2797 (2002).
[17]P. A. Nelson and J. F. W. Rose, "Errors in two-point sound reproduction," J. Acoust. Soc. Am. **118**(1), 193–204, 2005.
[18]M. R. Bai, C. W. Tung, and C. C. Lee, "Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the genetic algorithm," J. Acoust. Soc. Am. **117**, 2802–2813 (2005).
[19]W. G. Gardner, *3-D Audio Using Loudspeakers* (Kluwer Academic, Dordrecht, 1998).
[20]W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908 (1995).
[21]W. G. Gardner and K. D. Martin, *KEMAR HRTF Measurements* (MIT's Media Lab, http://sound.media.mit.edu/KEMAR.html, 1994).
[22]B. Noble, *Applied Linear Algebra* (Prentice-Hall, Englewoods, NJ, 1988).
[23]O. Kirkeby, P. A. Nelson, and H. Hamada, "Fast deconvolution of multichannel systems using regularization," IEEE Trans. Speech Audio Process. **6**, 189–194 (1998).
[24]A. Schuhmacher and J. Hald, "Sound source reconstruction using inverse boundary element calculations," J. Acoust. Soc. Am. **113**, 114–127 (2003).
[25]M. R. Bai and C. C. Lee, "Development and implementation of cross-talk cancellation system in spatial audio reproduction based on the subband filtering," J. Sound Vib. **290**, 1269–1289 (2006).
[26]V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Platz, New York, 2001).
[27]ITU-R Rec. BS.775–1, *Multi-Channel Stereonhonic Sound System With or Without Accompanying Picture* (International Telecommunications Union, Geneva, Switzerland, 1992–1994).

On the modeling of sound radiation from poroelastic materials

Noureddine Atalla[a)]
Department of Mechanical Engineering, Université de Sherbrooke, 2500 Boulevard Université, Sherbrooke, QC, J1K 2R1, Canada

Franck Sgard
Laboratoire des Sciences de l'Habitat, DGCB URA CNRS 1652, Ecole Nationale des Travaux Publics de l'Etat, 69518 Vaulx-en-Velin Cedex, France

Celse Kafui Amedin
Department of Mechanical Engineering, Université de Sherbrooke, 2500 Boulevard Université Sherbrooke, QC, J1K 2R1, Canada

(Received 14 March 2006; revised 29 May 2006; accepted 9 July 2006)

Numerical approaches based on finite element discretizations of Biot's poroelasticity equations provide efficient tools to solve problems where the porous material is coupled to elastic structures and finite extent acoustic cavities. Sometimes, it may be relevant to evaluate the radiation of a poroelastic material into an infinite fluid medium. Examples include (i) the evaluation of the diffuse field sound absorption coefficient of a porous material and/or the sound transmission loss of an elastic plate coupled to a porous sheet, (ii) the assessment of the acoustic radiation damping of a porous material coupled to a vibrating structure. The latter is particularly important for the correct experimental characterization of the intrinsic damping of the material's frame. Up to now, the acoustic radiation of a porous medium into an unbounded fluid medium has usually been neglected. The classical approaches for modeling free field radiation of porous materials (i) assumes the interstitial pressure at the radiation surface to be zero or (ii) fixes the radiation impedance to an approximate value. This paper discusses the limitations of these assumptions and presents a numerical formulation for evaluating the sound radiation of baffled poroelastic media including fluid loading effects. The problem is solved using a mixed FEM-BEM approach where the fluid loading is accounted for using an admittance matrix. Both numerical examples and a transmission loss test are presented to illustrate the performance of the technique and its applications. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2261244]

PACS number(s): 43.40.At, 43.40.Ey, 43.40.Rj [SFW] Pages: 1990–1995

I. INTRODUCTION

Numerical approaches based on finite element discretizations of Biot's poroelasticity equations provide efficient tools to solve problems where the porous material is coupled to elastic structures and finite extent acoustic cavities.[1–9] Often, it is relevant to evaluate the radiation of a poroelastic material into an infinite fluid medium. Examples include (i) the evaluation of the diffuse field sound absorption coefficient of a porous material, (ii) the calculation of the sound transmission loss of an elastic plate coupled to a porous sheet, (iii) the assessment of the acoustic radiation damping of a porous material coupled to a vibrating structure. The latter is particularly important for the correct experimental characterization of the intrinsic damping of the material's frame. Up to now, the acoustic radiation of a porous medium into an unbounded fluid medium has usually been neglected. The classical approach for modelling free field radiation of porous materials assumes the total stress tensor and the interstitial pressure at the radiation surface to be zero.[5] In this case, it is assumed that the coupling between the porous media and the surrounding acoustic medium is negligible; that is the porous system is assumed to be vibrating *in vacuo*. Other methods use a plane wave approximation in the same manner as in the Transfer Matrix Method.[10] This is mainly acceptable at high frequencies but is clearly erroneous at low frequencies since it will overestimate radiation damping. These approximations can be eliminated for specific problems. For instance in the case of radiation of a porous medium in a waveguide, the radiation impedance can be calculated accurately by expressing the radiated pressure in terms of the modal behavior of the waveguide.[11] In the case of a thin porous plate under flexural vibration, Horoshenkov and Sakagami[12] considered the absorption problem using a Helmholtz approach and presented a parameters study on the influence of the porous plate parameters on its absorption. Takahashi and Tanaka[13] considered the same problem and presented an analytical model to calculate the radiation impedance of a thin porous plate in flexure; in particular, they discussed the effects of plate permeability on its radiation damping based on numerical examples.

This paper tackles a special case of the radiation problem. It presents a numerical formulation for evaluating the sound radiation of baffled poroelastic media. The formulation is based on the mixed displacement-pressure formulation of Biot Allard's poroelasticity equations. The presenta-

[a)]Electronic mail: Noureddine.atalla@USherbrooke.ca

tion concentrates in expressing the free field condition using Rayleigh's integral, in terms of an admittance matrix and a solid phase-interstitial pressure coupling term. The approach is general and easy to implement. It can handle the special cases discussed in previous papers[12,13] and more complicated situations such as multilayer systems with various excitations. Numerical results are presented to illustrate the technique in three particular situations: (i) absorption of a porous sample, (ii) transmission loss of a porous sample, and (iii) the effect of radiation on the vibration of a plate-foam system. These examples highlight both the effect of the radiation impedance and size effects, especially at low frequencies.

II. THEORY

Consider a rectangular porous material sample inserted into a rigid planar baffle excited acoustically. The porous material is coupled to a semi-infinite fluid on one of its face (excitation side) and has specific boundary conditions on the other faces. The modified weak integral form associated to the porous material has been given in[7]

$$\begin{aligned}\int_{\Omega_p} [\tilde{\underline{\underline{\sigma}}}^s(\underline{u}):\underline{\underline{\varepsilon}}^s\,.\,(\delta\underline{u}) - \omega^2\tilde{\rho}\underline{u}\,.\,\delta\underline{u}]d\Omega \\ + \int_{\Omega_p}\left[\frac{\phi^2}{\omega^2\tilde{\rho}_{22}}\underline{\nabla}p\,.\,\underline{\nabla}\delta p - \frac{\phi^2}{\tilde{R}}p\delta p\right]d\Omega \\ - \int_{\Omega_p}\frac{\phi^2\rho_0}{\tilde{\rho}_{22}}\delta(\underline{\nabla}p\,.\,\underline{u})d\Omega - \int_{\Omega_p}\phi\left(1+\frac{\tilde{Q}}{\tilde{R}}\right) \\ \times\delta(p\underline{\nabla}\,.\,\underline{u})d\Omega - \int_{\partial\Omega_p}\phi[\underline{U}\,.\,\underline{n} - \underline{u}\,.\,\underline{n}]\delta p d\Gamma \\ - \int_{\partial\Omega_p}[\underline{\underline{\sigma}}^t\underline{n}]\,.\,\delta\underline{u}d\Gamma = 0 \quad \forall\,(\delta\underline{u},\delta p),\end{aligned} \tag{1}$$

Ω_p and $\partial\Omega_p$ refer to the porous-elastic domain and its bounding surface; $\underline{u}$ and p are the solid phase displacement vector and the interstitial pressure in the porous-elastic medium, respectively. $\underline{U}$ is the fluid macroscopic displacement vector. $\delta\underline{u}$ and δp refer to their admissible variation, respectively. $\underline{n}$ denotes the unit normal vector external to the bounding surface $\partial\Omega_p$. $\tilde{\underline{\underline{\sigma}}}^s$ and $\underline{\underline{\varepsilon}}^s$ are the *in vacuo* stress and strain tensors of the porous material. $\tilde{\underline{\underline{\sigma}}}^t$ is the total stress tensor of the material given by $\tilde{\underline{\underline{\sigma}}}^s = \underline{\underline{\sigma}}^t + \phi[1+(\tilde{Q}/\tilde{R})]p\underline{\underline{1}}$. Note that $\tilde{\underline{\underline{\sigma}}}^s$ accounts for structural damping in the skeleton through a complex Young's modulus $E(1+j\eta_s)$. ϕ stands for the porosity, $\tilde{\rho}_{22}$ is the modified Biot's density of the fluid phase accounting for viscous dissipation, $\tilde{\rho}$ is an effective density[10] given by $\tilde{\rho}=\tilde{\rho}_{11}-(\tilde{\rho}_{12}/\tilde{\rho}_{22})$ where $\tilde{\rho}_{11}$ is the modified Biot's density of the solid phase accounting for viscous dissipation. $\tilde{\rho}_{12}$ is the modified Biot's density which accounts for the interaction between the inertia forces of the solid and fluid phase together with viscous dissipation. $\tilde{Q}$ is an elastic coupling coefficient between the two phases, $\tilde{R}$ may be interpreted as the bulk modulus[10] of the air occupying a fraction of the unit volume aggregate, and $\underline{\underline{1}}$ is the identity matrix. In this formulation, the porous media couples to the semi-infinite fluid medium through the following boundary terms:

$$I_{\partial\Omega_p} = -\int_{\partial\Omega_p}[\underline{\underline{\sigma}}^t\,.\,\underline{n}]\delta\underline{u}d\Gamma - \int_{\partial\Omega_p}\phi(U_n-u_n)\delta p d\Gamma. \tag{2}$$

Since at the free surface, $\underline{\underline{\sigma}}^t.\underline{n} = -p\underline{n}$, Eq. (2) becomes

$$I_{\partial\Omega_p} = \int_{\partial\Omega_p}\delta(pu_n)d\Gamma - \int_{\partial\Omega_p}[\phi(U_n-u_n)+u_n]\delta p d\Gamma. \tag{3}$$

The continuity of the normal displacement at the surface results in the form[5]

$$\phi(U_n-u_n)+u_n = \frac{1}{\rho_0\omega^2}\frac{\partial p_a}{\partial n}.$$

Applying this relation, Eq. (3) becomes

$$I_{\partial\Omega_p} = \int_{\partial\Omega_p}\delta(pu_n)d\Gamma - \frac{1}{\rho_0\omega^2}\int_{\partial\Omega_p}\frac{\partial p_a}{\partial n}\delta p d\Gamma. \tag{4}$$

In the semi-infinite domain, the acoustic pressure p_a is the sum of the blocked pressure p_b and the radiated pressure p_r. So, $\partial p_a/\partial n = \partial p_r/\partial n$, and in the numerical implementation context, the discrete form associated with the second term of (4) reads:

$$\frac{-1}{\rho_0\omega^2}\int_{\partial\Omega_p}\frac{\partial p_a}{\partial n}\delta p d\Gamma = \frac{-1}{\rho_0\omega^2}\langle\delta p\rangle[C]\left\{\frac{\partial p_r}{\partial n}\right\}, \tag{5}$$

where $[C]$ is the classical coupling matrix given by

$$[C] = \int_{\partial\Omega_p}\langle N(M)\rangle\{N(M)\}d\Gamma(M) \tag{6}$$

with $\{N(M)\}$ denoting the vector of the used surface element shape functions and $\langle N(M)\rangle$ its transpose.[14]

The porous material being inserted into a rigid baffle, the radiated acoustic pressure, is related to the normal velocity via Rayleigh's integral

$$p_r(x,y,z) = -\int_{\partial\Omega_p}\frac{\partial p_r(x',y',0)}{\partial n}G(x,y,0,x',y',0)d\Gamma(M'), \tag{7}$$

where $G(x,y,0,x',y',0) = e^{-jk_0R}/2\pi R$ is the baffled Green's function, $k_0=\omega/c_0$, is the acoustic wave number in the fluid, c_0 is the associated speed of sound, and R is the distance between point $(x,y,0)$ and $(x',y',0)$ $R=\sqrt{(x-x')^2+(y-y')^2}$.

An associated integral form to Eq. (7) is given by

$$\begin{aligned}\int_{\partial\Omega_p}p_r(x,y,z)\delta p d\Gamma = -\int_{\partial\Omega_p}\int_{\partial\Omega_p}\frac{\partial p_r(x,y,0)}{\partial n} \\ \times G(x,y,0,x',y',0)\delta p d\Gamma(M)d\Gamma(M').\end{aligned} \tag{8}$$

The associated discrete form is

$$\langle \delta p \rangle [C] \langle p_r \rangle = - \langle \delta p \rangle [Z] \left\{ \frac{\partial p_r}{\partial n} \right\} \tag{9}$$

with

$$[Z] = \int_{\partial\Omega_p} \int_{\partial\Omega_p} \{N(M)\} G(M,M') \times \langle N(M') \rangle d\Gamma(M) d\Gamma(M'). \tag{10}$$

Since $\langle \delta p \rangle$ is arbitrary, one gets

$$\left\{ \frac{\partial p_r}{\partial n} \right\} = - [Z]^{-1} [C] \langle p_r \rangle. \tag{11}$$

Substituting Eq. (11) into Eq. (5), and recalling that on the interface $p = p_a = p_b + p_r$, the discrete form of Eq. (4) finally reads

$$I_{\partial\Omega_p} = \langle \delta u_n \rangle [C] \{p\} + \langle \delta p \rangle [C]^T \{u_n\} - \frac{1}{j\omega} \langle \delta p \rangle [\tilde{A}] \{p - p_b\}, \tag{12}$$

where

$$[\tilde{A}] = \frac{1}{j\omega\rho_0} [C][Z]^{-1}[C] \tag{13}$$

is an admittance matrix. In consequence, the radiation of the porous medium into the semi infinite fluid amounts to an admittance term added to the interface interstitial pressure degrees of freedom and to additional interface coupling terms between the solid phase and the interstitial pressure [first two terms in Eq. (12)]. Note that the last term involving p_b is the excitation term and disappears in the case of free radiation.

Using classical notations,[4,6] the discretized form of (1) combined with (12) leads to the following linear system:

$$\begin{bmatrix} -\omega^2 [\tilde{M}] + [K] & -[\tilde{C}] + [C] \\ -[\tilde{C}]^T + [C]^T & \dfrac{[\tilde{H}]}{\omega^2} - \dfrac{[\tilde{A}]}{j\omega} - [\tilde{Q}] \end{bmatrix} \begin{Bmatrix} u \\ p \end{Bmatrix} = \begin{Bmatrix} 0 \\ F_f \end{Bmatrix} \tag{14}$$

with

$$\{F_f\} = \frac{1}{j\omega} [\tilde{A}] \{p_b\}, \tag{15}$$

where $\{u\}$ and $\{p\}$ represent the solid phase and the fluid phase global nodal variables, respectively. $[\tilde{M}]$ and $[K]$ represent equivalent mass and stiffness matrices for the solid phase, $[\tilde{H}]$ and $[\tilde{Q}]$ represent equivalent kinetic and compression energy matrices for the fluid phase, and finally $[\tilde{C}]$ is a volume coupling matrix between the solid phase displacement variables and the fluid phase pressure variable. These matrices result from discretization of the various terms of Eq. (1); their expressions can be found in Ref. 4. The system of Eq. (14) is first solved in terms of the porous solid phase nodal displacements and interstitial nodal pressures. Next, the vibroacoustic indicators of interest can then be calculated.

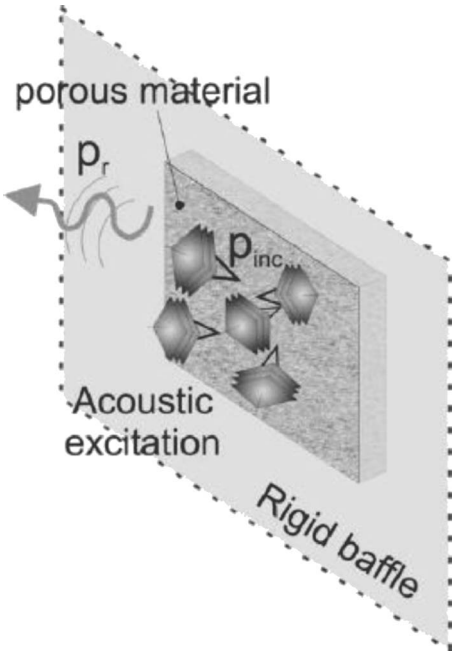


FIG. 1. Configuration of the problem.

III. NUMERICAL EXAMPLES

This section describes several examples illustrating the application and validity of the presented approach. The first two examples discuss the lateral size effects on the acoustic absorption and transmission loss of a foam. In particular, the presented method is compared to the classical Transfer Matrix Method and its extension accounting for finite size corrections (FTMM). The third example tackles the effects of radiation loss on a vibrating plate-foam system. And finally, an experimental validation, in the case of the random incidence transmission loss of a plate-foam system is presented in Sec. IV.

A. Oblique incidence absorption of a foam: Finite size effects

The first example considers an absorption problem and consists of a 0.5 m $\times$ 0.5 m $\times$ 5.08 cm rectangular sample of foam backed by a rigid wall and excited by a plane wave (see Fig. 1). The properties of the foam are given in Table I. A mesh of $22 \times 22 \times 15$ linear brick poroelastic elements is used for the foam for the normal incidence case (0°,0°) while a larger mesh of $32 \times 32 \times 15$ elements was used for the oblique plane wave (45°,0°). These meshes have been selected to assure convergence of calculations. The used meshing criteria are classical for the plate and are functions of the Biot's wavelengths for the foam.[15] Figures 2 and 3 show the corresponding absorption coefficients. The absorption coefficient is calculated using a power balance method[11]

$$\alpha(\theta, \varphi, \omega) = \frac{\Pi_{\text{diss}}}{\Pi_{\text{inc}}}. \tag{16}$$

In this expression Π_{diss} denotes the power dissipated in the foam and Π_{inc} denotes the incident power. The results obtained using the Transfer Matrix Method (TMM) (Ref. 10) is also shown for comparison. As expected, it is clearly

TABLE I. Properties of the plate and foam used in the examples.

Plate	
Lateral dimensions	0.35 m × 0.22 m
Thickness	1 mm
Mass density	2742 kg/m^3
Young's modulus	69 GPa
Poisson ratio	0.33
Loss factor	0.007
Foam	
Thickness	5 cm
Porosity	0.99
Flow resistivity	10900 N s/m^4
Tortuosity	1.02
Viscous characteristic length	130 μm
Thermal characteristic length	192 μm
Mass density	8.43 kg/m^3
Young's modulus	195 kPa
Poisson ratio	0.42
Loss factor	0.05

seen that at higher frequencies, the presented approach asymptotes to the infinite extent configuration since the normalized radiation impedance reduces to 1. Note that a finite size correction to the transfer matrix method can be obtained by replacing the infinite extent radiation efficiency $Z_{R,\mathrm{inf}}(\theta)=1/\cos\theta$ by the baffled window equivalence[16,17]

$$Z_R = \frac{ik_0}{S}\int_S\int_S e^{-jk_0 \sin\theta(\cos\varphi x_0+\sin\varphi y_0)} \times G(M,M_0)e^{jk_0 \sin\theta(\cos\varphi x+\sin\varphi y)}d\Gamma(M_0)d\Gamma(M). \tag{17}$$

The absorption coefficient accounting for the finite size is then given by

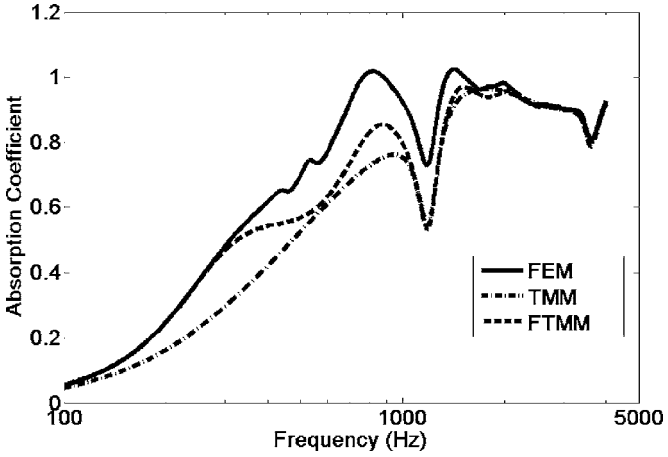


FIG. 2. Absorption coefficient of 2 in. thick foam slab of dimensions 0.5 m × 0.5 m excited by a normal incidence plane wave.

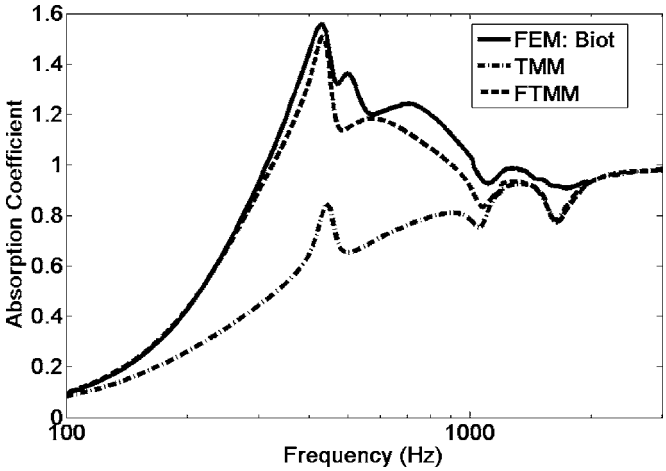


FIG. 3. Absorption coefficient of 2 in. thick foam slab of dimensions 0.5 m × 0.5 m excited by an oblique plane wave.

$$\alpha_f(\theta,\varphi) = \frac{\Pi_{\mathrm{abs},f}}{\Pi_{\mathrm{inc}}} = \frac{1}{\cos\theta}\frac{4\Re[Z_A]}{|Z_A+Z_R(\theta,\varphi)|^2}, \tag{18}$$

where Z_A is the normal surface impedance of the material. This expression extended to the diffuse field case has been shown to compare well with experimental data.[16] It is used here to compute the curve referred to as FTMM in Figs. 2 and 3. As expected it captures well the initial slope of the absorption curve which is governed by the size of the tested sample. Note that as the sample becomes larger, the FE and FTMM curves converge towards the TMM result. The same results are obtained for an oblique incidence plane wave, Fig. 3.

B. Oblique incidence transmission loss of a foam

The second example considers the transmission loss of the same foam sample, which is now between two semi-infinite media. The results are shown in Fig. 4 for an oblique incidence plane wave excitation (45°,0°) and compared to the TMM results. Once again and as expected the two methods lead to similar results at high frequencies. At low frequencies, the size effects lead to an increase in the transmis-

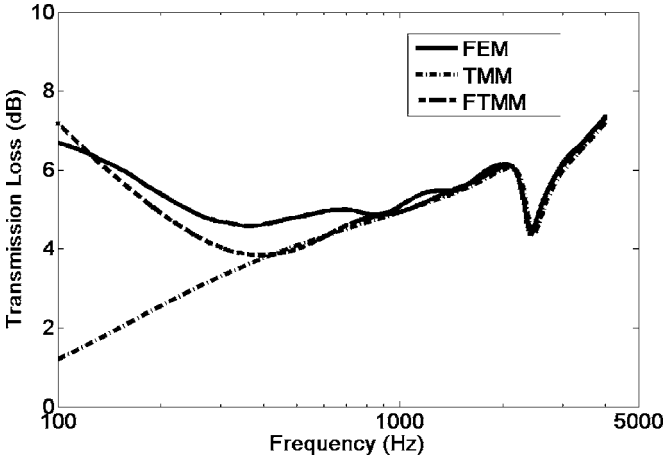


FIG. 4. Transmission loss of a 2 in. thick foam slab of dimensions 0.5 m × 0.5 m excited by an oblique incidence plane wave.

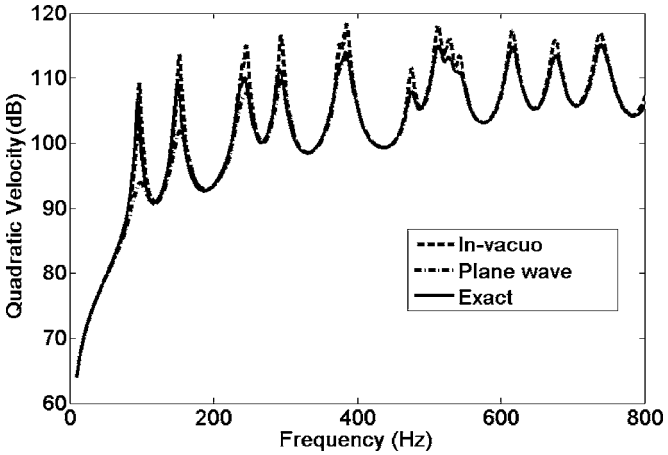


FIG. 5. Mean-square velocity of a plate-foam system radiating in air.

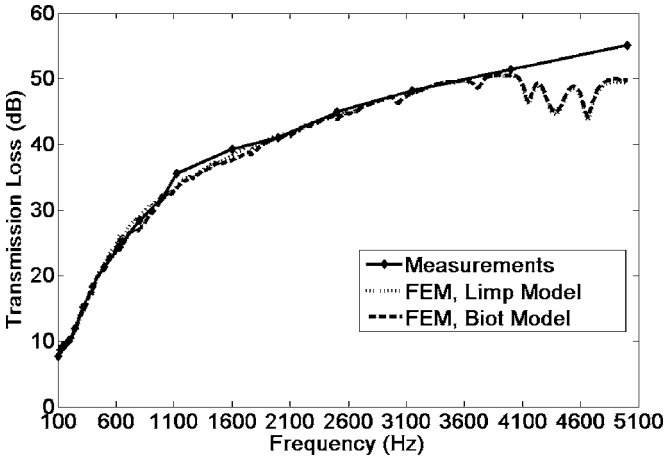


FIG. 6. Transmission loss of a plate-foam system. Tests versus FEM predictions: narrow band comparison using 1000 logarithmic frequency points.

sion loss. Again, a finite size correction can be applied to the TMM and used to check the predicted results at low frequencies. It is given by[16]

$$\tau_f(\theta,\varphi) = \tau_{\text{inf}}(\theta)\frac{\Re[Z_R(\theta,\varphi)]}{Z_{R,\text{inf}}(\theta)}\frac{|Z_A + Z_{R\ \text{inf}}(\theta)|^2}{|Z_A + Z_R(\theta,\varphi)|^2}, \tag{19}$$

where $\tau_{\text{inf}}(\theta)$ denotes the infinite extent transmission coefficient. It is seen that with this simple correction the low frequency behavior is acceptably predicted (within 1 dB). Again, for large samples, the results will asymptote towards the TMM result.

C. Radiation effects of a plate-foam system

The third example considers a plate-foam system excited by a point force. The material properties and dimensions of the plate and the foam are listed in Table I. The frequency range of interest is the range below 800 Hz. The plate is meshed using 24×15 thin shell elements and the porous material is meshed using $24\times15\times9$ linear brick poroelastic elements. The plate is assumed to be simply supported. The foam is bonded onto the plate, clamped (bonded to a hard baffle) along its edges and has a free face which can radiate into a semi-infinite space. The mean-square velocity of the plate is shown in Fig. 5 for three configurations: (i) using the radiation impedance condition (exact approach), (ii) neglecting the radiation condition (i.e., $p=0$ on the free face), and (iii) assuming a plane wave radiation condition ($Z_R=\rho_0 c$). It is clearly seen that the three results are similar except at the resonances where the effect of radiation damping is clearly highlighted, especially at low frequencies. Moreover, it is seen that the exact approach and the plane wave approximation lead to similar results at higher frequencies. However, the plane wave approximation clearly overestimates the radiation damping at low frequencies. The result also highlights the small effects of the radiation efficiency of the foam. Calculations using other foam materials with different properties (especially several range of flow resistivity) lead to similar conclusions.

IV. AN EXPERIMENTAL VALIDATION

A transmission loss test was performed on a flat aluminum panel with attached foam. The panel dimensions were $1.64\ \text{m}\times1.19\ \text{m}\times1.016$ mm and the foam dimensions are $1.64\ \text{m}\times1.19\ \text{m}\times7.62$ cm. The foam is attached along its edges to the panel using a double faced tape. The tests were performed at the Groupe Acoustique de l'Université de Sherbrooke (GAUS) transmission loss facility. The facility utilizes a semianechoic-reverberant transmission loss suite. The reverberation room dimensions are $7.5\ \text{m}\times6.2\ \text{m}\times3\ \text{m}$ with a Schroëder frequency of 200 Hz and a reverberation time of 5.3 s at 1000 Hz. The free volume of the semianechoic chamber is $6\ \text{m}\times7\ \text{m}\times3\ \text{m}$ with an operational frequency range from 200 Hz to 80 kHz. The plate is secured in a mounting window between the two chambers. The intensity technique is used to determine the transmission loss. The technique follows closely standard ISO 15186-1: 2000. The reverberation chamber is excited using six loudspeakers and sound power is captured using a rotating boom microphone. On the anechoic side, the sound intensity is measured in the reception side using an automated arm and an intensity probe with a 12 mm spacer between two $\frac{1}{2}$ in. microphones. This allows measurements to be carried out from 100 Hz to 5000 Hz.

The measured transmission loss and predictions with both the FTMM and the presented approach (FEM/BEM) are given in Figs. 6 and 7. In the latter method, a mesh of $45\times34\times9$ linear brick poroelastic elements was used for the foam. This mesh was compatible with the plate's mesh (45×34 Quad4 shell elements). Since the foam was only attached along its edges to the plate, the coupling boundary condition was modeled as an air gap inserted between the two components for both the FTMM approach and the FEM/BEM approach. The air gap was modeled with linear acoustic 8 nodes brick elements. The diffuse field was modeled as a superposition of plane waves using a GAUSS integration scheme of 6×6 points (6 plane waves along θ and 6 plane waves along φ). It should be noted at this stage that the mounted panel damping was not measured, and that a nominal modal damping ratio of 3% was assumed in the analysis (this value is justified by edge damping). Since the measure-

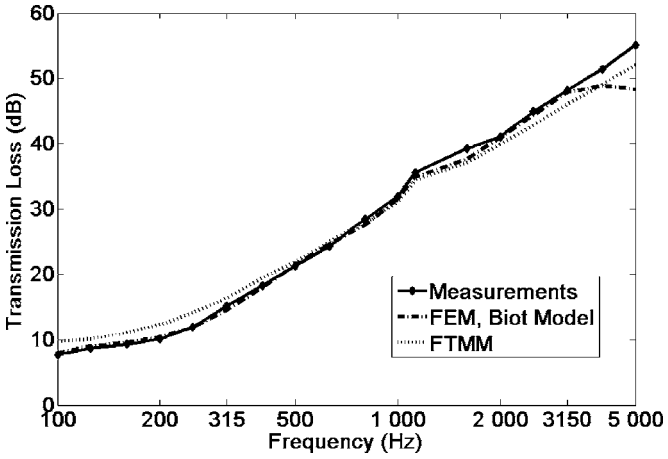


FIG. 7. Transmission loss of a plate-foam system. Tests versus FTMM and FEM predictions (1/3 octave band comparison).

ments were done in 1/3 octave bands, the FTMM results were calculated at the bands center frequencies while the FEM/BEM results were calculated at 1000 frequency points using a logarithmic frequency step (Fig. 6) and converted to 1/3 octave bands (Fig. 7). Figure 7 indicates that FTMM leads to a very good agreement throughout the frequency range of the test (up to 5000 Hz), apart from a slight overestimation at low frequencies (below 500 Hz). This corroborates the effectiveness of this method for predicting transmission loss of multilayer systems over the whole frequency range. Equally, the FEM method shows excellent comparison apart at higher frequencies, where it diverges as a consequence of the used mesh (Figs. 6 and 7). Still at these high frequencies there is not interest in using this deterministic method; the FTMM and even the TMM does an excellent job. In summary, at low to mid frequencies, both the FTMM and the present method lead to excellent results. Finally, it is worth mentioning that for this particular problem (unbonded foam) similar results can be obtained using a Limp model for the foam thus diminishing considerably the computational effort (Fig. 6).

V. CONCLUSION

This paper presented a numerical approach to predict the sound radiation of baffled poroelastic media including fluid loading effects. The approach uses a mixed FEM-BEM approach where the fluid loading is accounted for using an admittance matrix and solid phase-interstitial pressure coupling terms. The method is general and can be used as long as the porous material is inserted in a rigid baffle and radiates into a semi-infinite fluid, as demonstrated in the presented examples. Both numerical examples and a transmission loss test have been presented to validate the presented methodology and show the effect of radiation of porous materials.

ACKNOWLEDGMENTS

The authors acknowledge the financial support of AUTO21 (The Automobile of the 21st Century Canadian Network of Centers of Excellence) and the use of RQCHP (Quebec High Performance Computing Network) computational infrastructure for the calculations presented in this paper.

[1]R. Panneton and N. Atalla, "Numerical prediction of sound transmission through multilayer systems with isotropic poroelastic materials," J. Acoust. Soc. Am. **100**, 346–354 (1996).
[2]V. Easwaran, W. Lauriks, and J. P. Coyette, "Displacement-based finite element method for guided wave propagation problems: Application to poroelastic media," J. Acoust. Soc. Am. **100**, 2989–3002 (1996).
[3]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).
[4]N. Atalla, R. Panneton, and P. Debergue, "A mixed displacement pressure formulation for poroelastic materials," J. Acoust. Soc. Am. **104**, 1444–1452 (1998).
[5]P. Debergue, R. Panneton, and N. Atalla, "Boundary conditions for the weak formulation of the mixed (u,p) poroelasticity problem," J. Acoust. Soc. Am. **106**, 2383–2390 (1999).
[6]F. Sgard, N. Atalla, and J. Nicolas, "A numerical model for the low-frequency diffuse field sound transmission loss of double-wall sound barriers with elastic porous lining," J. Acoust. Soc. Am. **108**, 2865–2872 (2000).
[7]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).
[8]N. E. Hörlin, M. Nordström, and P. Göransson, "A 3D hierarchical FE formulation of Biot's equations for elastoacoutsic modeling of porous media," J. Sound Vib. **254**, 633–652 (2001).
[9]S. Rigobert, N. Atalla, and F. Sgard, "Investigation of the convergence of the mixed displacement pressure formulation for three-dimensional poroelastic materials using hierarchical elements," J. Acoust. Soc. Am. **114**, 2607–2617 (2003).
[10]J.-F. Allard, *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Elsevier, New York, 1993).
[11]N. Atalla, F. Sgard, X. Only, and R. Panneton, "Acoustic absorption from macro-perforated porous materials," J. Sound Vib. **243**, 659–678 (2001).
[12]K. V. Horoshenkov and K. Sakagami, "A method to calculate the acoustic response of a thin, baffled, simply supported poroelastic plate," J. Acoust. Soc. Am. **110**, 904–917 (2000).
[13]D. Takahashi and M. Tanaka, "Flexural vibration of perforated plates and porous elastic materials under acoustic loading," J. Acoust. Soc. Am. **112**, 1456–1464 (2002).
[14]J. L. Batoz, and G. Dhatt, *Modélisation des Structures par Éléments Finis, Volume 2, Poutres et Plaques* (Hermès, Paris, 1990).
[15]N. Dauchez, S. Sahraoui, and N. Atalla, "Convergence of poroelastic finite elements based on Biot displacement formulation," J. Acoust. Soc. Am. **109**, 33–41 (2001).
[16]N. Atalla, S. Ghinet, and O. Haisam, "Transmission Loss of Curved Sandwich Composite Panels," 18th ICA, Kyoto (2004).
[17]S. Ghinet and N. Atalla, "Vibro-acoustic behaviour of multi-layer orthotropic panels," Can. Acoust. **30**, 72–73 (2002).

Modal parameter estimation for fluid-loaded structures from reduced order models

Xianhui Li and Sheng Li
Department of Naval Architecture, Dalian University of Technology, Dalian 116024, People's Republic of China

(Received 12 December 2005; revised 30 June 2006; accepted 12 July 2006)

A model reduction method is developed to estimate modal parameters of fluid-loaded structures. The method uses a matrix-free formulation of rational Krylov projection to construct reduced order models of the fluid-loaded structures from forced responses at selected interpolation frequencies. Due to small sizes of the reduced order models, eigenpairs of the associated eigenvalue problems are available at a very low computational cost. Resonance frequencies, modal damping ratios, and mode shapes of the original systems are recovered from the forced responses and the eigenpairs of the reduced order models. Nonphysical modes introduced by the model reduction process are filtered out by a modal damping ratio test and double checked by the condition number of the dynamic stiffness matrix. Efficiency and accuracy of the present method are demonstrated on a benchmark model of a water-loaded plate clamped in a baffle. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2266574]

PACS number(s): 43.40.Rj [DSB] Pages: 1996–2003

I. INTRODUCTION

The problem of determining resonance frequencies, modal damping ratios, and mode shapes for fluid-loaded structures is of significant practical interest. Among many possible solution strategies, state-space methods provide a systematical way to determine these modal parameters by recasting the original systems into generalized eigenvalue problems. Based on the well-established coupled finite element and boundary element (FE/BE) analysis, Giordano and Koopmann[1] introduced a state-space method for the direct solution of fluid-loaded resonance frequencies and mode shapes. The method approximates the acoustic impedance matrix by a polynomial fit to remove its implicit frequency dependence. Coupling of the structural and acoustic matrices results in a canonical state-space form and a generalized eigenvalue problem is formulated. While Giordano and Koopmann used an acoustic impedance representation based on surface velocities for the least-squares fit, Cunefare and De Rosa[2] proposed a simple modification by using an impedance definition in terms of surface displacements. The modification reduces the order of the system and improves the computational efficiency and the numerical performance. Cunefare and De Rosa[2] also showed that the state-space method is capable of producing both correct and incorrect eigenvalues and both global and local correlation measures should be used to ensure the accuracy of the results. Recent work by Li[3] showed that a power series expansion or a least-squares fit of the frequency-dependent term in the Rayleigh integral can also lead to a polynomial approximation for the acoustic impedance matrix, by which a canonical state-space form is derived without direct polynomial fit for the acoustic impedance matrix.

This paper develops a model reduction method for determining modal parameters of fluid-loaded structures. The main idea is that if the dynamic behavior of a system can be reproduced with sufficient accuracy by a reduced order model (ROM), then the eigeninformation of the original system is preserved in the ROM. Once the high-fidelity ROM is constructed, modal parameters of the original system are readily recovered from the small-sized ROM. Although there exist many projection-based methods (e.g., Krylov subspace projection[4–6]) which can reduce a large linear system to a smaller one, these methods cannot be applied without modification to a nonlinear system having fully populated matrix with frequency dependence. In recent papers Ruhe[7,8] suggested a nonlinear rational Krylov method by knitting a secant method for linearizing the system and the rational Krylov method for the linearized problem. However, this method needs frequently updating shifts to get a better and better linear approximation of the original system, which is very time consuming due to the intensive integration involved in generating the frequency-dependent matrix. Moreover, this method requires explicit operation on the system matrices, which may restrict its application when such matrices are not accessible.

In order to construct the ROM efficiently, a matrix-free formulation of rational Krylov projection[9] is employed in this paper. It uses forced responses at selected interpolation frequencies to construct the ROM without directly referring to the system matrices of the original system. Frequency dependence of the acoustic impendence matrix in the coupled structural acoustic system is automatically removed and a generalized eigenvalue problem is derived. Eigenpairs of the ROM eigenvalue problem are available at a very low computational cost due to its small size. Resonance frequencies, modal damping ratios, and mode shapes of the original system are then recovered from the interpolated forced responses and the ROM eigenpairs. Similar to the earlier state-space methods,[1–3] some nonphysical modes are introduced by the model reduction process, therefore, modal damping ratio and condition number tests are used to discriminate the

true modes from the nonphysical ones. Details of ROM construction and recovery process are described in the rest of the paper.

II. ACOUSTIC MODEL

For acoustic radiation problems in free space, the boundary element formulation is based on the Helmholtz integral equation

$$C(P)p(P) = \int_S \left(\frac{\partial G(Q,P)}{\partial n} p(Q) - G(Q,P) \frac{\partial p(Q)}{\partial n} \right) dS(Q), \tag{1}$$

where $p(P)$ is the acoustic pressure at a field point P, which may be outside, inside, or on S; Q is any point on S; $G(Q,P)=e^{-ikR}/4\pi R$ is the free-space Green's function in which $R=|Q-P|$, $k=\omega/c$ is the wave number, ω is the circular frequency, and c is the speed of sound; n is the outward unit normal on S. The coefficient $C(P)$ is given by[10]

$$C(P) = 1, \quad P \text{ outside } S, \tag{2}$$

$$C(P) = 0, \quad P \text{ inside } S, \tag{3}$$

and

$$C(P) = 1 - \int_S \frac{\cos\beta}{4\pi R^2} \mathrm{d}S(Q), \quad P \text{ on } S, \tag{4}$$

where β is the angle between the normal n and the vector R. Equation (4) includes the possibility that the surface may have a nonsmooth geometry such as edges and corners. On the boundary, the normal derivative of the acoustic pressure is related to the normal velocity v_n through the momentum equation

$$\frac{\partial p}{\partial n} = -i\omega\rho v_n, \tag{5}$$

where ρ is the density of the acoustic medium. Convention of time-harmonic term $\exp\{i\omega t\}$ is followed throughout the paper.

The discretization of the surface Helmholtz integral equation (P on S) leads to

$$\mathbf{Ep} = \mathbf{Dv}_n, \tag{6}$$

where $\mathbf{E}$ and $\mathbf{D}$ are the assembled coefficient matrices, $\mathbf{p}$ and $\mathbf{v}_n$ are the surface pressure vector and the normal velocity vector.

It should be noted that for an exterior problem the previous formulation based on the surface Helmholtz integral equation may fail to yield a unique solution at characteristic frequencies of the associated interior Dirichlet problem. To avoid the difficulty, the combined Helmholtz integral equation formulation (CHIEF) method[11,12] is employed. The CHIEF method augments the surface Helmholtz integral equation with the interior Helmholtz equation (P inside S) which yields an overdetermined system of equations. A least-squares solution of $\mathbf{p}$ is given by

$$\mathbf{p} = \mathbf{E}^{+}\mathbf{Dv}_n, \tag{7}$$

where $\mathbf{E}^{+}$ stands for the Moore-Penrose inverse of $\mathbf{E}$.

If a planar surface extends over an infinite half-space, the acoustic pressure at any field point P according to the Rayleigh integral is described as follows:[13]

$$p(P) = \mathrm{i}\omega\rho \int_S e^{-\mathrm{i}kR} v_n(Q)/2\pi R \mathrm{d}S, \tag{8}$$

where $p(P)$ is the acoustic pressure at a field point P and $v_n(Q)$ is the normal velocity of the vibrating surface at a point Q on the plate surface S. Discretizing the plate surface into elements and interpolating the structural normal velocity and surface pressure, Eq. (8) is rewritten as

$$\mathbf{p} = \mathbf{Dv}_n, \tag{9}$$

where $\mathbf{D}$ denotes the acoustic impedance matrix.

In summary, the surface acoustic pressure is related to the structural normal velocity by

$$\mathbf{p} = \mathbf{Zv}_n, \tag{10}$$

where the acoustic impedance matrix $\mathbf{Z}=\mathbf{E}^{+}\mathbf{D}$ for an arbitrarily shaped three-dimensional body and $\mathbf{Z}=\mathbf{D}$ for a planar surface extending over an infinite half-space.

III. COUPLED STRUCTURAL ACOUSTIC MODEL

For a fluid-loaded structures subjected to time-harmonic excitations, if the structure is modeled with finite elements, the governing equation for the structure is given by[14]

$$[-\omega^2\mathbf{M} + \mathrm{i}\omega\mathbf{C} + \mathbf{K}]\mathbf{x} = \mathbf{Fu} - \mathbf{GAp}, \tag{11}$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are the structural mass, damping, stiffness matrices, $\mathbf{x}$ is the structural displacement vector, $\mathbf{F}$ is the force matrix, $\mathbf{u}$ is the load spectra, $\mathbf{G}$ is a matrix mapping the interface degrees of freedom (DOFs) to the structural DOFs, and matrix $\mathbf{A}=\int_S \mathbf{N}^T\mathbf{N}\mathrm{d}S$, where $\mathbf{N}$ is the matrix of interpolation functions.

For the time-harmonic excitation, the normal velocity vector $\mathbf{v}_n$ is related to the structural displacement vector $\mathbf{x}$ by

$$\mathbf{v}_n = \mathrm{i}\omega\mathbf{G}^T\mathbf{x}. \tag{12}$$

Substituting Eqs. (10) and (12) into Eq. (11) yields

$$[-\omega^2\mathbf{M} + \mathrm{i}\omega(\mathbf{C} + \mathbf{GAZG}^T) + \mathbf{K}]\mathbf{x} = \mathbf{Fu}, \tag{13}$$

which contains the structural displacements as the only unknown variables to describe the coupled structural acoustic system. Although the physical model of the coupled system is linear, the eigenvalue problem associated with the governing Eq. (13) is nonlinear due to the frequency dependence of the acoustic-impedance matrix. Existing eigensolvers for linear systems cannot be directly applied to such a nonlinear eigenvalue problem.

IV. CONSTRUCTION OF ROM

The earlier state-space methods rely on a polynomial fit to remove the implicit frequency dependence in the acoustic impedance matrix. For a coupled system with ND structural DOFs, when l terms are used in the fitting polynomial, the resulting state-space equation is of order $l\cdot ND$ for the Giordano and Koopmann method,[1] $(l-1)\cdot ND$ for the Cunefare and De Rosa method,[2] or $(l+1)\cdot ND$ for Li's method.[3]

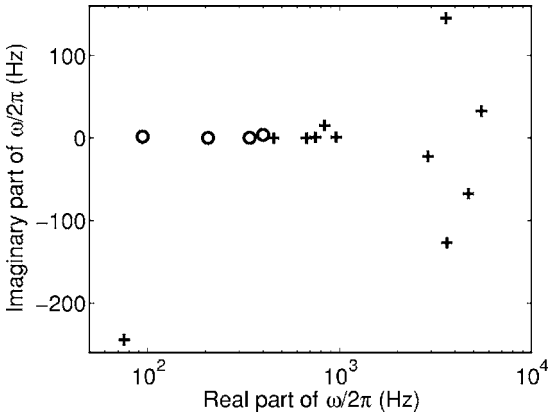


FIG. 1. Distribution of the complex resonance frequencies for the ROM of order 15.

The associated eigenvalue problem is neither symmetric nor sparse, making it computationally expensive to solve for a system with a large number of structural DOFs.

In this paper, a linear time-invariant (LTI) system of reduced order is sought to relieve the above computational difficulty. Suppose there is a large LTI system which can reproduce the dynamic behavior of the coupled system with sufficient accuracy and one of its transfer functions is given by

$$\mathbf{H}(\omega) = \mathbf{L}^H[\bar{\mathbf{K}} - \omega^2\bar{\mathbf{M}}]^{-1}\mathbf{R}, \tag{14}$$

where $\bar{\mathbf{K}}$ and $\bar{\mathbf{M}}$ are the system and description matrices and $\mathbf{R}$ and $\mathbf{L}$ are the input and output matrices. To enhance the numerical efficiency in modal parameter estimation, a projection-based model reduction is applied to the large LTI system. According to the rational Krylov projection,[4,9] the left and right projection subspaces can be defined as

$$\bigcup_{\omega_\alpha \in S_\alpha} \{[\bar{\mathbf{K}} - \omega_\alpha^2\bar{\mathbf{M}}]^{-H}\mathbf{L}\} \subseteq \mathbf{colsp}\{\mathbf{W}\}, \tag{15}$$

$$\bigcup_{\omega_\beta \in S_\beta} \{[\bar{\mathbf{K}} - \omega_\beta^2\bar{\mathbf{M}}]^{-1}\mathbf{R}\} \subseteq \mathbf{colsp}\{\mathbf{V}\}, \tag{16}$$

where the disjoint sets S_α and S_β contain the interpolation frequencies used to generate the projection subspaces and **colsp** denotes the subspace spanned by column vectors. Projecting the large LTI system onto the above rational Krylov subspaces results in a ROM with a transfer function

$$\hat{\mathbf{H}}(\omega) = \hat{\mathbf{L}}^H[\hat{\mathbf{K}} - \omega^2\hat{\mathbf{M}}]^{-1}\hat{\mathbf{R}}, \tag{17}$$

where

$$\hat{\mathbf{K}} = \mathbf{W}^H\bar{\mathbf{K}}\mathbf{V}, \quad \hat{\mathbf{M}} = \mathbf{W}^H\bar{\mathbf{M}}\mathbf{V}, \tag{18}$$

$$\hat{\mathbf{R}} = \mathbf{W}^H\mathbf{R}, \quad \hat{\mathbf{L}} = \mathbf{V}^H\mathbf{L}, \tag{19}$$

are the system, description, input, and output matrices of the ROM and $\hat{\mathbf{H}}$ interpolates the original transfer function $\mathbf{H}$ at the frequencies in $S_\alpha \cup S_\beta$. It should be noted that the explicit form of the large LTI system is not available in the coupled FE/BE formulation, therefore, the ROM cannot be constructed by Eqs. (18) and (19) directly.

To overcome the difficulty, a matrix-free formulation of the rational Krylov projection[9] is employed to reduce the large LTI system in the frequency band of interest. Denote $\mathbf{W}_\alpha$ the left projection subspace corresponding to the interpolation frequency ω_α in S_α and $\mathbf{V}_\beta$ the right projection subspace corresponding to the interpolation frequency ω_β in S_β. According to the matrix-free formation,[9] the (α,β) blocks of the ROM system and description matrices are given by

$$[\hat{\mathbf{K}}]_{\alpha,\beta} = \mathbf{W}_\alpha^H\bar{\mathbf{K}}\mathbf{V}_\beta = \frac{\omega_\alpha^2\mathbf{H}(\omega_\alpha) - \omega_\beta^2\mathbf{H}(\omega_\beta)}{\omega_\alpha^2 - \omega_\beta^2},$$
$$[\hat{\mathbf{M}}]_{\alpha,\beta} = \mathbf{W}_\alpha^H\bar{\mathbf{M}}\mathbf{V}_\beta = \frac{\mathbf{H}(\omega_\alpha) - \mathbf{H}(\omega_\beta)}{\omega_\alpha^2 - \omega_\beta^2}, \tag{20}$$

and the α-th and β-th blocks of the ROM input and output matrices are

$$\hat{\mathbf{R}}_\alpha = \mathbf{W}_\alpha^H\mathbf{R} = \mathbf{H}(\omega_\alpha), \quad \hat{\mathbf{L}}_\beta^H = \mathbf{L}^H\mathbf{V}_\beta = \mathbf{H}(\omega_\beta). \tag{21}$$

Once the transfer functions at the selected interpolation frequencies $\mathbf{H}(\omega_\alpha)_{\omega_\alpha \in S_\alpha}$ and $\mathbf{H}(\omega_\beta)_{\omega_\beta \in S_\beta}$ are known, the ROM can be constructed by Eqs. (20) and (21).

V. RECOVERY OF MODAL PARAMETERS

To identify the modal parameter of the coupled system, the transfer function of the large LTI system is chosen to be

$$\mathbf{H}(\omega) = \mathbf{F}^T[-\omega^2\mathbf{M} + \mathrm{i}\omega(\mathbf{C} + \mathbf{GAZG}^T) + \mathbf{K}]^{-1}\mathbf{F}. \tag{22}$$

Any other transfer function can also be used as long as it captures the dynamics of the coupled system. Constructing the ROM according to Eq. (20) leads to the ROM eigenvalue problem

$$\hat{\mathbf{K}}\hat{\mathbf{\Phi}} = \hat{\mathbf{M}}\hat{\mathbf{\Phi}}\hat{\mathbf{\Lambda}}, \tag{23}$$

where $\hat{\mathbf{\Lambda}} = \mathbf{diag}\{\hat{\omega}_1^2, \hat{\omega}_2^2, \ldots, \hat{\omega}_{mn}^2\}$ is the ROM eigenvalue matrix and $\hat{\mathbf{\Phi}} = [\hat{\boldsymbol{\varphi}}_1, \hat{\boldsymbol{\varphi}}_2, \ldots, \hat{\boldsymbol{\varphi}}_{mn}]$ is the ROM eigenvector

TABLE I. The resonance frequencies and modal damping ratios of the water-loaded modes below 450 Hz.

	ROM (15×15)			State-space method (Ref. 3)			Iterative method (Ref. 15)		
	Resonance frequency (Hz)			Resonance frequency (Hz)			Resonance frequency (Hz)		
Mode index	Vacuum	Water	Damping ratio (%)	Vacuum	Water	Damping ratio (%)	Vacuum	Water	Damping ratio (%)
1	276.0	94.2	1.693	276.0	95.3	1.690	275.2	94.6	1.681
2	441.1	206.4	0.022	441.1	206.4	0.024	440.0	202.4	0.011
3	675.0	339.7	0.057	675.0	329.8	0.042	663.8	329.5	0.034
4	731.4	399.7	0.933	731.4	395.0	0.644	715.2	371.2	0.839

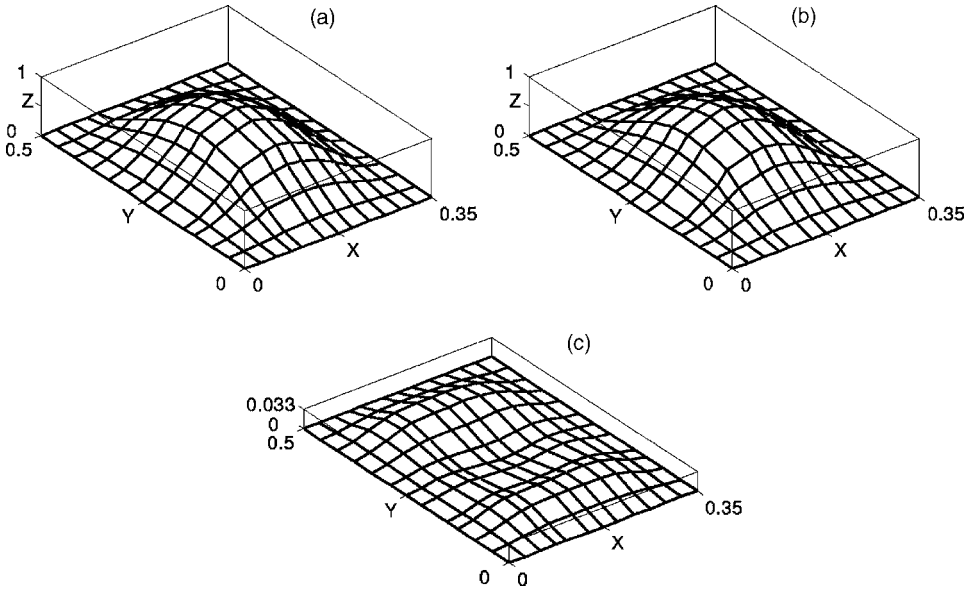


FIG. 2. The first mode shape (a) normalized water-loaded mode, (b) normalized *in vacuo* mode, and (c) difference between the normalized *in vacuo* and the normalized water-loaded mode.

matrix. If the force matrix $\mathbf{F}$ and the interpolation frequencies are appropriately chosen, the eigenvalues of the original system are well approximated by $\hat{\mathbf{\Lambda}}^{1/2}$ and the eigenvectors are recovered by

$$\mathbf{\Phi} = [\mathbf{X}(\omega_{\beta 1}), \mathbf{X}(\omega_{\beta 2}), \ldots, \mathbf{X}(\omega_{\beta n})]\hat{\mathbf{\Phi}}, \tag{24}$$

where

$$\mathbf{X}(\omega_\beta) = [-\omega_\beta^2 \mathbf{M} + \mathrm{i}\omega_\beta(\mathbf{C} + \mathbf{GAZG}^T) + \mathbf{K}]^{-1}\mathbf{F}, \quad \omega_\beta \in S_\beta, \tag{25}$$

are the forced responses at the selected interpolation frequencies.

Suppose n interpolation frequencies are used in each interpolation set and the force matrix $\mathbf{F}$ consists of m probing vectors. The resulting ROM is of the order mn, much smaller than the models obtained by the earlier state-space methods for a fluid-loaded structure with a large number of structural DOFs. In most practical applications, a high-quality ROM can be constructed following the rules:

(1) The interpolation frequencies evenly distributed in the frequency band of interest.
(2) The total number of interpolation frequencies is about the number of fluid-loaded modes in the interpolated frequency band.
(3) The number of the probing vectors in the force matrix is about 5 to 10.

It is obvious that using more interpolation frequencies and probing vectors will improve the ROM quality, but this improvement is achieved at an extra computational cost. Adaptive interpolation based on the current error estimates[4,9] helps to construct the ROM with a quasioptimal placement of the interpolation frequencies, but it is out of the scope of this paper and will not be addressed herein.

It should be emphasized that the eigenpairs recovered from the ROM only approximate those of the original system in the interpolated frequency band, depending on the choice

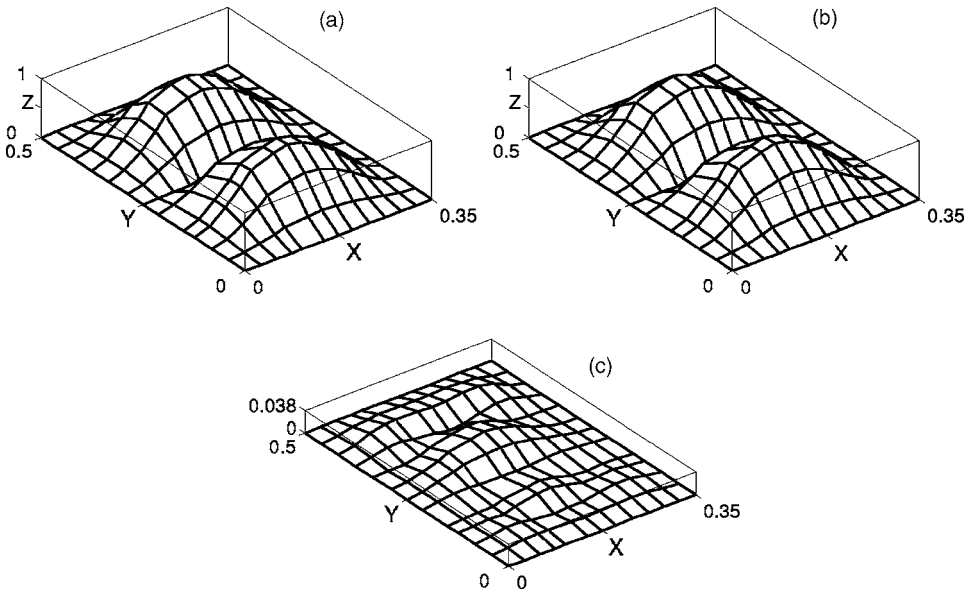


FIG. 3. The second mode shape (a) normalized water-loaded mode, (b) normalized *in vacuo* mode, and (c) difference between the normalized *in vacuo* and the normalized water-loaded mode.

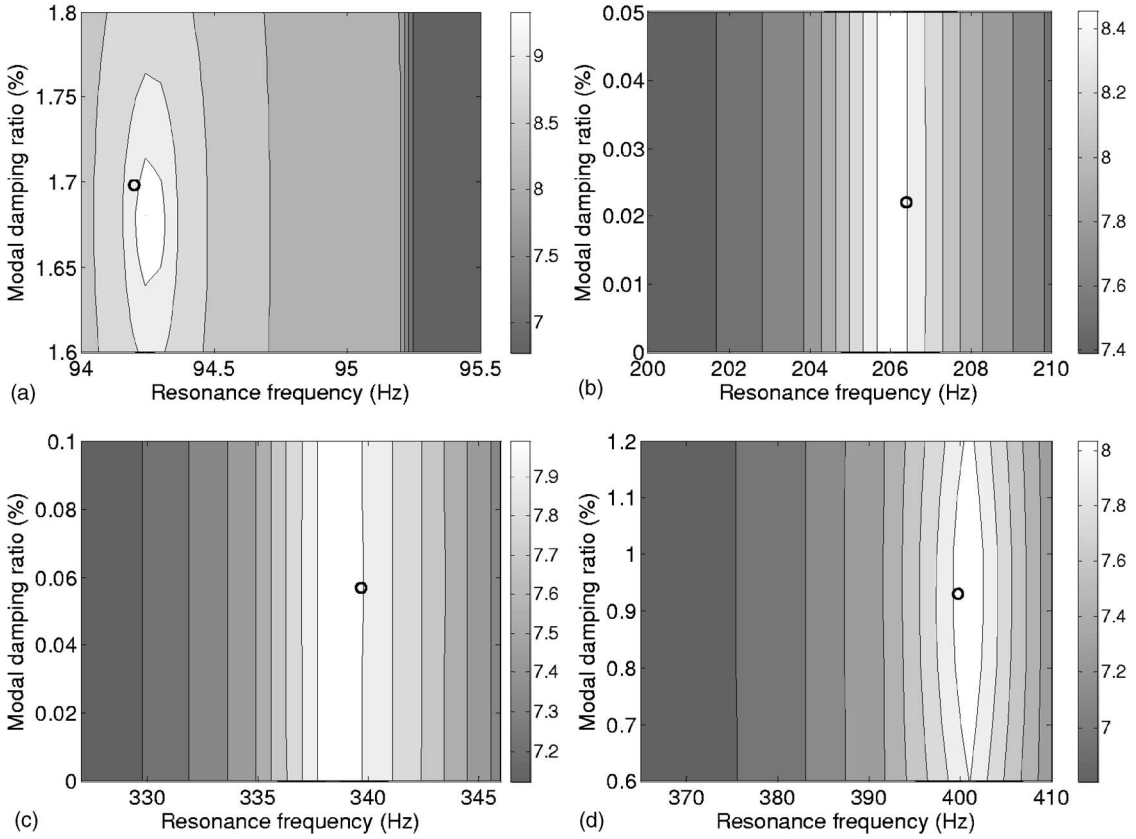


FIG. 4. Contour plot for the logarithm of the condition number of the dynamic stiffness matrix in the vicinity of (a) the first water-loaded mode, (b) the second water-loaded mode, (c) the third water-loaded mode, and (d) the fourth water-loaded mode; "o" denotes the resonance frequency and modal damping ratio recovered from the ROM.

of the interpolation frequencies and the probing vectors. All out-of-band modes should be discarded because of the large approximation error. Moreover, the true modes should have reasonable modal damping ratios (e.g., less than unity according to Ref. 2). For a lightly damped underwater structure (without any special damping treatment), this constraint is quite conservative. In this paper, a reasonable modal damping ratio is chosen to be less than 0.3. In practice, the above criteria can only filter out some obvious nonphysical modes. The remaining modes need to be double checked by the condition number of the dynamic stiffness matrix

$$\mathbf{K}_d(\hat{\omega}_i) = [-\hat{\omega}_i^2\mathbf{M} + \mathrm{i}\hat{\omega}_i(\mathbf{C} + \mathbf{GAZ}(\hat{\omega}_i)\mathbf{G}^T) + \mathbf{K}]. \quad (26)$$

A large condition number of $\mathbf{K}_d(\hat{\omega}_i)$ means that it approaches a singular matrix, therefore, $\hat{\omega}_i$ is a good approximation of the true eigenvalue.

After the true modes in the interpolated frequency band are identified, the resonance frequencies and modal damping ratios are obtained from the ROM eigenvalues by

$$f_i = \mathrm{Re}\{\hat{\omega}_i/2\pi\}, \quad \zeta_i = \frac{\mathrm{Im}\{\hat{\omega}_i\}}{\mathrm{Re}\{\hat{\omega}_i\}}. \quad (27)$$

VI. NUMERICAL RESULTS

Efficiency and accuracy of the present method are demonstrated on a benchmark model of a water-loaded plate clamped in a baffle. The dimensions of the plate are $0.350\ \mathrm{m} \times 0.500\ \mathrm{m} \times 0.005\ \mathrm{m}$. The structural material has a density $7800\ \mathrm{kg/m^3}$, Young's modulus 200 GPa, and Poisson's ratio 0.3. The water has a density $1000\ \mathrm{kg/m^3}$ and sound speed 1500 m/s. No structural damping is considered

TABLE II. The resonance frequencies and modal damping ratios of the water-loaded modes below 750 Hz.

	ROM (20×20)		State-space method (Ref. 2)		Iterative method (Ref. 15)	
Mode index	Resonance frequency (Hz)	Damping ratio (%)	Resonance frequency (Hz)	Damping ratio (%)	Resonance frequency (Hz)	Damping ratio (%)
1	94.2	1.693	95.8	0.947	94.6	1.681
2	206.4	0.022	206.3	0.127	202.4	0.011
3	339.7	0.057	339.8	0.011	329.5	0.034
4	399.7	0.933	399.7	0.950	371.2	0.839
5	453.9	0.004	454.0	0.002	441.5	0.027
6	655.6	0.061	655.6	0.060	624.4	0.001
7	701.2	0.290	701.7	0.304	628.6	0.158

TABLE III. Modal assurance criteria for (1) the water-loaded modes from the ROM and the state-space method (Ref. 2); (2) the water-loaded modes from the ROM and the *in vacuo* modes.

Mode index	1	2	3	4	5	6	7
MAC(1)	1.000	1.000	1.000	1.000	1.000	1.000	0.998
MAC(2)	0.997	0.998	0.998	0.985	0.997	0.992	0.982

so as to compare the results with those obtained by Habault and Filippi,[15] where the water-loaded modes are solved with an iterative method and the first 20 resonance frequencies and modal damping ratios are tabulated. The plate is modeled by Mindlin plate elements and a 12×12 mesh is used for both structural finite element and Rayleigh integral surface discretization, leading to a coupled FE/BE model with 363 DOFs.

In the first example, water-loaded modes below 450 Hz are solved by the present method. Five probing vectors are generated with random entries from a uniform distribution in the interval (0, 1). Six interpolation frequencies are evenly distributed in the frequency band (80,450) Hz. The ROM constructed using Eq. (20) is of the order 15. The ROM eigenvalue problem is solved and the complex resonance frequencies are plotted in Fig. 1. There are four modes located in the interpolated frequency band, which are labeled as the circles in the figure. The corresponding modal parameters are compared with those obtained from the state-space method by Li[3] (with power series $m=3$). The results obtained by the iterative method[15] are also listed in Table I for reference. Good agreement is observed in the resonance frequencies, while the modal damping ratios are agreed within the same magnitude. The mode shapes are recovered from the ROM according to Eq. (24). Figures 2(a) and 3(a) show the magnitudes of the first two water-loaded mode shapes normalized by the maximum normal displacements, which are very similar to those to the corresponding *in vacuo* mode shapes shown in Figs. 2(b) and 3(b). It is observed from Figs. 2(c) and 3(c) that the fluid loading has little influence on the mode shapes for this example.

The accuracy of the recovered resonance frequencies and modal damping ratios are double checked by the condition number of the dynamic stiffness matrix. The contour plots for the logarithms of the condition numbers in the vicinity of the first four water-loaded modes are shown in Fig. 4. The resonance frequencies and modal damping ratios recovered from the ROM are located in the high condition number regions, indicating the accuracy of the present method.

In order to identify water-loaded modes in a wider frequency band, more interpolation frequencies are needed to construct the ROM. In this example, water-loaded modes below 750 Hz are solved by the present method using eight evenly distributed interpolation frequencies and five probing vectors. The resulting ROM is of the order 20. It takes 116.8 s to construct the ROM and only 1.9 s to recover the modal parameters of all the in-band modes. For comparison, the same problem is solved with the state-space method by Cunefare and De Rosa.[2] It takes 117.3 s to fit the acoustic impedance matrix at eight frequencies, resulting in a generalized eigenvalue problem of the order 1089. It takes an additional 232.1 s to find all the in-band modes. The resonance frequencies obtained from both methods are in good agreement as shown in Table II. The modal damping ratios of the first two modes have relatively large discrepancies, but the

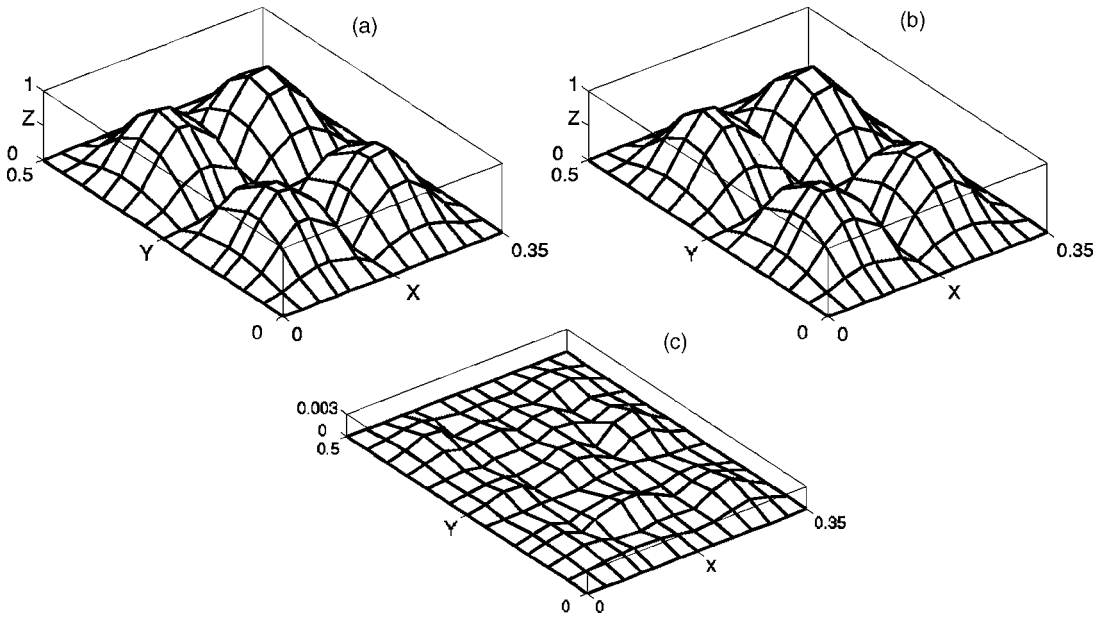


FIG. 5. The fifth water-loaded mode shape (a) normalized water-loaded mode from the ROM, (b) normalized water-loaded mode from the state-space method (Ref. 2); and (c) difference between the normalized water-loaded modes from both methods.

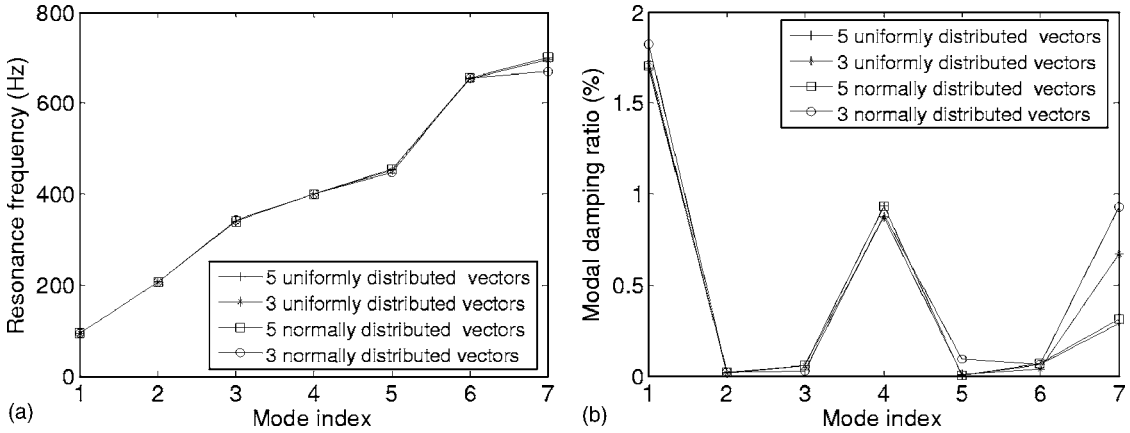


FIG. 6. Modal parameters recovered from the ROM using eight interpolation frequencies and four different force matrices (a) resonance frequencies, and (b) modal damping ratios.

ROM results are closer to those obtained by the iterative method.[15] Modal assurance criteria (MAC) are used to test the correlation between the water-loaded mode shapes obtained from the present method and the state-space method.[2] It is observed from Table III that the MAC is very close to unity, indicating that the difference between the mode shapes from the above methods are negligible, as shown in Fig. 5. MAC between water-loaded mode shapes and *in vacuo* mode shapes are also very close to unity, which confirms the previous observation that fluid loading has little influence on the mode shapes.

The ROM quality is mainly determined by the selection of the interpolation frequencies and the probing vectors. For the case of eight interpolation frequencies evenly distributed in the frequency band (0,750) Hz, force matrices consisting of three or five random probing vectors with entries uniformly or normally distributed in the interval (0, 1) are used to construct the ROM. The recovered resonance frequencies and modal damping ratios are shown in Fig. 6. It is observed that results from the ROM using five probing vectors are in good agreement for all the in-band modes, while the ROM using three probing vectors can only identify the first six modes correctly. If ten interpolation frequencies are used to construct the ROM, then the results using three and five probing vectors are in good agreement for all the in-band modes, as shown in Fig. 7. The total computation time for the ROM using eight interpolation frequencies is about 20% less than that for the ROM using ten interpolation frequencies, therefore, it is advisable to use fewer interpolation frequencies (about the number of in-band modes) and more probing vectors (about five to ten random vectors) in order to construct a high quality ROM efficiently.

VII. CONCLUSIONS

A model reduction method has been developed for determining modal parameters of fluid-loaded structures. The main idea is to approximate the original system with a high-fidelity ROM, which preserves the original eigeninformation in the frequency band of interest. A matrix-free formulation of rational Krylov projection is employed to construct the ROM from the forced responses at selected interpolation frequencies. Frequency dependence of the acoustic impendence matrix in the coupled structural acoustic system is automatically removed and a generalized eigenvalue problem is derived. As compared with the nonlinear rational Krylov method, the present method does not need an intermediate linearization process, reducing the computational cost and the memory requirement in the construction of the ROM. Furthermore, the present method could use transfer functions

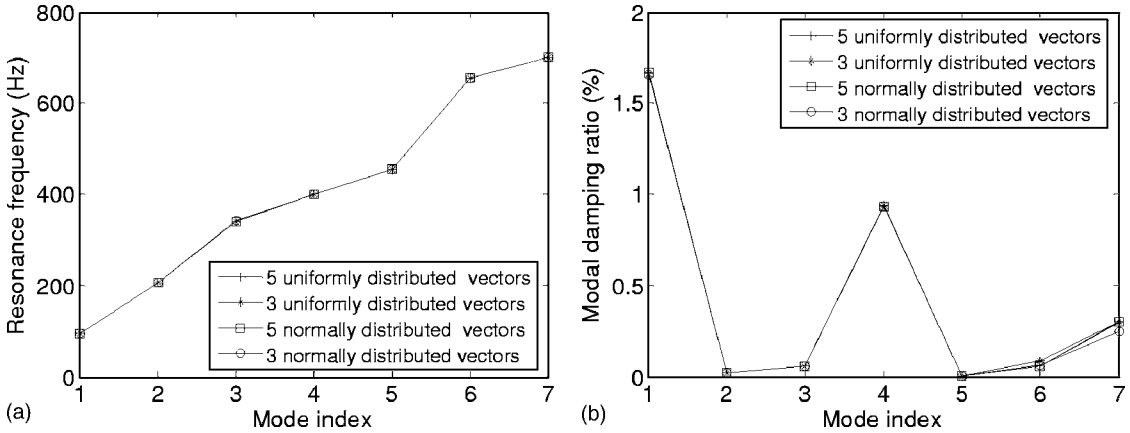


FIG. 7. Modal parameters recovered from the ROM using ten interpolation frequencies and four different force matrices (a) resonance frequencies, and (b) modal damping ratios.

obtained from other methods to construct the ROM, extending its application to the problem when a coupled FE/BE model is not available. For a fluid-loaded structure with a large number of structural DOFs, the ROM eigenvalue problem is usually much smaller than the models obtained by the earlier state-space methods, hence significantly reducing the computational cost. Resonance frequencies, modal damping ratios, and mode shapes of the original system are recovered from the interpolated forced responses and the ROM eigenpairs. Some nonphysical modes are introduced by the model reduction process, which can be effectively filtered out by the modal damping ratio and condition number tests. Several numerical experiments are conducted on a benchmark model of a water-loaded plate clamped in a baffle. Although rigorous bounds of the model reduction error and the recovered modal parameters are not provided in this paper, it is believed that the results are accurate if the ROMs using different combinations of interpolation frequencies and probing vectors yield similar results. Comparison with the state-space methods and the iterative method validates the efficiency and accuracy of the present method in estimating the modal parameters of the fluid-loaded structures.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China. (No. 10402004). The first author is also grateful for the support from Young Teacher Cultivation Foundation of Dalian University of Technology.

[1]J. A. Giordano and G. H. Koopmann, "State space boundary element-finite element coupling for fluid-structure interaction analysis," J. Acoust. Soc. Am. **98**, 363–372 (1995).
[2]K. A. Cunefare and S. De Rosa, "An improved state-space method for coupled fluid-structure interaction analysis," J. Acoust. Soc. Am. **105**, 206–210 (1999).
[3]S. Li, "A state-space coupling method for fluid-structure interaction analysis of plates," J. Acoust. Soc. Am. **118**, 800–805 (2005).
[4]E. J. Grimme, *Krylov projection methods for model reduction*, Ph.D. dissertation, University of Illinois at Urbana-Champaign, IL (1997).
[5]R. W. Freund, "Krylov-subspace methods for reduced-order modeling in circuit simulation," J. Comput. Appl. Math. **123**, 395–421 (2000).
[6]K. H. A. Olsson and A. Ruhe, *Rational Krylov for Model Order Reduction and Eigenvalue Computation*, Technical Report, TRITA-NA-0520 (Royal Institute of Technology (KTH), Stockholm, Sweden, 2005).
[7]A. Ruhe, "Computing nonlinear eigenvalues with spectral transformation Arnoldi," Z. Angew. Math. Mech. **76**, 17–20 (1996).
[8]A. Ruhe, "A rational Krylov algorithm for nonlinear matrix eigenvalue problems," Zap. Nauchn. Semin. POMI **268**, 176–180 (2000).
[9]X. Li, *Power flow prediction in vibrating systems via model reduction*, Ph. D. dissertation, Boston University, MA (2004).
[10]A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, "An advanced computational method for radiation and scattering of acoustic waves in three dimensions," J. Acoust. Soc. Am. **77**, 362–368 (1985).
[11]H. A. Schenck, "Improved Integral Formulation for Acoustic Radiation Problems," J. Acoust. Soc. Am. **44**, 41–58 (1968).
[12]W. Benthien and A. Schenck, "Nonexistence and nonuniqueness problems associated with integral equation methods in acoustics," Comput. Struct. **65**, 295–305 (1997).
[13]F. Fahy, *Sound and Structural Vibration: Radiation, Transmission and Response* (Academic Press, London, 1985).
[14]G. C. Everstine and F. M. Henderson, "Coupled finite element/boundary element approach for fluid-structure interaction," J. Acoust. Soc. Am. **87**, 1938–1947 (1990).
[15]D. Habault and P. J. T. Filippi, "A numerical method for the computation of the resonance frequencies and modes of a fluid-loaded plate: Application to the transient response of the system," J. Sound Vib. **270**, 207–231 (2004).

Active vibration isolation experiments using translational and rotational power transmission as a cost function

Carl Q. Howard[a] and Colin H. Hansen
Active Noise and Vibration Control Group, School of Mechanical Engineering, The University of Adelaide, SA 5005, Australia

(Received 28 February 2005; revised 15 May 2006; accepted 25 June 2006)

Active vibration isolation experiments were conducted using a transducer that measures translational and rotational power transmission from a vibrating mass, through a single-axis active isolator and into a beam. The transducer is capable of measuring forces and moments along six axes and an accelerometer array measures its motion. By combining the measured force and velocity signals the translational and rotational power transmission was measured. Comparisons were made of the effectiveness of several cost functions for minimizing the vibration transmitted into the beam. The results show that active vibration isolation using power transmission as a cost function to be minimized is limited by the phase accuracy of the transducers. The best results were obtained from the minimization of the weighted sum of force and velocity. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2228839]

PACS number(s): 43.40.Vn, 43.40.At [KC] Pages: 2004–2016

I. INTRODUCTION

Vibrating machinery usually generates vibration forces in more than one direction and vibration isolators are often used to reduce the transmission of vibration from the machine into the supporting structure. Typically, vibration isolators are selected to maximize vibration attenuation in the predominant vibrating direction, which is often the translational vertical axis. However, previous research has shown that the vibrational power transmission from rotational moments cannot be neglected when considering the total vibrational power transmitted into the receiving structure.

Here, results of an experimental investigation of the active vibration isolation of a vibrating rigid mass from a simply supported beam are presented. The active vibration isolator used for the investigation has a single control actuator which is orientated vertically. The six-axis vibratory power transducer described in Howard[1] is used to measure the vibratory power transmitted from a vibrating rigid mass, through a vibration isolator, and into the simply supported beam. Several cost functions are compared in terms of their effectiveness at reducing the vibration transmitted into the simply supported beam. The cost functions that are compared are various combinations of squared translational and rotational accelerations, the weighted sum of squared translational force and velocity and rotational moments and velocities, signed translational and rotational power transmission, and squared translational and rotational power transmission. Predictions of the vibration isolation attenuation are made using the theory described in Howard.[1] The predictions are compared with experimental measurements of the active vibration isolation performance using a single-axis active vibration isolator.

The novel work presented in this paper is the active vibration isolation experiments involving the minimization of cost functions that include translational forces and rotational moments, and translational and rotational power transmission. The reason why this work has not been attempted previously is because of the lack of suitable transducers capable of measuring power transmission by moments. The results from experiments presented here provide experimental evidence to support previously published theoretical predictions on the power circulation (or negative power flow) phenomenon that can occur in active vibration isolation implementations when the contribution of rotational power transmission is omitted. In addition, the experimental results demonstrate that the phase errors in the transducers used to measure power transmission limit the usefulness of power transmission as a cost function to be minimized. A better cost function, which is not sensitive to phase inaccuracies, is the weighted sum of the squared translational forces and velocities, and rotational moments and rotational velocities.

II. PREVIOUS WORK

Active vibration isolation experiments have been conducted by several researchers.[2–5] However, in almost all previous work they have neglected the contribution and measurement of power transmission by moments, because suitable transducers were not available. Instead, researchers have attempted to indirectly estimate the vibrational power flowing through the support structure rather than directly measuring the vibrational power flowing into the support structure. Pinnington[6] considered the power transmitted from a machine into a longitudinally stiffened plate, using a multipole expansion technique. Power transmission through four passive isolators was measured using two techniques which did not require the measurement of force at the bottom of the isolator, and was compared with a reference technique, which measured the force and acceleration at the mounting point of each isolator. The first practical method of measuring power transmission through each isolator was to measure

[a]Electronic address: carl.howard@adelaide.edu.au

the source acceleration. The second method was to estimate the magnitude of power generated by all sources of vibration, including airborne noise. It was shown that the two measurement techniques agreed with the reference technique.

Gardonio *et al.*[7,8] theoretically examined the power transmission of a vibrating rigid mass isolated from a plate using two active mounts. They showed that minimization of the out-of-plane component of power, when power transmission due to moments was omitted, caused a "power circulation" phenomena (see also Refs. 9 and 10), where power was drawn into the support plate and then reabsorbed by the active mounts. Power circulation caused greater vibration levels in the plate than without active control. Gardonio's work used two different types of cost function. The first was the out-of-plane power transmission, which was capable of negative values and the second was the weighted sum of the out-of-plane squared velocity and squared force, which is positive definite. A weighting factor was applied to the squared force error signal so that it was the same order of magnitude as the squared velocity signal. In this case, the weighting factor was chosen to be the square of the point mobility of the receiving structure. Gardonio *et al.* reported that the second cost function gave better results than the first. This result is not surprising as the second cost function is always positive and by the definition of power transmission, if the squared velocity or squared force is reduced to zero, then the power transmission along a vertical (out-of-plane) axis is also reduced to zero. The surprising result was that the second cost function gave results close to the minimization of total power transmission, except at a few frequencies where active plus passive isolation was worse than just passive isolation.

Although active vibration isolators have been considered in the past, previous authors have used vibration amplitude squared as the cost function, which does not necessarily relate to the power transmission into the support structure.[11,12] Work which deals with the active vibration isolation of machinery from flexible supports, which uses the power transmitted into the structure as the cost function to be minimized, has also been reported.[2,4] In this work, the power transmission was optimized by manual adjustment of the control forces to minimize the product of force and velocity. However, only the power transmission along a single translational axis was considered, whereas previous research[13–18] has shown the importance of considering power transmission from both translational and rotation axes.

Moorhouse[19] discusses theoretical aspects of the relative importance of force and moment loading on several structural systems such as finite and infinite plates. The methods can be used to identify potential locations for active vibration control sources on a structure.

Ji *et al.*[20] describe a numerical "power mode" approach to estimate power transmission from forces and moments. Their concept is similar to the "radiation mode" approach that is often used in active noise control analyses.

Royston and Singh[21] have considered the active isolation of a vibrating rigid body from a simply supported beam which used a nonlinear spring as a passive vibration isolating element and an "active force input" as a control actuator to cancel the primary disturbance. The active force input was aligned in the vertical axis with the spring and excitation force. Royston and Singh neglected any rotational or horizontal motion because of the difficulty in measuring the rotational dynamics of the system, but noted in the literature review that power transmission by rotational motion was considered important by previous authors.

While there are many theoretical studies that highlight the importance of rotational power transmission in the measurement of total power transmission, few researchers conduct experimental measurements of rotational power transmission.

To measure the power transmission along six axes (three rotational and three translational), a unique "impedance head" is needed to measure the force and acceleration in each of these directions. Although previous authors have considered multiple axis vibration isolators, for use in the aerospace industry[22–25] and machinery vibration isolation,[26–28] there has not been any experimental work reported that uses an active vibration isolator to minimize both translational and rotational vibration. Although Sanderson[29] has measured moment mobilities in structures, there has not been any experimental work in active vibration isolation which minimizes the transmission of rotational moment loads.

Howard *et al.*[9] showed that passive plus active vibration isolation, using vibrational power along a vertical axis as the cost function to be minimized, can increase the vibrational power transmission into the support structure compared with just passive isolation. A similar problem has been examined by Gardonio *et al.*[8] for a plate, but the problem has not been examined for a beam or a cylinder.

This paper presents results from active vibration isolation experiments where translational force and accelerations as well as rotational moments and accelerations are used as cost functions to be minimized. Experimental results showing the active vibration isolation performance are derived from measured transfer function data, which is described in the next section.

III. TRANSFER FUNCTION METHOD TO PREDICT THE VIBRATION ISOLATION USING ACTIVE CONTROL

The method used here to predict the isolation performance of the system uses measured transfer function data. A similar method has been used by Dorling *et al.*[30,31] where measured acoustic transfer function data were used to predict the sound-pressure levels inside an aircraft cabin as a result of active noise control.

Transfer functions were measured between the driving force on the structure and the response at the error sensors. The driving force was measured by placing a force transducer between a primary shaker and the structure. Response measurements were made at the six-axis force transducer and the acceleration transducers. Transfer functions were also measured between the primary shaker and the error sensors and between the control shaker and the error sensors.

The error signals from the error sensors can be written in matrix form as [Ref. 32, Appendix A.5],

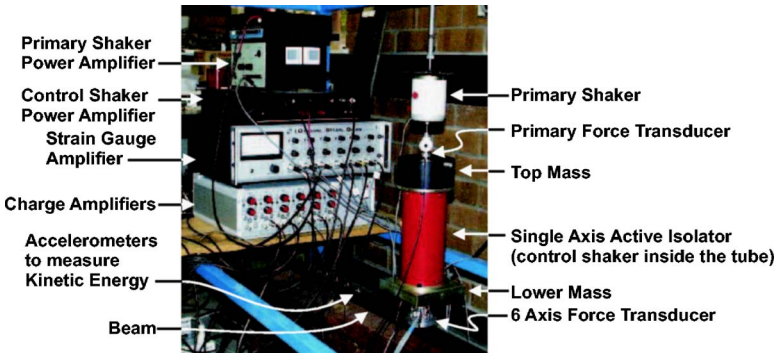


FIG. 1. (Color online) Experimental setup for the single axis isolator on the simply supported beam.

$$\mathbf{e} = \mathbf{d} + \mathbf{C}\mathbf{x}, \tag{1}$$

where $\mathbf{e}$ is an $(n_e \times 1)$ vector of n_e error signals, $\mathbf{x}$ is a $(n_c \times 1)$ vector of control signals, $\mathbf{d}$ is an $(n_e \times 1)$ vector of the error signals resulting from passive control, and $\mathbf{C}$ is a $(n_e \times n_c)$ matrix of the transfer functions between the control signals and the error signals when the primary disturbance is turned off. The usual goal of active control systems is to determine the amplitude and phase of the control signals which will cancel the primary disturbance, and is given by rearrangement of Eq. (1) as

$$\mathbf{x}_0 = -(\mathbf{C})^{-1}\mathbf{d}. \tag{2}$$

Equation (2) can be solved when there are an equal number of control signals and error signals $(n_c = n_e)$. If there are more error signals than control signals $(n_e > n_c)$ then the problem is said to be overdetermined. The matrix $\mathbf{C}$ is not square and cannot be inverted, and generally it is not possible to achieve complete cancellation at all of the error sensors. The problem can be transformed into a least-squares problem such that the cost function J which is minimized is the squared amplitude of the error signals $\mathbf{e}$, which can be written as

$$J = \mathbf{e}^{\mathrm{H}}\mathbf{e} \tag{3}$$

$$= \mathbf{x}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\mathbf{x} + \mathbf{x}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{d} + \mathbf{d}^{\mathrm{H}}\mathbf{C}\mathbf{x} + \mathbf{d}^{\mathrm{H}}\mathbf{d}. \tag{4}$$

Equation (4) is in the general Hermitian quadratic form, and has a minimum value when the control signals are given by

$$\mathbf{x}_0 = -(\mathbf{C}^{\mathrm{H}}\mathbf{C})^{-1}(\mathbf{C}^{\mathrm{H}}\mathbf{d}). \tag{5}$$

When there are more control sources than error sensors $(n_c > n_e)$, the minimization problem becomes underdetermined and there are an infinite number of solutions for the control sources which will minimize the error signals. The problem can be redefined to include a control effort term, such as $\mathbf{x}^{\mathrm{H}}\mathbf{x}$, so that the cost function J is minimized with the least amount of control effort. The cost function J is minimized when the control source is given by

$$\mathbf{x}_0 = -\mathbf{C}^{\mathrm{H}}(\mathbf{C}\mathbf{C}^{\mathrm{H}})^{-1}\mathbf{d}. \tag{6}$$

Consider the system shown in Fig. 1, where the velocity along the vertical axis at the connection between the six-axis force transducer and the beam is to be minimized when the top rigid body is subjected to a harmonic vertical primary force. A transfer function measurement is taken over the frequency range of interest, between the primary driving force and the velocity along the vertical axis at the base of the isolator and this transfer function is called $\mathbf{Z}_{vp}$. The primary driving force is then turned off and a transfer function measurement is taken between the force exerted by the control shaker and the velocity along the vertical axis at the base of the isolator; this transfer function is called $\mathbf{Z}_{vc}$. The terms $\mathbf{d}$ and $\mathbf{C}$ become

$$\mathbf{d} = \mathbf{Z}_{vp}\mathbf{f}_p, \tag{7}$$

$$\mathbf{C} = \mathbf{Z}_{vc}, \tag{8}$$

where $\mathbf{f}_p$ is the $(n_p \times 1)$ column vector of primary forces, which for this example is $\mathbf{f}_p = 1$.

In the experiments that follow, the optimal control forces are calculated by using Eq. (5) and (6) depending on the number of error sensors and control forces. In the experiments where signed power transmission is minimized, the optimal control forces are calculated by a similar method, which is explained later in this section.

Gardonio *et al.* suggested minimizing the weighted sum of squared velocity and squared force along the vertical axis to actively control vibration transmission through an active isolator. They gave the vector of optimal control forces as

$$\mathbf{x}_0 = -(\mathbf{A})^{-1}\mathbf{b}, \tag{9}$$

where

$$\mathbf{A} = \mathbf{Z}_{vc}^{\mathrm{H}}\mathbf{Z}_{vc} + \mu\mathbf{Z}_{fc}^{\mathrm{H}}\mathbf{Z}_{fc}, \tag{10}$$

$$\mathbf{b} = \mathbf{Z}_{vc}^{\mathrm{H}}\mathbf{Z}_{vp}\mathbf{f}_p + \mu\mathbf{Z}_{fc}^{\mathrm{H}}\mathbf{Z}_{vp}\mathbf{Z}_{fp}\mathbf{f}_p, \tag{11}$$

where μ is the weighting factor which is applied to the squared force signal so that the amplitudes of the squared velocity signals and squared force signals are similar, $\mathbf{Z}_{ij}$ is a transfer function between velocity or force, i, and primary or control force, j. For example, $\mathbf{Z}_{vc}$ is the transfer function matrix of dimensions $(n_e \times n_c)$ between the velocity measured at an error sensor and the driving control force.

When there are more error sensors than control forces, Eqs. (9) and (11) presented in Ref. 8 cannot be solved and have to be rewritten in terms of the least-squares problem formulation. The velocities and forces at the error sensors can be written as

$$\mathbf{v} = \mathbf{Z}_{vp}\mathbf{f}_p + \mathbf{Z}_{vc}\mathbf{f}_c, \tag{12}$$

$$\mathbf{f} = \mathbf{Z}_{fp}\mathbf{f}_p + \mathbf{Z}_{fc}\mathbf{f}_c. \tag{13}$$

The terms $\mathbf{d}$ and $\mathbf{C}$ become

$$\mathbf{d} = \begin{bmatrix} \mathbf{Z}_{vp}\mathbf{f}_p \\ \sqrt{\mu}\mathbf{Z}_{fp}\mathbf{f}_p \end{bmatrix} \tag{14}$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{Z}_{vc} \\ \sqrt{\mu}\mathbf{Z}_{fc} \end{bmatrix}. \tag{15}$$

Equation (5) and (6) can now be used to calculate the optimal control forces depending on the number of error sensors and control forces.

The method described here can always be used to calculate the theoretical control force that will minimize the theoretical cost function based on measured transfer function data. However, whether the control force can actually be implemented in practice depends on the primary disturbance and the *causality* of the transfer functions and the causality of Eqs. (5) and (6). If the primary disturbance is tonal (periodic), then the causality issues are not of concern. However, if the primary disturbance is unpredictable (random), then the causality issues are important. The causality issues are further discussed in Ref. 32, (Chap. 8.6).

The calculation of the optimum control forces for the minimization of signed power transmission can be derived in a similar manner as the previous derivation. The velocity and force at the n_e error sensors can be described by vectors $\mathbf{v}_t$ and $\mathbf{f}_t$ which have length n_e. The velocity and force vectors are given by

$$\mathbf{v}_t = \mathbf{Z}_{vp}\mathbf{f}_p + \mathbf{Z}_{vc}\mathbf{f}_c, \tag{16}$$

$$\mathbf{f}_t = \mathbf{Z}_{fp}\mathbf{f}_p + \mathbf{Z}_{fc}\mathbf{f}_c, \tag{17}$$

where $\mathbf{f}_p$ and $\mathbf{f}_c$ are the primary and control force column vectors of length n_p and n_c, respectively, $\mathbf{Z}_{ij}$ is a transfer function between velocity or force, i, and primary or control force, j. For example, $\mathbf{Z}_{fc}$ is the transfer function matrix of dimensions $(n_e \times n_c)$ between the forces measured at the error sensors and the driving control force. These definitions can be used to define the time-averaged harmonic vibrational power transmission into the structure as

$$\text{Power} = \frac{1}{2}\,\text{Re}(\mathbf{v}_t^{\text{H}}\mathbf{f}_t), \tag{18}$$

where the superscript H is the Hermitian transpose. Substitution of Eqs. (16) and (17) into Eq. (18) and rearranging results in a quadratic expression in terms of the control force $\mathbf{q}_c$,

$$\text{Power} = \frac{1}{2}(\mathbf{q}_c^{\text{H}}\boldsymbol{\alpha}\mathbf{q}_c + \mathbf{q}_c^{\text{H}}\boldsymbol{\beta} + \boldsymbol{\beta}^{\text{H}}\mathbf{q}_c + c^i), \tag{19}$$

where

$$\mathbf{q}_c = \begin{bmatrix} \mathbf{f}_c^r \\ \mathbf{f}_c^i \end{bmatrix} \tag{20}$$

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}^T = \frac{1}{2}\begin{bmatrix} \mathbf{a}^i + (\mathbf{a}^i)^T & \mathbf{a}^r - (\mathbf{a}^r)^T \\ -\mathbf{a}^r + (\mathbf{a}^r)^T & \mathbf{a}^i + (\mathbf{a}^i)^T \end{bmatrix} \tag{21}$$

$$\boldsymbol{\beta} = \frac{1}{2}\begin{bmatrix} (\mathbf{b}_2^i)^T + \mathbf{b}_1^i \\ (\mathbf{b}_2^r)^T - \mathbf{b}_1^r \end{bmatrix}, \tag{22}$$

and the real matrices $\mathbf{f}_c^r$, $\mathbf{f}_c^i$, $\mathbf{a}^r$, $\mathbf{a}^i$, $\mathbf{b}_1^r$,… represent, respectively, the real and imaginary parts of the complex matrices $\mathbf{f}_c$, $\mathbf{a}$, $\mathbf{b}_1$, and $\mathbf{b}_2$ and the complex constant c, which are defined as

$$\mathbf{a} = \mathbf{Z}_{vc}^{\text{H}}\mathbf{Z}_{fc}, \tag{23}$$

$$\mathbf{b}_1 = \mathbf{Z}_{vc}^{\text{H}}\mathbf{Z}_{fp}\mathbf{f}_p, \tag{24}$$

$$\mathbf{b}_2 = \mathbf{f}_p^{\text{H}}\mathbf{Z}_{vp}^{\text{H}}\mathbf{Z}_{fc}, \tag{25}$$

$$c = \mathbf{f}_p^{H}\mathbf{Z}_{vp}^{\text{H}}\mathbf{Z}_{fp}\mathbf{f}_p. \tag{26}$$

The power transmission into the system for passive vibration isolation ($\mathbf{q}_c = [0,0]^{\text{T}}$) is given by $c^i/2$. The minimum of Eq. (19) is given by

$$\text{Power}_{\min} = -\frac{1}{2}(\boldsymbol{\beta}^{\text{T}}\boldsymbol{\alpha}^{-1}\boldsymbol{\beta} + c^i), \tag{27}$$

corresponding to an optimum control force vector given by

$$(\mathbf{q}_c)_{\text{opt}} = -\boldsymbol{\alpha}^{-1}\boldsymbol{\beta}. \tag{28}$$

The derivation of the optimum control forces to minimize squared power transmission can be achieved by following the same process as above and has not been included here. Readers are referred to Howard[1] for a complete derivation.

IV. DESCRIPTION OF THE EXPERIMENTAL SETUP

Figure 1 shows a picture of the experimental rig and Fig. 2 shows how the instruments were connected. A steel beam, of dimensions 1.55 m long by 25 mm square, was mounted between two knife edges which provided simply supported end conditions. The six-axis force transducer, described in Howard,[1] was bolted to the beam, 0.75 m from the end of the beam. Attached to the top of the force transducer was the lower mass, which was used to support the end of the vibration isolator. The vibration isolator was a cylindrical polyurethane tube and inside the tube was a Ling Dynamics V203 shaker which provided a canceling force to counteract the vibrations which passed through the outer tube. On top of the vibration isolator was a solid-steel cylindrical mass which weighed 7.4 kg. The control shaker was connected to the rigid mass with a "stinger" that is stiff along the axis of the shaker but is rotationally flexible. Five accelerometers were attached to the beam to measure its residual vibration when active control was applied. The five accelerometers were used to measure the velocity of the beam and were mounted at 0.30, 0.35, 0.40, 0.45, and 0.50 m from its end. For the experimental system considered here, it was found that there were three vibrational axes of greatest importance for power transmission: vertical translation and two rotational axes which did not include the drilling axis through the isolator. Four accelerometers were attached to the six-axis force transducer and used to calculate the acceleration of the material beneath the strain gauges, using the method described in Howard.[1] The use of four accelerometers enables the mea-

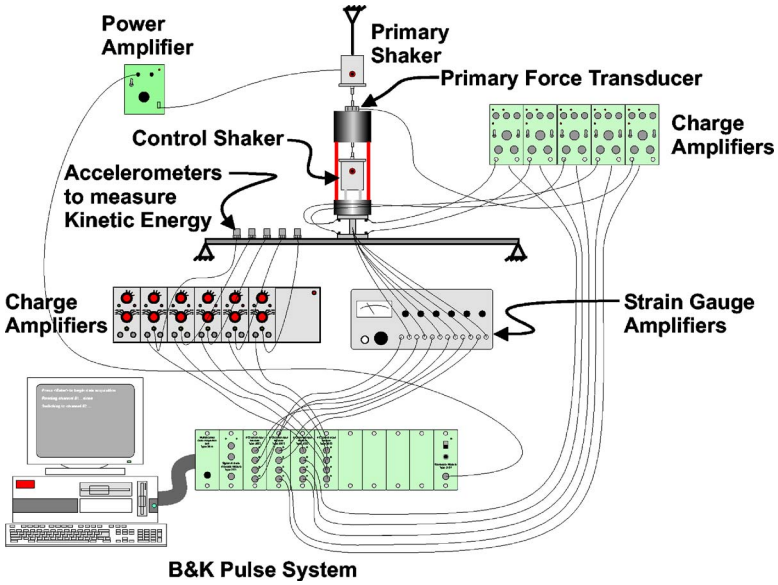


FIG. 2. (Color online) Setup of the instruments for the experiment of the single axis isolator and the simply supported beam.

surement of translational motion and of rotational about one axis. These four accelerometers were repositioned so that motion about a second rotational axis could be measured.

Figure 2 shows how the instrumentation was connected. All the transducers were connected to amplifiers which were connected to the Brüel & Kjær Pulse System, which in turn measured the transfer functions. The primary shaker was connected to the top mass through a B&K force transducer, which was used to measure the primary translational force, and applied a harmonic force which swept in frequency between 5 Hz and 200 Hz.

The vibration isolation performance described here is quantified by the change in the average of the squared velocity of the beam measured using five accelerometers. The average of the squared velocity of the beam is proportional to the kinetic-energy (KE) of the beam. The true value of the KE is calculated by the summation of an infinite number of squared velocity measurements over the length of the beam to measure the translational and rotational accelerations, multiplied by the mass of the beam, and has units of joules. The metric used to describe the relative reduction in beam vibration in the experimental results presented here is proportional to the KE, as the mass term has been neglected, and a finite number of accelerometers was used. This measurement is not affected by phase errors and provides a reasonable approximation of the global KE of the beam. It also provides an independent measure of the isolation performance. Comparisons of the isolation performance using a single sensor, for example the acceleration at the base of the isolator, do not provide a good measure because it is possible to minimize the vibration at the sensor and increase the vibration elsewhere on the supporting structure. Some of the theoretical predictions and experimental results to follow are limited in validity as only five accelerometers were used to measure the average of the squared velocity. This limitation is addressed when it is apparent that it affects the results.

The physical properties of the simply supported beam and isolator system are shown in Table I and are used with the theoretical analysis presented in Howard.[1] The resonance frequencies of the simply supported beam, without the isolator attached, were measured to be 29, 103, and 234 Hz.

V. EXPERIMENTAL RESULTS

In this section, the vibration isolation performance of an active vibration isolation system when various cost functions are minimized is compared with and without active control. It should be noted that no attempt was made to optimize the design of the polyurethane tube that provided passive vibration isolation. It is possible to obtain much higher vibration isolation using a properly designed passive vibration isolator than the results presented here. The isolation performance is measured by monitoring the average of the squared velocity of the beam. The transfer function method, which was described in Sec. III, was used to calculate the cost functions and the average of the squared velocity of the beam.

A. No control

Figure 3 shows that the experimentally measured values of power transmission into the beam do not match the theoretically predicted values. The difference is attributed to the phase errors in the transducers. The phase accuracy of the force transducer and accelerometer combination was measured to be about ±2°.[1] Figure 4 shows that there is a random ±2° phase error of the relative phase angle between force and displacement for theory and experiment. The difference in phase angles between force and displacement is close to 180°, which means that the difference between the force and velocity would be very close to 90°, and hence the small

TABLE I. Parameters used in the active isolator and beam system.

Beam length	1.550 m	Beamwidth	0.025 m
Beam thickness	0.025 m	Isolator location	0.750 m
Young's modulus	207 GPa	Moment of inertia	1.6×10^{-5} m^4
Beam density	7800 kg/m^3	Beam damping	7.48×10^{-6} sN/m
Isolator stiffness k_z	45 870 N/m	Isolator damping c_z	140 sN/m
Isolator stiffness k_{θ_y}	216 N/rad	Isolator damping c_{θ_y}	140 sN/rad
Top mass	7.4 kg	Bottom mass	8.2 kg

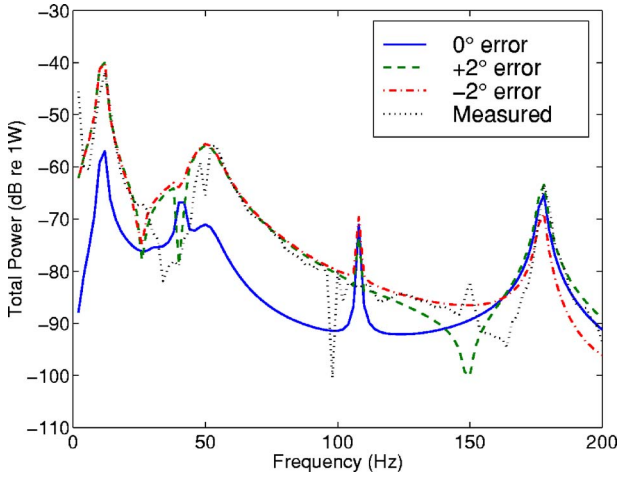


FIG. 3. (Color online) Theoretically predicted and experimentally measured power transmission along the vertical Z axis for a vertical primary force of $F_z=1$ N. The theoretical values of power transmission are shown for 0°, +2°, and −2° phase errors.

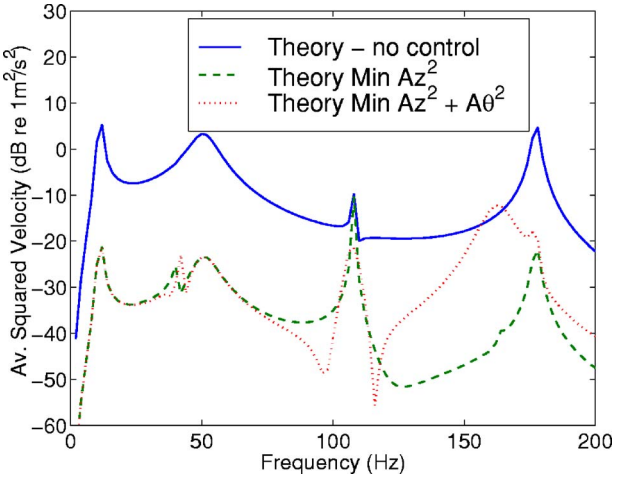


FIG. 5. (Color online) Theoretically predicted average of the squared velocity of the beam for no control, minimization of squared acceleration A_z^2 along the vertical axis and the minimization of the sum of squared accelerations $A_z^2+A_{\theta y}^2$ along the vertical and rotational axes.

errors in the phase measurements have led to the erroneous measurements of power transmission. Theoretical predictions of the total power transmission were made when there was an artificially imposed error in the relative phase between force and velocity of ±2° and −2°, and the results are shown in Fig. 3. The results show that the theoretical predictions with the imposed phase errors appear similar to the experimentally measured results for power transmission.

B. Active control

Theoretical predictions and experimental results are presented for the cases of with and without active isolation of a vibrating rigid mass that is actively isolated from the beam. It is theoretically possible to stop the vibration from the rigid mass from reaching the simply supported beam if the primary force is exactly aligned with the control actuator. In reality, this is difficult to achieve as there is usually a small misalignment between the primary shaker and the centroid of the rigid mass. For the theoretical results presented in this section, it is assumed that there is 2 mm of misalignment, so that the primary load on the top mass is $F_z=1$ N and $M_y=0.002$ Nm.

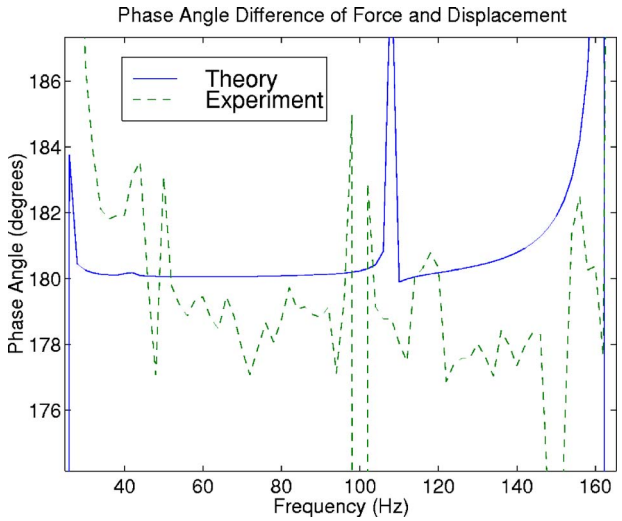


FIG. 4. (Color online) Theoretically predicted and experimentally measured phase angle between force and displacement corresponding to the same configuration shown as shown in Fig. 3.

A reasonable approach to the active vibration isolation of this system is to minimize the squared acceleration along the vertical axis at the base of the isolator. Figure 5 shows the theoretically predicted average of the squared velocity of the beam for no control, minimization of squared acceleration A_z^2 along the vertical axis, and the minimization of the sum of the squared accelerations $A_z^2+A_{\theta y}^2$ along the vertical and rotational axes.

The minimization of the sum of the squared accelerations $A_z^2+A_{\theta y}^2$ along the vertical and rotational axes generally results in smaller reductions in the average of the squared velocity of the beam than the minimization of squared acceleration A_z^2 along the vertical axis. At first glance this result appears to be counterintuitive, as controlling two axes might be expected to produce better results than controlling one axis. However, when minimizing the sum of the squared accelerations $A_z^2+A_{\theta y}^2$ there are two error sensors and one control source so the cost function is overdetermined $(n_e>n_c)$. In this case, it is not possible to calculate a control force such that the amplitude of both error signals will equal zero. Instead, the sum of the squared accelerations $A_z^2+A_{\theta y}^2$ along the translational and rotational axes is minimized by increasing the squared acceleration along the vertical axis compared to when minimizing only the squared acceleration A_z^2 along the vertical axis. The one exception where this resulted in better performance was at 108 Hz, where it can be seen in Fig. 5 the reduction in average of the squared velocity of the beam at the rotational resonance is greater when controlling the sum of the squared accelerations along the vertical and rotational axes than when controlling the squared acceleration along the vertical axis.

In control systems which have more error signals than control sources $(n_e>n_c)$, it is possible to weight the contributions of each error signal to the overall cost function, by multiplying each error signal by a weighting factor, thus providing a mechanism to optimism the results. A large weight-

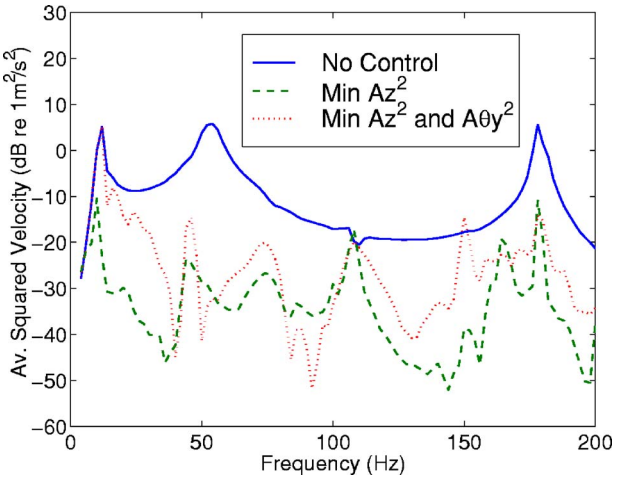


FIG. 6. (Color online) Experimental results for no control, minimization of squared acceleration A_z^2 along the vertical axis and the minimization of the sum of the squared accelerations $A_z^2+A_{\theta y}^2$ along the vertical and rotational axes.

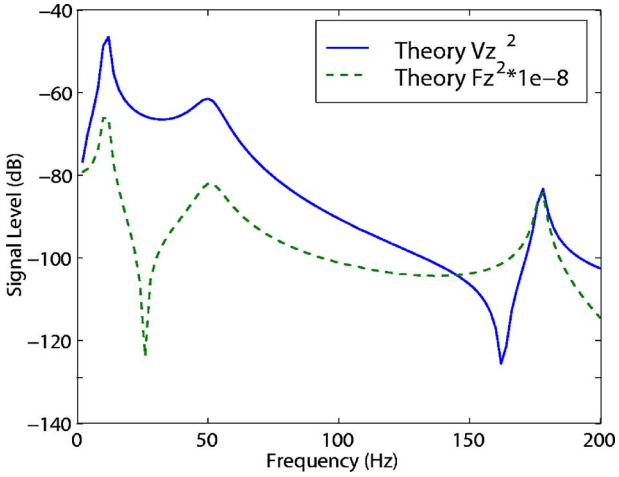


FIG. 8. (Color online) Theoretical signal levels of squared velocity and squared force with the squared force multiplied by a factor of $\mu = 10^{-8}$ s^2/kg^2.

ing factor places greater emphasis on the corresponding error signal in the cost function. Cost functions which use a weighted sum of the error signals are further discussed later in this section.

An experiment was conducted to verify the theoretical predictions shown in Fig. 5 and the results are shown in Fig. 6.

These experimental results confirm the two theoretical predictions that: (1) in general the reduction in the average of the squared velocity of the beam for minimization of the sum of the squared accelerations along the vertical and rotational axes is less than that obtained by minimizing the squared acceleration along the vertical axis; and (2) the reduction in average of the squared velocity of the beam at the rotational resonance is greater when the sum of the squared accelerations along the vertical and rotational axes is minimized than when only the squared acceleration along the vertical axis is minimized.

Figures 5 and 7 show the limitation of using only five accelerometers to measure the average of the squared velocity of the beam. Figure 5 shows that at 95 and 115 Hz the average of the squared velocity for the minimization of the sum of the squared accelerations along the vertical and rotational axes is lower than that obtained for the minimization of squared acceleration along the vertical axis. Figure 7 shows the same theoretical predictions as Fig. 5, but this time ten accelerometers mounted along the beam and four on the force transducer were used to calculate the average of the squared velocity of the beam. The theoretical results in Fig. 7 show that by using 14 accelerometers to measure the average of the squared velocity of the beam, the isolation performance at 95 and 115 HZ when minimizing the sum of the squared accelerations $A_z^2+A_{\theta y}^2$ along the vertical and rotational axes is similar to minimizing the squared acceleration A_z^2 along the vertical axis. The power transmission spectrum can be related to the KE spectrum by a frequency-dependent function as shown by Pavić.[33]

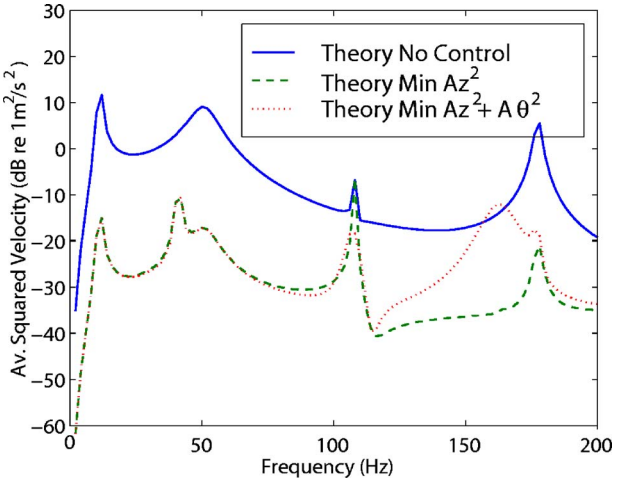


FIG. 7. (Color online) Theoretically predicted average of the squared velocity of the beam (using 14 accelerometers) for no control, minimization of squared acceleration A_z^2 along the vertical axis and the minimization of the sum of the squared accelerations $A_z^2+A_{\theta y}^2$ along the vertical and rotational axes.

1. Weighted sum of force and velocity

It has been suggested by Gardonio *et al.*[8] that the minimization of a weighted sum of the squared velocity and squared force along the vertical axis will have a similar result to the minimization of total power transmission. The purpose of using a weighted sum of squared velocity and squared force is to adjust the signal levels to be a similar order of magnitude. Figure 8 shows that when the theoretically predicted value of squared force is reduced in amplitude by multiplying by 10^{-8} s^2/kg^2 it has a similar signal level to the theoretically predicted squared velocity. Figure 9 shows the corresponding experimentally measured squared velocity signal and the experimentally measured squared force signal multiplied by $\mu = 10^{-8}$ s^2/kg^2.

It can be seen in Figs. 8 and 9 that the squared velocity signal level is greater than the weighted squared force signal, except in the frequency range between about 150 and 170 Hz. It is then reasonable to expect that the predicted theoretical and experimental results of the minimization of the weighted sum of squared velocity and squared force should follow the response of the squared velocity except in

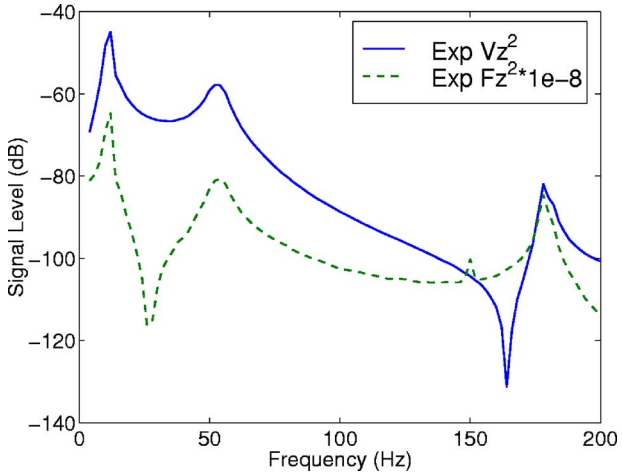


FIG. 9. (Color online) Experimental results of scaling the squared force signal to be a similar order of magnitude as the squared velocity signal.

the frequency range between 150 and 170 Hz, where is should follow the results for the minimization of the squared force.

Figure 10 shows the theoretically predicted average of the squared velocity of the beam for no control, minimization of squared velocity V_z^2, minimization of squared force F_z^2, and the minimization of the weighted sum of squared velocity and squared force $V_z^2+\mu F_z^2$, where $\mu=10^{-8}$ s^2/kg^2.

An experiment was conducted to verify the results from Fig. 10 and the results are shown in Fig. 11. These results show that there is some improvement in results when the minimization of the weighted sum of squared velocity and squared force is used rather than the minimization of squared acceleration.

Gardonio *et al.*[8] suggested minimizing the weighted sum of squared velocity and squared force along the vertical axis. Another possibility is to minimize the weighted sum of the squared velocities along translational and rotational axes, squared forces and squared moments. Figure 12 shows the experimentally measured average of the squared velocity of the beam for no control, the minimization of squared velocity along the vertical axis, and the minimization of the weighted sum of V_t^2 (sum of the squared velocities along the vertical axis V_z^2 and around the rotational axis $V_{\theta y}^2$) and F_t^2 (the sum of the squared force along the vertical axis F_z^2 and the squared moment M_y^2). The minimization of $V_t^2+\mu F_t^2$, the weighted sum of squared velocities, squared forces, and squared moments along translational and rotational axes, results in slightly better vibration isolation performance than $V_z^2+\mu F_z^2$, the weighted sum of the squared velocity and squared force along the vertical axis.

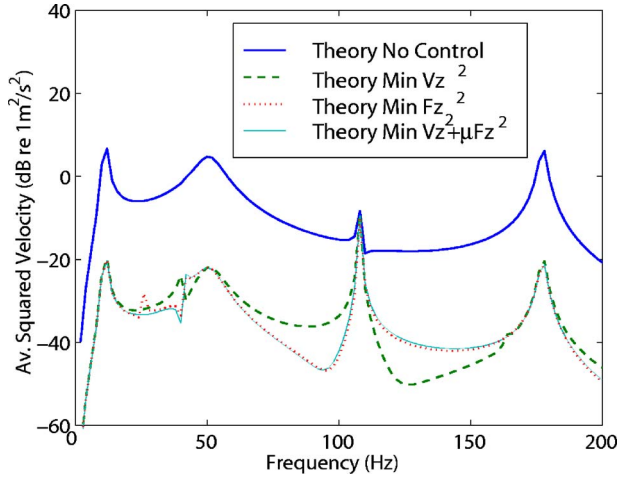


FIG. 10. (Color online) Theoretically predicted average of the squared velocity of the beam for no control, minimization of squared velocity V_z^2 along the vertical axis, minimization of squared force along the vertical axis F_z^2 and the minimization of the weighted sum of $V_z^2+\mu F_z^2$ the squared velocity and squared force where $\mu=10^{-8}$ s^2/kg^2.

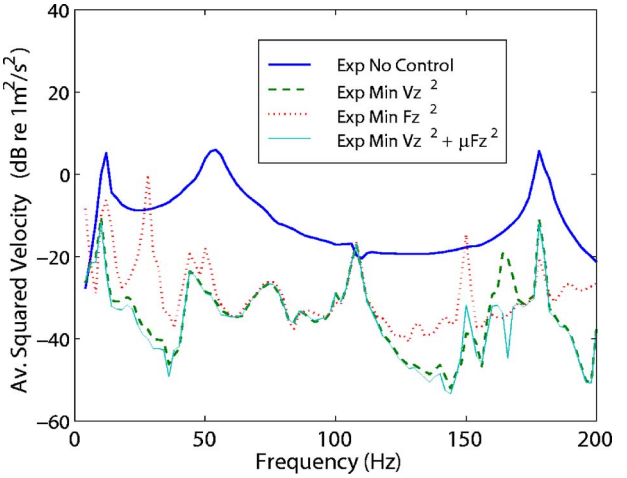


FIG. 11. (Color online) Experimental results of the average of the squared velocity of the beam for no control, minimization of squared velocity along the vertical axis V_z^2, minimization of squared force F_z^2 along the vertical axis and the minimization of the weighted sum of $V_z^2+\mu F_z^2$ the squared velocity and squared force.

2. Signed power transmission

The results from Howard *et al.*[9] show that active control using signed power transmission as a cost function to be minimized will converge to a negative value if moments are present and could result in the overall vibration response of the receiving structure being greater than it was with only passive isolation. Signed power transmission is a measure of

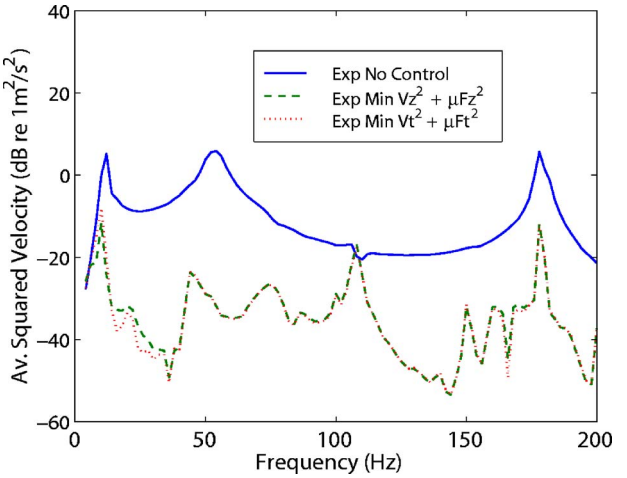


FIG. 12. (Color online) Experimental results of the average of the squared velocity of the beam for no control, minimization of the weighted sum of $V_z^2+\mu F_z^2$ the squared velocity and squared force along the vertical axis and the minimization of the weighted sum of $V_t^2+\mu F_t^2$, squared velocities, squared forces and squared moments along translational and rotational axes.

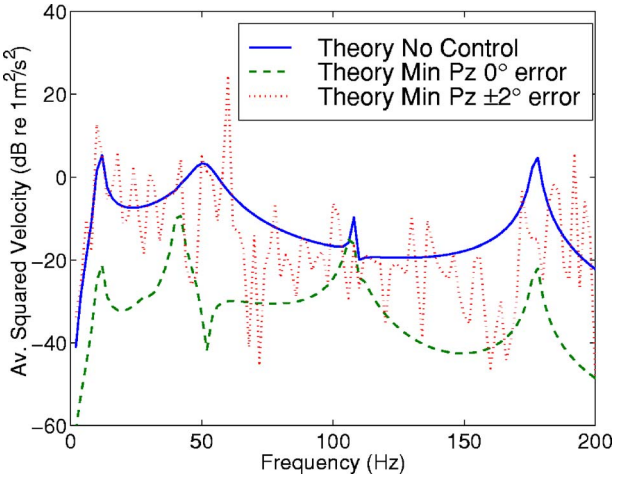


FIG. 13. (Color online) Theoretically predicted average of the squared velocity of the beam for no control, minimization of signed power transmission P_z along the vertical axis and when there is ±2° phase error.

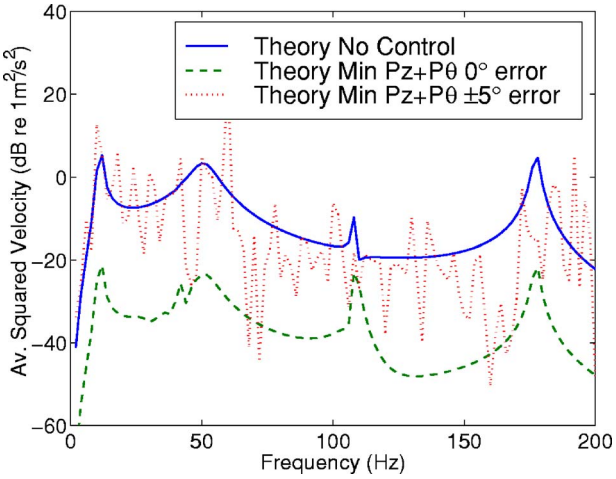


FIG. 15. (Color online) Theoretically predicted average of the squared velocity of the beam for no control, minimization of the sum of $P_z+P_{\theta y}$, the signed power transmission along the vertical axes and when there is ±2° phase error.

power transmission that takes into account its direction: positive from the vibration source to the support structure and negative for the opposite direction. Minimizing signed power transmission results in the most negative value being optimum. It has been of concern to researchers that small phase errors in the measurement of power transmission can corrupt its true measure such that attempts to reduce vibration transmission using active vibration control, with signed power transmission as a cost function, will be unsatisfactory.

Figure 13 shows a theoretical prediction of the approximated average of the squared velocity of the beam for no control and minimization of signed power transmission and the minimization of signed power transmission when there is a random ±2° phase error. A phase error between ±2° was applied to the transfer function measurement between the force response of the structure and the primary load $\mathbf{Z}_{fp}$ and another phase error between ±2° was applied to the transfer function measurement between the force response of the structure and the force applied by the control actuator $\mathbf{Z}_{fc}$. The use of two different values of phase error for the primary and control actuator responses simulates a random phase error that varies with time.

The theoretical result in Fig. 13 shows that the minimization of signed power transmission along a vertical axis with a small phase error will produce unsatisfactory results. This was confirmed in an experiment as shown in Fig. 14.

Similar results occur when the signed total power transmission is minimized. Figure 15 shows the theoretical approximate average of the squared velocity of the beam when the signed power transmission along both the vertical axis and the rotational axis are minimized for an accurate measurement of power and when there is a ±2° phase error in the measurement of force. Figure 16 shows the corresponding experimental result.

The results presented in Figs. 13–16 verify that attempts to minimize signed power transmission along either a vertical axis or along the sum of the vertical and rotational axes will be limited by the phase accuracy of the transducers. This result agrees with the comments by Henriksen[34] and Gardonio *et al.*[8]

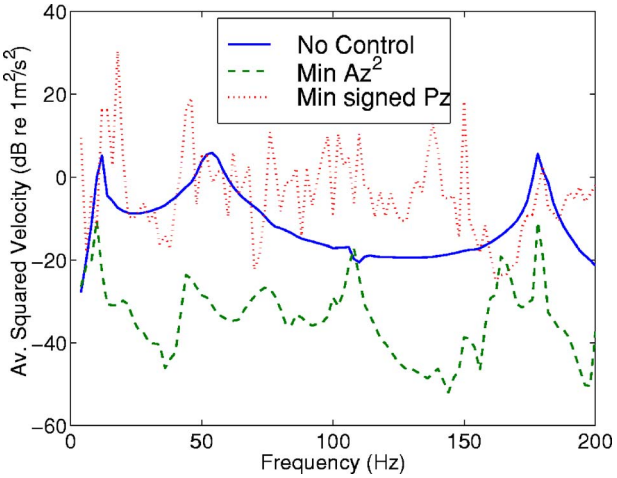


FIG. 14. (Color online) Experimental results of the average of the squared velocity of the beam for no control, minimization of A_z^2 and minimization of the signed power transmission P_z.

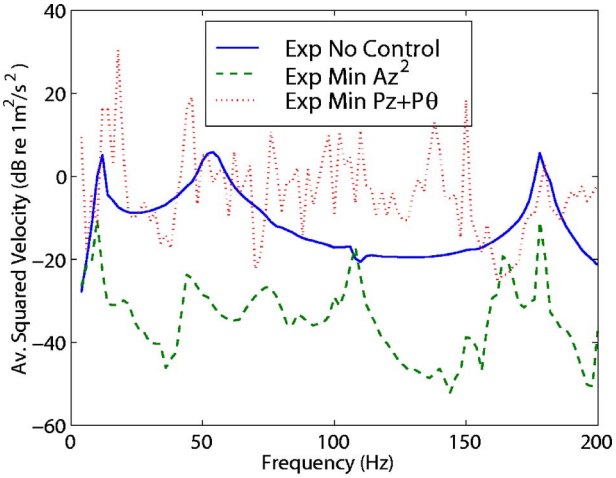


FIG. 16. (Color online) Experimental results of the average of the squared velocity of the beam for no control, minimization of A_z^2 and minimization of the sum of the signed power transmission P_z and $P_{\theta y}$.

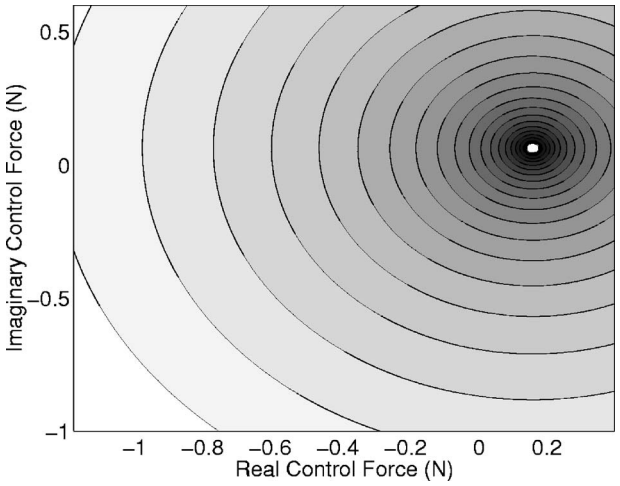


FIG. 17. Contour plot of the theoretical squared power transmission $P_z^2 + P_{\theta y}^2$ along the vertical and rotational axes with no phase error. The white dot shows the control force which minimizes the squared total power transmission. Power transmission is inversely proportional to the darkness of the contour.

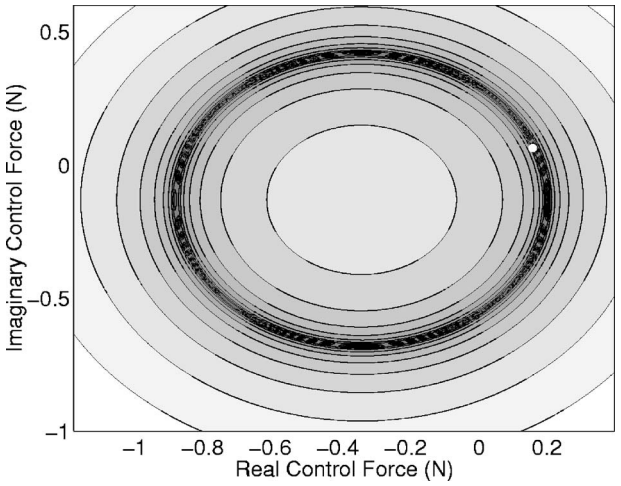


FIG. 18. Contour plot of the theoretical squared power transmission $P_z^2 + P_{\theta y}^2$ along the vertical and rotational axes with a ±2° phase error. The white dot shows the control force which minimizes the squared total power transmission with no phase error. Power transmission is inversely proportional to the darkness of the contour. Minimum power transmission is represented by the black ring.

3. Squared power transmission

The results from Howard[9] show that the minimization of squared power transmission gives results better than the minimization of signed power transmission when negative values of signed power transmission are possible, even though random phase errors also cause an error in the measurement of squared power transmission. The theoretical model can be used to show the effect of phase errors on the isolation performance for the minimization of squared power transmission. Figure 17 shows a contour plot of the squared total power transmission at 100 Hz for the theoretical model when transducers have no phase errors. The shading indicates constant levels of squared power transmission, and darker shading indicates values that are closer to zero. The axes are the real and imaginary parts of the control force, and the white dot at the center of the rings corresponds to the value of the control force which minimizes the squared total power transmission. This result shows that if the transducers had no phase errors, then the error surface would resemble a parabolic bowl. If the transducers have phase errors then the error surface will not have a unique global minimum but will have an infinite number of solutions for the control force which will minimize the erroneous measure of squared power transmission, as shown in Fig. 18 as the dark ring. Figure 18 shows a white dot which is at the same location as the white dot in Fig. 17. This is the control force which an adaptive controller should converge towards. The error surface shown in Fig. 18 resembles a parabolic bowl with an inverted bowl at the center of the parabola. Figure 19 shows a close-up of Fig. 18 around the control force which minimizes the total power transmission with no phase error. Figure 19 shows that the control force which minimizes the true value of squared total power transmission does not lie on the ring of solutions which minimizes the erroneous measure of squared total power transmission. Obviously an adaptive controller should converge towards the true value, but the controller could converge to any solution on the dark ring shown in Fig. 18. The controller needs to be guided towards the true value. Figure 19 also shows the control force which minimizes the squared acceleration along the vertical Z axis, which has a value near to the control force which minimizes the true value of squared total power transmission. An adaptive controller could be guided towards minimizing the squared acceleration, which will start the adaptation process in the correct direction towards minimizing the true value of squared power transmission. Once the controller had minimized the cost function of squared acceleration, the cost function was altered so that it minimized the squared power transmission. This technique was used here and the control force which was calculated is shown in Fig. 19 as a white dot which lies on the ring of solutions where the squared total power transmission (with phase errors) equals zero. These same three solutions for the control force are shown in Fig.

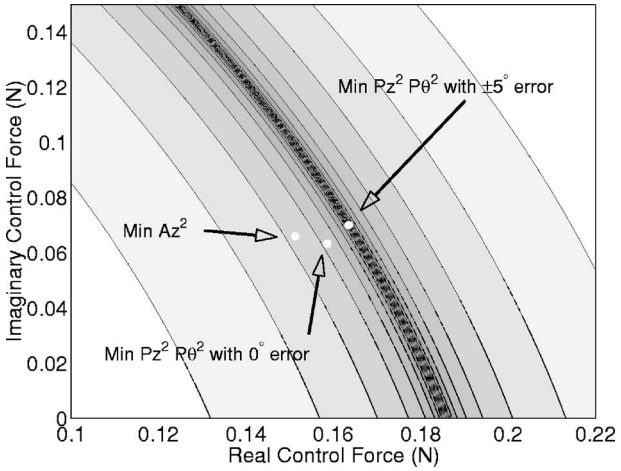


FIG. 19. Contour plot of the theoretical squared power transmission $P_z^2 + P_{\theta y}^2$ along the vertical and rotational axes with a ±2° phase error showing the three different control forces which minimize the squared acceleration along the vertical axis, the squared total power transmission with no phase error, and the squared total power transmission with ±2° phase error. Power transmission is inversely proportional to the darkness of the contour. Minimum power transmission is represented by the black ring.

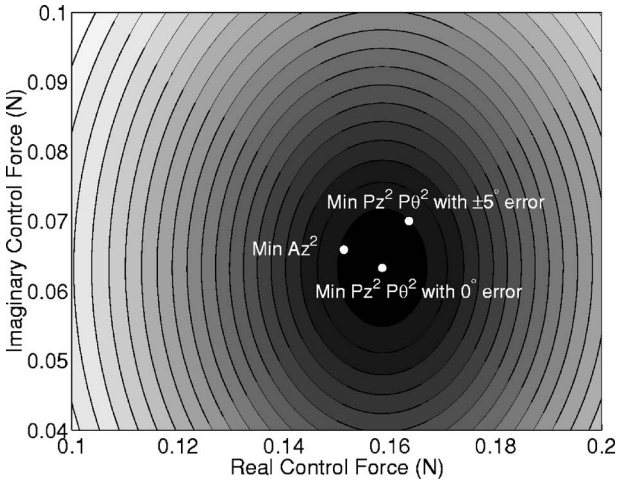


FIG. 20. Contour plot of the theoretical squared power transmission $P_z^2 + P_{\theta y}^2$ along the vertical and rotational axes with no phase error, showing the three different control forces which minimize, respectively, the squared acceleration along the vertical axis, the squared total power transmission with no phase error, and the squared total power transmission with ±2° phase error. Power transmission is inversely proportional to the darkness of the contour.

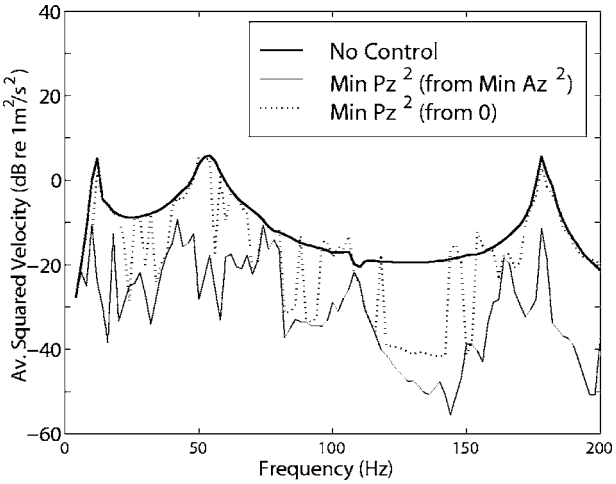


FIG. 21. Experimentally measured average of the squared velocity of the beam when the adaptive controller starts to minimize the squared power transmission along the vertical axis from zero control force and when the adaptation starts from the minimization of squared acceleration A_z^2.

20, where the contours show the true value of total power transmission, that is with no phase error. The control force which is closest to the control force which minimizes the true value of the squared total power transmission is the better solution. In this case the control force which minimizes the squared acceleration along the vertical axis and the control force which minimizes the squared total power transmission with phase errors have about the same value of total power transmission as they are both on the same contour level.

It is not possible to experimentally demonstrate this phenomenon as the force transducers and accelerometers used in the experiments have phase errors and cannot be compared with an experiment without phase errors. It is possible to experimentally demonstrate the technique described above where the adaptation is guided towards the minimization of squared acceleration. Figure 21 shows the experimental results for the average of the squared velocity of the simply supported beam when the adaptive controller starts to minimize the squared power transmission along the vertical Z axis from zero control force and when the controller starts from a control force which minimizes the squared acceleration along the vertical Z axis. This result confirms that the controller must be guided towards minimizing the true value of total power transmission (with no phase error).

The results which follow, in which squared power transmission has been minimized, were obtained using this technique to initially guide the solution towards minimizing the squared acceleration.

Figure 22 shows the theoretically predicted average of the squared velocity of the beam for no control, when the squared power transmission P_z^2 along the vertical axis is minimized, and when the sum of the squared power transmission $P_z^2 + P_{\theta y}^2$ along the vertical and rotational axes is minimized when there is a random ±2° phase error. This result shows that phase errors associated with the measurement of power will not greatly affect the minimization of squared power transmission. This prediction was confirmed by experiment as shown in Fig. 23. It can be seen that the minimization of squared power transmission along the vertical and rotational axes results in a lower average of the squared velocity of the beam at the rotational resonance of 108 Hz.

Another experiment was conducted for the case where the rigid mass was excited along both the vertical axis and the horizontal axis aligned with the beam. The results from these experiments are not presented as they are similar to those described above, except that the peak corresponding to the rotational resonance at 108 Hz is larger for both the controlled and uncontrolled cases.

Figure 24 shows that the cost functions considered so far provide similar levels of vibration isolation. However, the greatest vibration isolation obtained by the minimization of

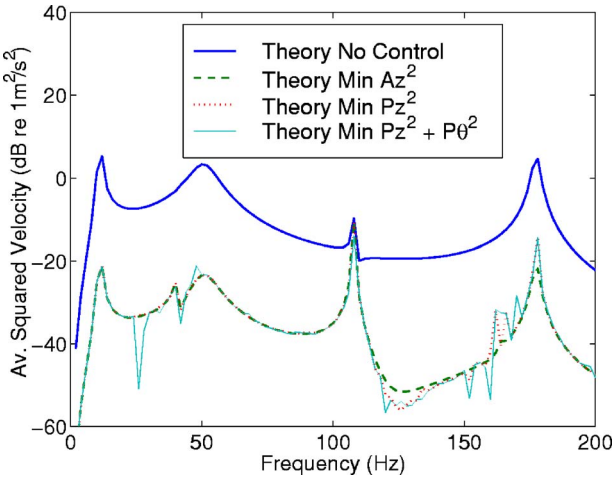


FIG. 22. (Color online) Theoretical prediction of the average of the squared velocity of the beam for no control, minimization of squared acceleration A_z^2 along the vertical axis, minimization of squared power transmission P_z^2 along the vertical axis with a random ±2° phase error and the minimization of the sum of the squared power transmissions $P_z^2 + P_{\theta y}^2$ along the vertical and rotational axes with a random ±2° phase error.

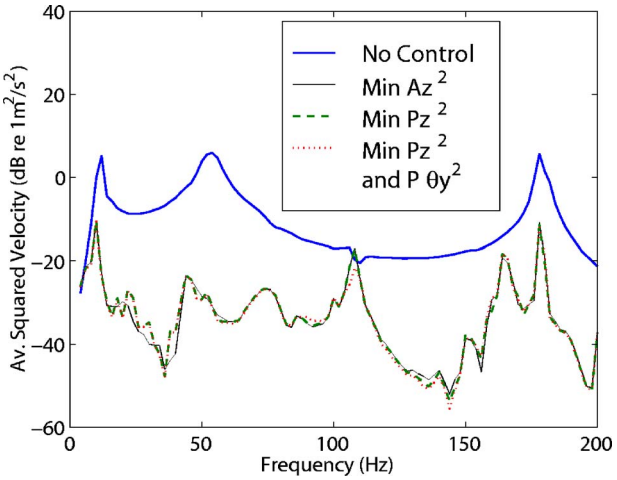


FIG. 23. (Color online) Experimental results of the average of the squared velocity of the beam for no control, minimization of squared acceleration A_z^2 along the vertical axis, minimization of squared power transmission P_z^2 along the vertical axis and the minimization of the sum of the squared power transmissions $P_z^2+P_{\theta y}^2$ along the vertical and rotational axes.

the weighted sum of squared velocity and force along translational and rotational axes was slightly better than that obtained using the other cost functions.

VI. CONCLUSIONS

A novel transducer was used to investigate the effectiveness of various cost functions for actively minimizing the transmission of vibration from a vibrating rigid mass to a simply supported beam. The active isolator was intended to control vibration transmission only along the vertical axis, and the transducer was used as an error sensor allowing minimization of vibration along any translational or rotational axis or combination of axes. The effectiveness of each

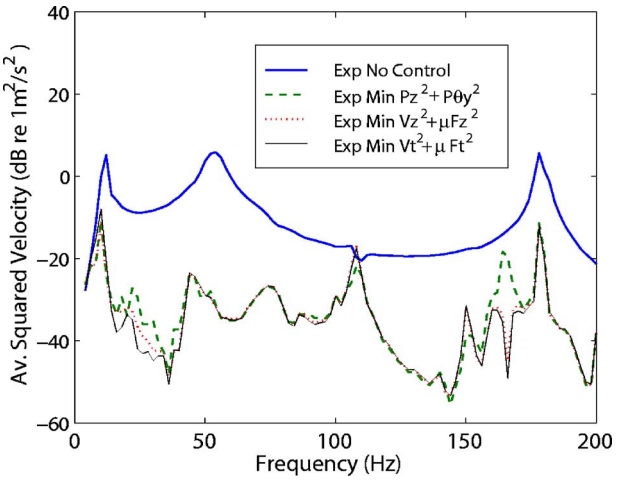


FIG. 24. (Color online) Experimental results of the average of the squared velocity of the beam for no control, minimization of squared acceleration A_z^2 along the vertical axis, minimization of the sum of the squared power transmissions $P_z^2+P_{\theta y}^2$ along the vertical and rotational axes, minimization of the weighted sum of $V_z^2+\mu F_z^2$, the squared velocity and force along the vertical axis and the minimization of $V_t^2+\mu F_z^2$, the weighted sum of squared velocities, forces, and moments along translational and rotational axes.

cost function was evaluated by measuring the vibration levels in the simply supported beam which acted as the receiving structure.

The experimental results showed that the minimization of the signed value of power transmission was ineffective in minimizing the vibration transmitted into the simply supported beam, which was due to insufficient phase accuracy of the transducer used to derive the signed power transmission for the error signal for the controller. Theoretical predictions were made which included random phase errors for the error transducer, and these predicted results were comparable to the measured experimental results. The results obtained by minimizing the squared value of power transmission were an improvement over those obtained by minimizing the signed value of power transmission. However, the phase accuracy of the transducer still limited the maximum vibration attenuation that could be achieved. Although the measurement of total vibrational power transmission is theoretically appealing because the vibrational energy from the contribution of translational and rotational vibration uses consistent units of watts, this metric will unfortunately always be limited by the phase accuracies of the transducers used to determine power transmission.

The best vibration isolation performance was obtained from the minimization of the weighted sum of the squared translational forces and velocities and squared rotational moments and velocities. This cost function is not limited by the phase accuracies of the transducers and hence it is more practical than the measurement of vibrational power transmission. Similar results have been shown for the acoustic equivalent where energy density sensing has been shown to be more practical than the minimization of sound intensity.[35] The combined force, moment, and velocity signals had to be weighted appropriately so that the amplitude of each signal was similar, so as not to favor the attenuation of one vibration along or around one axis over that corresponding to another axis. The appropriate weighting factor is a function of the structural impedance measured at the error transducer.

It can be justifiably claimed that the adjustment of the signal amplitudes constitutes the creation of an artificially optimum cost function, whereas the cost function of the total vibrational power transmission is independent of the system configuration under investigation. Although there is elegance in the academic power transmission approach, the use of squared force and velocity signals is a realizable and much more practical solution.

[1]C. Q. Howard, "Active isolation of machinery vibration from flexible structures," Ph.D. thesis, University of Adelaide, Australia, 1999.
[2]J. Q. Pan, C. H. Hansen, and J. Pan, "Active isolation of a vibration source from a thin beam using a single active mount," in *Proceedings of Inter-Noise '91* (INCE, Sydney, Australia, 1991), pp. 683–686.
[3]J. Q. Pan, C. H. Hansen, and J. Pan, "Active isolation of a vibration source from a thin beam using a single active mount," J. Acoust. Soc. Am. **94**(3), 1425–1434 (1993).
[4]J. Q. Pan and C. H. Hansen, "Active control of vibratory power flow from a vibrating rigid body to a flexible panel through two active isolators," J. Acoust. Soc. Am. **93**, 1947–1953 (1993).
[5]T. Royston and R. Singh, "Optimization of passive and active non-linear vibration mounting systems based on vibratory power transmission," J. Sound Vib. **194**(3), 295–316 (1996).
[6]R. J. Pinnington, "The measurement of vibrational power input to a struc-

ture by a multipole expansion," in *Structural Intensity and Vibrational Energy Flow—3rd International Congress on Intensity Techniques* (CETIM, Senlis, France, 1990), pp. 441–448.
[7]P. Gardonio, S. J. Elliot, and R. J. Pinnington, "Active isolation of structural vibration on a multiple degree of freedom system. I. The dynamics of the system," J. Sound Vib. **207**(1), 61–93 (1997).
[8]P. Gardonio, S. J. Elliot, and R. J. Pinnington, "Active isolation of structural vibration on a multiple degree of freedom system. II. Effectiveness of active control strategies," J. Sound Vib. **207**(1), 95–121 (1997).
[9]C. Q. Howard and C. H. Hansen, "Finite element analysis of active vibration isolation using vibrational power as a cost function," Int. J. Acoust. Vib. **4**(1), 23–36 (1999).
[10]B. Gibbs and S. Yap,"The contribution of forces and moments to the structure-borne sound power from machines rigidly attached to supporting floors," in *Proceedings of Inter-Noise '98* (INCE, Christchurch, New Zealand, 1998).
[11]M. D. Jenkins, "Active control of periodic machinery vibrations," Ph.D. thesis, University of Southampton, 1989.
[12]S. D. Sommerfeldt, "Multi-channel adaptive control of structural vibration," Noise Control Eng. J. **37**(2), 77–89 (1991).
[13]B. A. T. Petersson and B. M. Gibbs, "The influence of source location with respect to vibrational energy transmission," in *3rd International Congress on Intensity Techniques* (CETIM, Senlis, France, 1990), pp. 449–456.
[14]B. A. T. Petersson, "Moment and force excitation at edges and corners of beam and plate like structures," in *International Conference on Recent Advances in Structural Dynamics* (ISVR, Southampton, 1991), pp. 148–157.
[15]Y. K. Koh and R. G. White, "Analysis and control of vibrational power transmission to machinery supporting structures subjected to a multi-excitation system. I. Driving point mobility matrix of beams and rectangular plates," J. Sound Vib. **196**(4), 469–493 (1997).
[16]Y. K. Koh and R. G. White, "Analysis and control of vibrational power transmission to machinery supporting structures subjected to a multi-excitation system. II. Vibrational power analysis and control schemes," J. Sound Vib. **196**(4), 495–508 (1997).
[17]Y. K. Koh and R. G. White, "Analysis and control of vibrational power transmission to machinery supporting structures subjected to a multi-excitation system. III. Vibrational power cancellation and control experiments," J. Sound Vib. **196**(4), 509–522 (1997).
[18]M. Sanderson, "Vibration isolation: Moments and rotations included," J. Sound Vib. **198**, 171–191 (1996).
[19]A. Moorhouse, "A dimensionless mobility formulation for evaluation of force and moment excitation," J. Acoust. Soc. Am. **112**(3), 972–980 (2002).
[20]L. Ji, B. Mace, and R. Pinnington, "A power mode approach to estimating vibrational power transmitted by multiple moments," J. Sound Vib. **265**, 387–399 (2003).
[21]T. J. Royston and R. Singh, "Optimization of passive and active non-linear vibration mounting systems based on vibratory power transmission," J. Sound Vib. **194**(3), 295–316 (1996).
[22]K. M. Misovec, F. F. Flynn, B. G. Johnson, and J. K. Hedrick, "Sliding mode control of magnetic suspensions for precision pointing and tracking applications," in *American Society of Mechanical Engineers, Winter Annual Meeting - Active Noise and Vibration Control* (ASME, Dallas, TX, 1990), Vol. NCA - **8**, pp. 75–82.
[23]R. C. Fenn, J. R. Downer, V. Gondhalekar, and B. G. Johnson, "An active magnetic suspension for space-based microgravity vibration isolation," in *American Society of Mechanical Engineers, Winter Annual Meeting - Active Noise and Vibration Control* (ASME, Dallas, TX, 1990), Vol. NCA - **8**, pp. 49–56.
[24]C. H. Gerhold and R. Rocha, "Active vibration control in microgravity environment," Trans. ASME, J. Vib., Acoust., Stress, Reliab. Des. **110**, 30–35 (1988).
[25]D. A. Kienholz, "Defying gravity with active test article suspension systems," Sound Vib., **14**, 14–21 (1994).
[26]C. Ross, J. Scott, and S. Sutcliffe, Active Control of Vibration, International Patent Application No. PCT/GB87/00902, 1988.
[27]C. Ross, J. Scott, and S. Sutcliffe, Active Control of Vibration, US Patent No. 5,433,422, 1989.
[28]S. Sutcliffe, G. Eatwell, and S. Hutchins, Active Control of Vibration, UK Patent No. GB 2,222,657, 1989.
[29]M. A. Sanderson, "Direct measurement of moment mobility. II. An experimental study," J. Sound Vib. **179**(4), 685–696 (1995).
[30]C. Dorling, G. Eatwell, S. Hutchins, C. Ross, and S. Sutcliff, "A demonstration of active noise reduction in an aircraft," J. Sound Vib. **112**(2), 389–395 (1987).
[31]C. Dorling, B. Eatwell, S. Hutchins, C. Ross, and S. Sutcliffe, "A demonstration of active noise reduction in an aircraft cabin," J. Sound Vib. **128**(2), 358–360 (1989).
[32]P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, San Diego, 1992).
[33]G. Pavić, "On the relationship between the energy flow into a structure and its vibration," in *Structural Intensity and Vibrational Energy Flow - 4th International Congress on Intensity Techniques* (CETIM, Senlis, France, 1993), pp. 95–99.
[34]E. Henriksen, "Adaptive active control of structural vibration by minimisation of total supplied power," in *Inter-Noise 96* (INCE, Liverpool, England, 1996), pp. 1615–1618.
[35]Y. Park and S. Sommerfeldt, "Global attenuation of broadband noise fields using energy density control," J. Acoust. Soc. Am. **101**(1), 350–359 (1997).

Broadband noise reduction of piezoelectric smart panel featuring negative-capacitive-converter shunt circuit

Jaehwan Kim[a] and Young-Chae Jung[b]
Department of Mechanical Engineering, Inha University, 253 Younghyun-Dong, Namku, Incheon 402-751, Korea

(Received 25 October 2005; revised 29 March 2006; accepted 8 July 2006)

A broadband noise reduction of a piezoelectric smart panel featuring a negative capacitance converter (NCC) shunt circuit is experimentally investigated. Piezoelectric shunt damping utilized on the panel structure is attractive for noise reduction especially at low resonance frequencies of the structure. To achieve a broadband noise reduction, however, a multimode shunt is necessary. The NCC circuit can be an ideal broadband shunt circuit by nullifying the capacitance of the piezoelectric patch with the circuit. Since the intrinsic capacitance of the patch is not constant with the frequency, the broadband shunt performance of the NCC can be deteriorated. Thus, we introduce the dual-patch NCC circuit on the smart panel. The proposed concept is explained and the tuning and implementation procedures are addressed. The noise reduction performance of the panel is tested in terms of transmission loss according to the standard transmitted noise measurement. The broadband damping performance of the smart panel featuring a dual-patch NCC shunt is compared with the panels featuring resonant shunt circuit and ordinary NCC shunt circuit in terms of acceleration and noise transmission loss. It is found that the dual-patch NCC shunt is more efficient than ordinary NCC and resonant shunt for achieving broadband noise reduction with smart panels. [DOI: 10.1121/1.2259791]

PACS number(s): 43.50.Gf, 43.50.Ki, 43.55.Wk, 43.55.Rg [KAC] Pages: 2017–2025

I. INTRODUCTION

Noise reduction from vibrating structures is a crucial issue in various fields such as automobiles, airplanes, ships, and buildings. There have been many attempts to reduce or control the noise radiating from structures. Designs of material properties or thickness of the structures have reduced the radiated noise.[1,2] The use of sound absorbing layers on the structures is a conventional method in passive noise control.[3] These passive methods allow a simple configuration, are easy to install, low cost, and robust in harsh environments. However, when the structure or acoustic cavity coincides with resonance frequencies, especially low modes, these methods are not efficient. Furthermore, when the structural characteristics are changed, they are not effective in adapting to change. In contrast, active methods use sensors and actuators along with a proper control so as to minimize the noise from the structures at a certain frequency band. Successful noise reductions have been obtained by using piezoelectric sensors and actuators such that active structural acoustic control has been achieved.[4,5] Usually, active methods can easily adapt to the changes in noise problems but their cost is high to deal with high frequency band due to the increased complexity of the controller to take into account many radiating modes of the structure.

Moving away from these limitations, the use of smart materials has been proposed.[6,7] Recently, piezoelectric shunt damping has been utilized on panel structures reducing radiated noise from it.[8] This system is composed of piezoelectric patches and simple shunt circuits. To briefly explain, the concept of piezoelectric shunt damping is energy conversion and energy absorption, similar to the mechanical vibration absorber. Advantages of the shunt damping method in noise problems are effective noise reduction at low frequencies, easy to install, lightweight, and low cost. Hagood and von Flotow have investigated the possibility of dissipating mechanical energy with passive electrical circuits.[9] They optimally tuned an electrical resonance of a shunt circuit to structural resonance in a manner analogous to the mechanical vibration absorber for a selected model. Recently, the electrical impedance model for piezoelectric structures coupled with the shunt impedance model has been used along with maximum energy dissipation to determine tuning parameters.[10] A suppression of the transmitted noise was achieved for broadband frequencies by utilizing a hybrid concept that combines the use of sound absorbing materials for the mid-frequency region and piezoelectric shunt damping for the low frequency region.[8] However, several piezoelectric patches have been used to take into account the multiple vibration modes of the panel. Hollcamp has expended the theory of piezoelectric shunting for the single mode so that a single piezoelectric element can be used to suppress two modes by optimally designing the shunt parameters.[11] Wu accomplished multimodes shunt damping with the blocking circuit.[12] The blocking circuit consists of one parallel capacitor and inductor antiresonance circuit, and is designed to produce infinite electrical impedance at the natural frequencies of all other resonant shunt circuits. A multimode shunt damping of the piezoelectric smart panel has been

[a]Electronic mail: jaehwan@inha.ac.kr; Tel: +82-32-860-7326; Fax: +82-32-868-1716.
[b]Electronic mail: young036@hanmail.net; Tel: +82-32-860-8846; Fax: +82-32-868-1716.

studied for the noise reduction of the panel.[13] On a single piezoelectric patch, a blocked shunt circuit has been connected to implement the multimode shunt damping. However, the shunt circuit tuning is complicated for achieving multimode damping.

Therefore, the use of a new shunt circuit is proposed for smart panels that can reduce the noise of structures in broadband. In this paper, the negative capacitor converter (NCC) circuit is used in a piezoelectric smart panel, which is a kind of negative impedance converter.[14,15] By implementing the negative capacitance of the piezoelectric patch on the smart panel with the NCC circuit, a shunt damping can be achieved. The most attractive benefit on this idea is that this shunt damping can work at all resonance frequencies. Theoretically, the piezoelectric patch can be considered as a capacitor, and this capacitor can be nullified by the NCC circuit. When the capacitor is exactly nullified, the vibration energy in the structure can be totally absorbed through the circuit in all frequency bands, and in turn it results in broadband noise reduction from the structure. This approach has many advantages. First, it can achieve broadband noise reduction with one piezoelectric patch and one shunt circuit. Second, no additional tuning is necessary when the structural characteristics such as natural frequencies are changed. Thus, this method is considerably robust and suitable for multimode shunt damping for broadband noise reduction. The use of a negative capacitor shunt in smart panels is very simple and theoretically perfect.

However, there is a problem in implementing this idea on a real structure with the piezoelectric patch. The capacitance of the piezoelectric patch is not constant at all resonance frequencies. This intrinsic capacitance can have some deviation at resonance frequencies due to the piezoelectric material behavior. Thus, when the NCC circuit is tuned at a resonance frequency, the tuned circuit would not be effective at different resonance frequencies. This can cause some deterioration of the shunt damping performance of panel structures for broadband noise reduction. To resolve this problem, an additional piezoelectric patch that is identical with the shunt piezoelectric patch is introduced in the NCC circuit to mimic the same electric property of the original piezoelectric patch. This paper deals with the implementation process of the NCC circuit with a single piezoelectric patch on the smart panel. The concept of the piezoelectric smart panel with the NCC shunt is introduced and the tuning and implementation procedures are explained. Finally the noise reduction performance of the panel is tested in terms of transmission loss based upon the SAE J1400 test method.[16] The experimental process of the noise test is addressed.

II. APPROACH

A. Piezoelectric smart panel

Figure 1 illustrates the concept of the piezoelectric smart panel. The smart panel is basically a plate structure on which the piezoelectric patch with a shunt circuit is connected. The connected shunt circuit can absorb the electrical energy produced from the piezoelectric patch. The piezoelectric smart panel was designed to reduce the transmitted noise at the low frequency region, for example, below five resonance modes. A $300\times300\times1.5$ mm aluminum plate was used as a host panel structure, and a piezoelectric patch (PZT-5H, $100\times50\times0.5$ mm) was bonded on the plate with epoxy adhesives. Generally, strong radiation modes of the rectangular plate are odd modes. By locating the piezoceramic patch at the center of the panel, the odd symmetric modes can be dealt with.

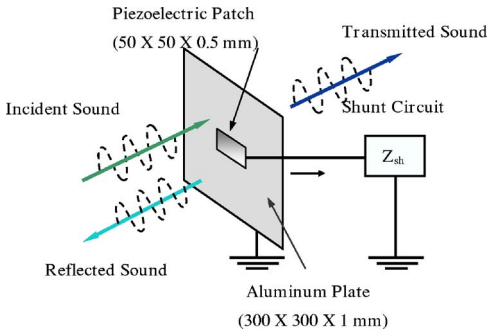


FIG. 1. Schematic diagram of the piezoelectric smart panel for broadband noise reduction. The piezoelectric patch is bonded on the host structure and a shunt circuit is connected to it such that the vibration energy converted from the incident sound will be absorbed in the circuit.

B. Piezoelectric shunt damping

The concept of piezoelectric shunt damping is the energy conversion by a direct piezoelectric effect. Piezoelectric material converts mechanical vibration energy into electrical energy and the converted energy is absorbed or dissipated through the shunt circuit. To evaluate the energy conversion and absorption, the electrical impedance model of piezoelectric structures have been derived from the analogy between mechanics and electrics. This section has a brief summary of the electrical circuit model followed by the resonance shunt and NCC shunt. And, the implementation of the NCC shunt with double piezoelectric patches is addressed.

C. Electric circuit model of the piezoelectric material

Piezoelectric materials can be approximately represented as an equivalent electric circuit at a resonant frequency. Van Dyke's model is well known for equivalent model.[17] This model is generally represented with four parameters; C_0, L_1, R_1, and C_1. C_0 describes an inherent dielectric capacity of the piezoelectric material, while L_1, R_1, and C_1 imply mass, damping, and compliance of the structure, respectively. By invoking the Van Dyke's model, the piezoelectric smart structure on which the piezoelectric patch is bonded along with the shunt circuit can be modeled as Fig. 2. This is an equivalent circuit model for the piezoelectric smart panel. Total impedance of the equivalent circuit can be written as

$$Z = Z_1 + \frac{Z_2 Z_3}{Z_2 + Z_3}. \tag{1}$$

Here, Z_1 and Z_2 are the impedance of the first and second systems, respectively. Z_3 is the impedance of shunt circuit,

$$Z_1 = j\omega L_1 + \frac{1}{j\omega C_1} + R_1,$$

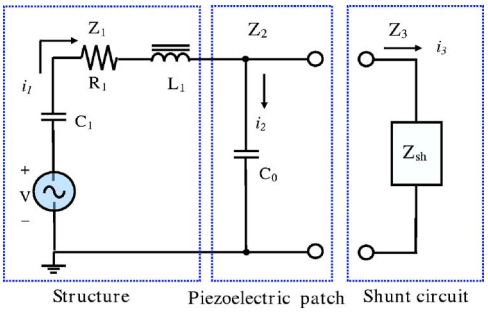


FIG. 2. Electric impedance model for the piezoelectric smart panel. The structure with the piezoelectric patch is modeled as Van Dyke's circuit model.

$$Z_2 = \frac{1}{j\omega C_0}, \tag{2}$$

$$Z_3 = Z_{\text{sh}}.$$

Here, j represents the imaginary variable and ω is the circular frequency. The transfer function of the total system can be expressed in terms of electrical admittance of the piezoelectric structure including shunt circuit as shown in Eq. (3). This is analogous to the ratio of velocity output to applied force at the mechanical system,

$$T_r = \left|\frac{v}{F}\right| = \left|\frac{I}{V}\right| = \frac{1}{|Z|} = |Y| = \left|\frac{Z_2 + Z_3}{Z_1(Z_2 + Z_3) + Z_2 Z_3}\right|. \tag{3}$$

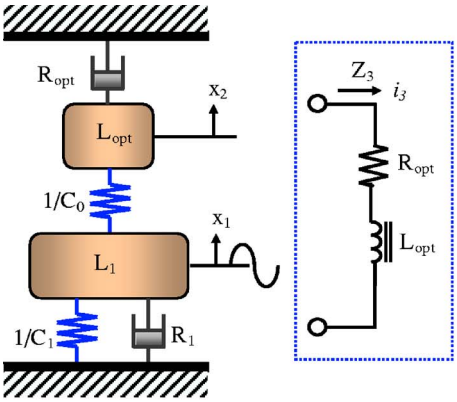
To use the electrical impedance model, coefficients of the Van Dyke's model should be determined. This can be done by measuring the electrical impedance of the piezoelectric patch bonded on the panel structure. The electrical impedance can be measured by using the impedance analyzer (HP4192A), and the equivalent parameters can be determined according to the IEEE Standard on Piezoelectricity.[18]

D. Resonant and NCC shunt dampings

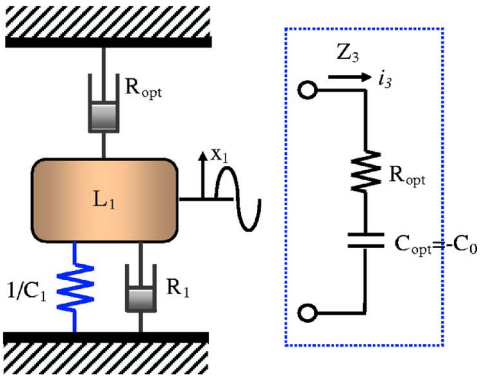
In this paper, resonant shunt circuit and negative capacitance converter shunt circuit are utilized to minimize the transfer function shown in the Eq. (3). The resonant shunt circuit consists of a resistor and an inductor in series. In conjunction with the resonant shunt circuit, the piezoelectric structure can be modeled as the circuit shown in Fig. 2. When the inductor (L_{opt}) of the shunt circuit is connected in parallel to the capacitance (C_0) of the piezoelectric structure, the reactive component of the piezoelectric shunt system is cancelled and the phase between the current and voltage is made zero, which assists in dissipating the energy as much as possible through the resistance. Figure 3(a) shows the mechanical model of the resonant shunt circuit and its electrical impedance model. The resonant shunt circuit is analogous to the mechanical vibration absorber. The impedance of the resonant circuit, Z_{res} can be written as

$$Z_{\text{sh}} = Z_{\text{res}} = R_{\text{opt}} + j\omega L_{\text{opt}}. \tag{4}$$

Then the denominator of the total impedance can be written



(a) Resonant Shunt Damping.



(b) NCC Shunt Damping

FIG. 3. Mechanical model and shunt circuits for resonant damping and NCC shunt damping: (a) resonant shunt damping is composed of resistance R_{opt} and inductance L_{opt}, which are analogous to the mechanical damper and mechanical mass. (b) NCC shunt damping is composed of resistance R_{opt} and negative capacitance $C_{\text{opt}}=-C_0$, which cancels out the mechanical compliance and the damper only remains.

$$Z_2 + Z_{\text{sh}} = R_{\text{opt}} + j\left(\omega L_{\text{opt}} - \frac{1}{\omega C_0}\right). \tag{5}$$

Similarly, the impedance of the NCC shunt circuit, Z_{ncc} can be written as

$$Z_{\text{sh}} = Z_{\text{ncc}} = R_{\text{opt}} + \frac{1}{j\omega C_{\text{opt}}}. \tag{6}$$

And, the numerator of the total impedance is

$$Z_2 + Z_{\text{sh}} = R_{\text{opt}} - j\left(\frac{1}{\omega C_{\text{opt}}} + \frac{1}{\omega C_0}\right). \tag{7}$$

Figure 3(b) shows the mechanical model of the NCC shunt circuit and its electrical impedance. Contrary to resonant shunt, the NCC shunt can be mechanically modeled as a skyhook damper system. If the capacitance of the NCC shunt circuit (C_{opt}) is tuned to the negative capacitance of the piezoelectric structure ($-C_0$), the real part only remains in the numerator of the transfer function, and the shunt damping is independent of the frequency. Since the NCC shunt damping is not necessarily tuned at a particular resonance frequency, multimode shunt damping can be

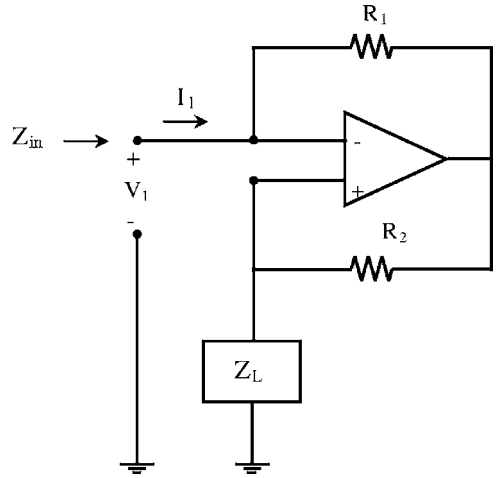


FIG. 4. Negative impedance converter and load impedance. By the ratio of R_1 and R_2, the input impedance can be converted into negative impedance. For NCC, the load impedance is substituted in the capacitance of the piezoelectric shunt patch.

easily realized with a single piezoelectric patch. Thus, this NCC shunt damping can be theoretically robust for the vibration or noise reduction of piezoelectric smart panels.

E. Creating the negative capacitance

The NCC shunt circuit requires a negative capacitance element. It can be implemented by a general class of circuits known as a negative impedance converter. Figure 4 shows the negative impedance converter with load impedance (Z_L). The load impedance is the capacitance of the piezoelectric patch in this research. Using the nodal equilibrium at each node, the input impedance (Z_{in}) of the negative capacitance converter can be described as

$$Z_{\text{in}} = -\frac{R_1}{R_2} Z_L. \tag{8}$$

If $Z_L = 1/j\omega C$, then $Z_{\text{in}} = -1/j\omega C$, such that we can create a negative capacitance, which is scaled by the ratio of the resistors R_1 and R_2.

The NCC shunt damping can ideally attain multimode shunt damping. However, due to the different intrinsic capacitances of the piezoelectric patch at resonance frequencies, the NCC shunt has some deviation in its performance of shunt damping. Therefore, a double-patch NCC shunt is proposed. Figure 5 shows the schematic diagram of the double-patch the NCC shunt. In order to produce the optimal capacitance for the NCC circuit, an identical piezoelectric patch is attached to the opposite side of the plate and is insulated from the plate.

F. Parameter tuning

It is essential to maximize the performance of the piezoelectric shunt damping by adjusting parameters of the shunt circuit. Shunt circuits are composed of resistor, inductor, and capacitor. Values of the shunt parameters should be optimized to achieve effective noise reduction, which is called the optimal parameter tuning. Instead of tuning the transfer function geometrically as used in the conventional tuning

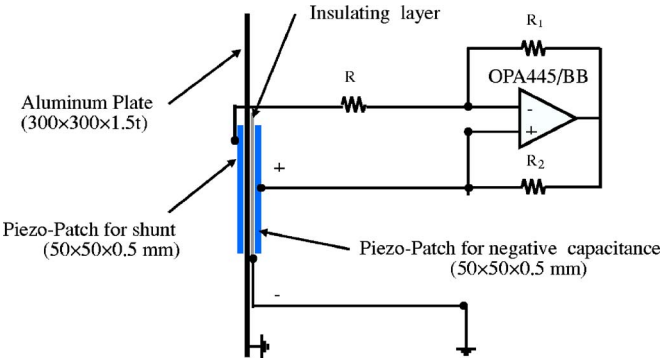


FIG. 5. Schematic diagram of the double-patch NCC shunt circuit. By using the identical piezoelectric patch, the deviation of capacitance of the piezoelectric shunt patch can be eliminated in the NCC shunt circuit.

method for the dynamic absorber, the new parameter tuning method associated with maximum dissipated energy at the shunt circuit is adopted.[10] Conventional tuning method is not valid for negative shunt. The ratio of the dissipated energy to the input energy is

$$J = \frac{P_D}{P_{\text{IN}}} = \frac{\text{Re}(Z_3) \cdot \left| \left(\frac{Z_2}{Z_2 + Z_3} \right) \right|^2}{|Z|}. \tag{9}$$

This ratio is given at a specific frequency near resonance. In the tuning process, however, this should be maximized by optimally changing the shunt circuit parameters. Thus, the objective function in the optimization is taken as the averaged J at a certain frequency band near the targeted resonance frequency. The optimal design variables L^* and R^* in resonant shunt are found by maximizing the objective function,

$$[L^*, R^*] = \max_{\substack{L \\ R}} \left[\frac{1}{n} \sum_{k=1}^{n} |J_k| \right]. \tag{10}$$

Here, n is the number of single frequency points in the frequency band. In the case of the NCC shunt circuit, a negative capacitance element is tuned to the capacitance found from the admittance curve and the resistor is determined by the optimal tuning process,

$$[R^*] = \max_{R} \left[\frac{1}{n} \sum_{k=1}^{n} |J_k| \right]. \tag{11}$$

III. EXPERIMENTS

The tests were conducted at the sound transmission test facility built according to the standardized test specification in SAE J1400.[16] Figure 6 shows the test facility. The transmission loss test facility has two adjacent chambers, a reverberation chamber and a semianechoic chamber. A smart panel was placed on a window located between two chambers. The reverberation chamber and semianechoic chamber are designed to realize diffused fields and free fields, respectively. The reverberation chamber was suitable for use at 300 Hz and above of the volume limitation. Sound was gen-

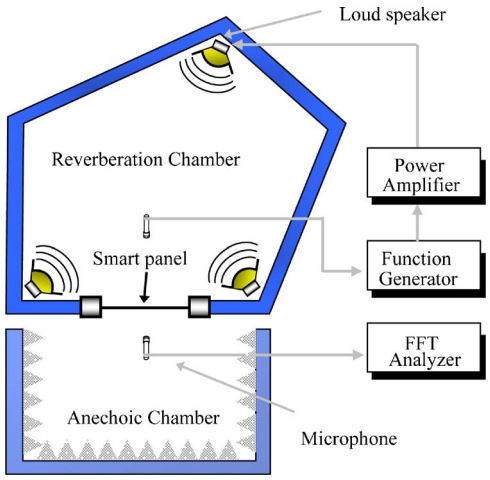


FIG. 6. Schematic diagram of experimental apparatus for transmission loss measurement. The facility has a reverberation chamber and a semianechoic chamber. The smart panel is placed on a window located between two chambers.

erated in the reverberation chamber and the amount of sound transmitted through the window was measured.

Before conducting shunt-damping tests, modal analysis of the panel structure was performed numerically by using a finite element program. Figure 7 shows mode shapes of the panel structure. From the results, the location of the piezoelectric patch was chosen at the center of the panel, and target modes were determined as the fifth (504 Hz), ninth (860 Hz), 11th (1255 Hz), and 21st (1629 Hz) modes. Although the sound was strongly radiated at the first mode, the first mode is not considered as the target mode because diffuse fields in the reverberation chamber were not guaranteed below 300 Hz. This is not the shunt damping limitation. The shunt damping for the transmitted noise reduction at the first mode has been demonstrated.[8] After bonding the piezoceramic patch at the center of the panel, the admittance at the patch was measured to tune the shunt circuit.

The transmission loss is expressed in terms of the sound power in the reverberation chamber with respect to the sound power in the semianechoic chamber. In other words, it can be described with the difference of a sound pressure and a material property of the test specimen as follows:

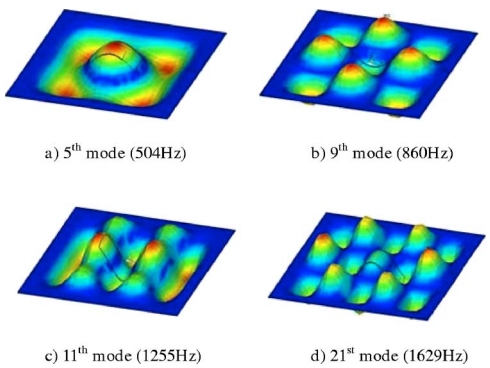


FIG. 7. Modal analysis results of the panel. Since odd modes are strong radiators, they are only shown. The first and third modes are not shown since they are below 300 Hz, the lower limit frequency of the transmission loss measurement setup.

$$\mathrm{TL} = \mathrm{MNR} - 10\log(A/S\alpha), \tag{12}$$

where TL is the transmission loss of the panel, MNR the measured noise reduction between the reverberation chamber and receiving room, $S\alpha$ the Sabine absorption of the receiving room, and A is the area of the test window. The expression $10\log_{10}(A/S\alpha)$ is constant for any test panel with the same area.

Using a random noise with a bandwidth of 300–5000 Hz, we measured the noise level in the reverberation and the receiving chambers in order to determine the measured noise reduction (MNR). Since the TL difference was found with and without the shunt, the last term in Eq. (12) is cancelled out in the TL difference calculation, and is not necessarily found.

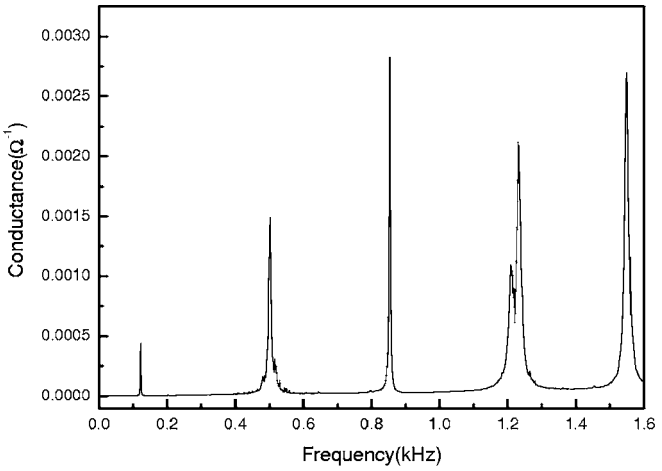
IV. RESULTS AND DISCUSSIONS

A. Resonant shunt damping

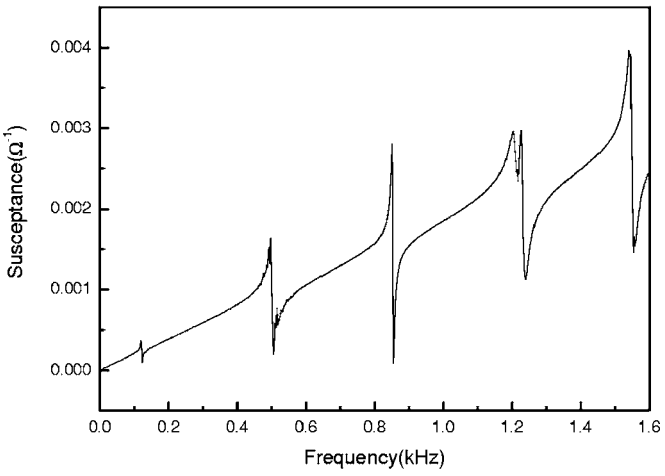
At first, the resonant shunt experiment was accomplished for the transmitted noise reduction for fifth and ninth modes separately. $100\times50\times0.5$ mm piezoelectric patch (PZT-5H) was bonded at the center of the plate structure. Figure 8 shows the measured admittance of the piezoelectric smart panel in terms of conductance (real part) and susceptance (imaginary part). As we can see, the fifth and ninth resonance frequencies were found to be 502 Hz and 852 Hz, respectively. Compared to the finite element analysis results, these values are quite accurate. From the measured admittance curves, the parameters for the equivalent impedance model were extracted for the fifth and ninth modes, respectively. The second column in Table I shows these values. The optimally tuned resonant shunt parameters were shown in the third column of Table I. As a reference performance of the shunt damping, the acceleration signal of the panel structure was measured at the center of the structure. As Fig. 9 shows, the acceleration on the structure was reduced by 18 dB at the fifth and ninth modes. With the same shunt parameters, transmission losses at the fifth and ninth modes were measured individually. Figure 10 shows the transmission loss difference between shunt and no shunt cases. The TL difference was found by subtracting the MNR values between the shunt and no shunt cases. Positive TL difference means that the noise reduction is increased by the shunt damping. When the resonant shunt was tuned at 502 Hz, 8 dB of the transmission loss was achieved at 500 Hz third-octave band, while 4 dB of the transmission loss was achieved at 800 Hz third-octave band when the shunt was tuned at 852 Hz, separately.

B. Negative capacitance converter shunt damping

Second, the multimode shunt experiment was performed using the NCC shunt circuit. A piezoelectric patch (PZT-5H) of $50\times50\times0.5$ mm was bonded at the center of the plate structure and the identical piezoelectric patch was bonded on the other side to comprise the NCC shunt circuit. Table II shows the equivalent circuit (Van Dyke) parameters and optimal shunt parameters for the NCC circuit. The capacitance in the NCC shunt circuit was tuned in the capacitance found from the impedance curve and the resistance was determined

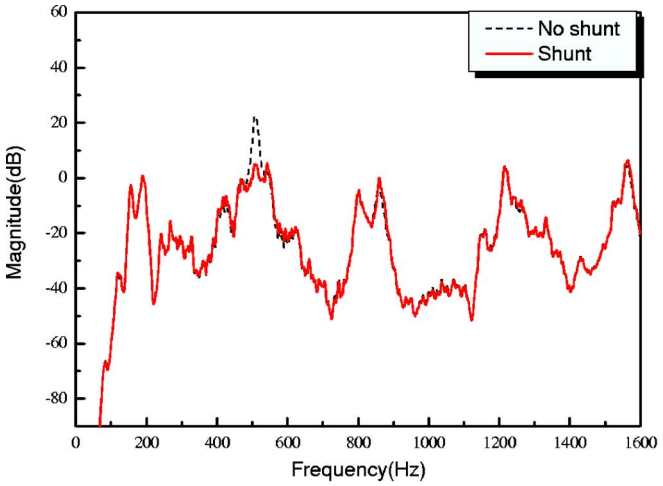


(a) Conductance

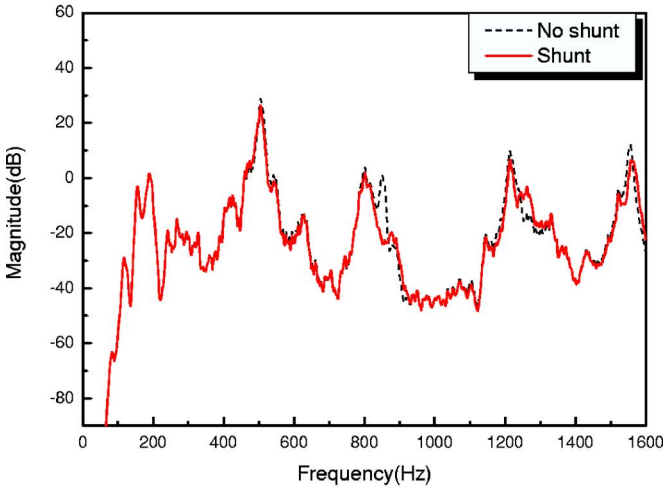


(a) Susceptance

FIG. 8. Measured admittance of the piezoelectric panel.



a) 5th mode



b) 9th mode

FIG. 9. Acceleration of resonant shunt damping. The resonant shunt was tuned at the fifth and ninth modes separately. The acceleration level was reduced by 18 dB at these modes.

by the optimal tuning process at the fifth mode. Figure 11 shows the acceleration and the transmission loss difference of the panel structure. The acceleration levels at the fifth, ninth, 11th, and 21st modes were reduced by 15 dB, 7 dB, 4 dB, and 3 dB, respectively. The transmission loss was increased by 7–2 dB at 500–2000 Hz third-octave band. As one can see in Fig. 11, the shunt damping effect was reduced as frequency was increased. To investigate this effect, the capacitance of the piezoelectric patch was found from the impedance curve (second column in Table III). Because the capacitance value of the piezoelectric patch was decreased with the frequency, the tuned NCC circuit at the fifth mode was not so effective at higher modes. This is the difficulty in tuning the NCC shunt circuit.

TABLE I. Van Dyke's coefficients (C_0, C_1, L_1, R_1) found at the fifth and ninth modes, and the optimal parameters (L_{opt}, R_{opt}) for resonant shunt damping.

	Van Dyke coefficients		
Frequency	Coefficient	Values	Shunt parameters (optimal)
Fifth mode (502 Hz)	C_0 (F)	3.000e−7	L_{opt}=0.3543 H
	C_1 (F)	8.766e−9	
	L_1 (H)	11.49	R_{opt}=130Ω
	R_1 (Ω)	658.7	
Ninth mode (852 Hz)	C_0 (F)	3.053e−7	L_{opt}=0.1204 H
	C_1 (F)	1.474e−9	
	L_1 (H)	25.09	R_{opt}=34.4Ω
	R_1 (Ω)	576.3	

Therefore, the double-patch NCC shunt was used as shown in Fig. 5. In order to produce the optimal capacitance for the NCC circuit, an identical piezoelectric patch was attached to the opposite side of the panel structure. Figure 12 shows the measured admittance of both piezoelectric patches. Table III shows the capacitance values founded from the impedance curve. The capacitance values of both sides were not exactly the same, which might be caused by the differences in the material, thickness, size or/and bonding layer. Nevertheless, the ratio of two capacitance values at

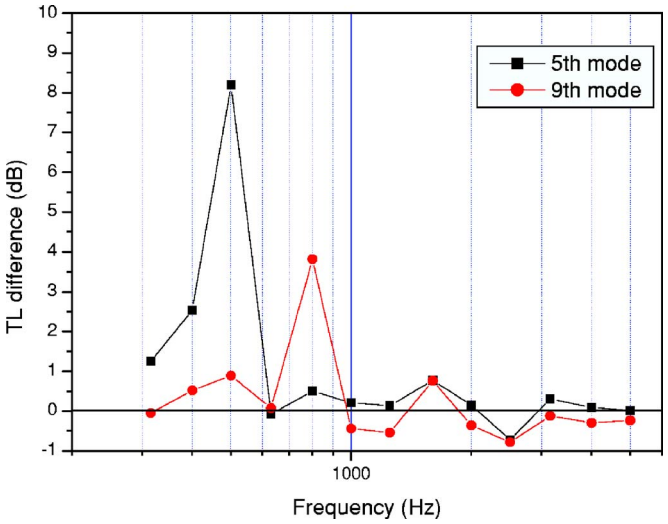


FIG. 10. Transmission loss difference of resonant shunt damping. When the fifth and ninth modes were shunted separately, 8 dB and 4 dB TL improvements were achieved at 500 and 800 Hz third-octave band, respectively.

each resonance frequency was constant. Figure 13 shows the acceleration and the transmission loss before and after switching on the double-patch NCC shunt circuit. Compared to the resonant shunt, more acceleration reduction was obtained by 10–5 dB at four resonance frequencies. The shunt damping performance at 130 Hz was less than other resonance mode. Possible reasons of this phenomenon might be the low induced voltage of the front PZT patch at the low resonance frequency, and the interference of the induced voltage of the back PZT patch. Once the large PZT patch is used, the induced voltage from the front patch can be increased. In Fig. 13(b), the transmission loss was increased by 7–4 dB at 500–2000 Hz third-octave band. Since it was accomplished with the same piezoelectric patch area of the resonant shunt damping, this approach is more efficient for broadband noise reduction.

This NCC shunt damping method has merits in terms of easy to tune, simple in configuration and robust in resonance frequency change. The resonant shunt method and multimode shunt dampings are sensitive to the change of resonance frequencies. Although the NCC shunt can be used for broadband frequency, the combination of resonant shunt at the lower mode and NCC shunt for mid-frequency modes may be a good engineering solution for broadband noise reduction because the optimal location and size of the PZT patches for the shunt damping would be different for each resonance mode.

TABLE II. Van Dyke's coefficients (C_0, C_1, L_1, R_1) found at the fifth mode, and the optimal shunt parameters C_{opt} and R_{opt} for the NCC shunt damping.

Freq. (Hz)	Van Dyke coefficients		Shunt parameters (optimal)
NCC Shunt (fifth mode)	C_0 (F)	3.000e−7	C_{opt}=3.04e−7 F
	C_1 (F)	8.766e−9	
	L_1 (H)	11.49	R_{opt}=40 Ω
	R_1 (Ω)	658.7	

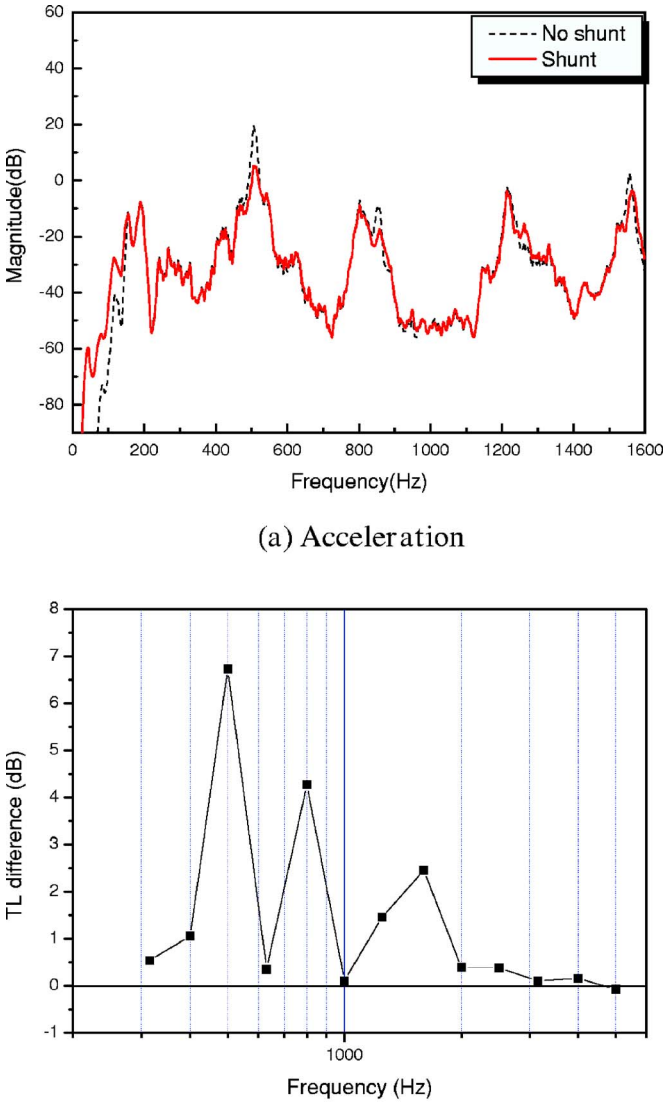


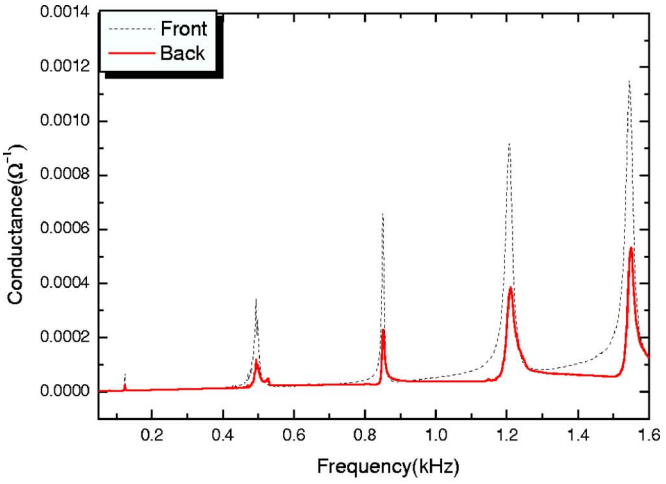
FIG. 11. Acceleration and transmission loss of ordinary NCC shunt damping. The acceleration levels were reduced by 15–3 dB at above the fifth modes, while the transmission loss was increased by 7–2 dB at 500–2000 Hz third-octave band.

V. CONCLUSIONS

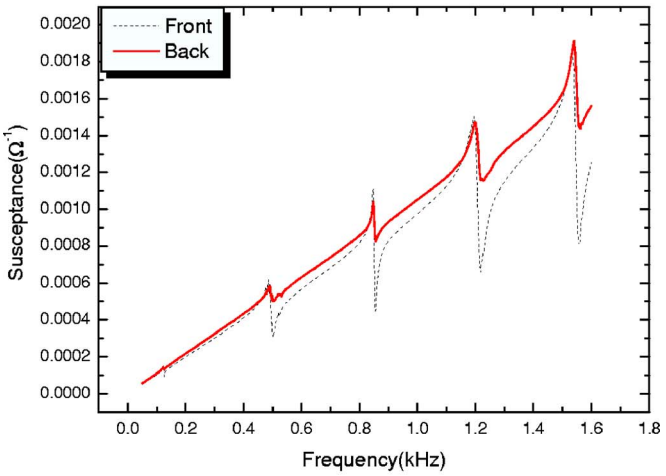
In this paper, a broadband noise reduction of the smart panel featuring negative capacitance converter shunt circuit was experimentally investigated. To eliminate the variation of capacitance of piezoelectric patch, dual-patch NCC shunt circuit was introduced. The performance smart panel on which the dual-patch NCC shunt circuit was attached was compared with the panels featuring resonant shunt circuit

TABLE III. The capacitance values for front and back piezoelectric patches.

Frequency (Hz)	Parameter	Front	Back
502	C_0 (F)	1.624e^{-7} F	1.736e^{-7} F
852	C_0 (F)	1.610e^{-7} F	1.719e^{-7} F
1230	C_0 (F)	1.601e^{-7} F	1.703e^{-7} F
1549	C_0 (F)	1.582e^{-7} F	1.679e^{-7} F

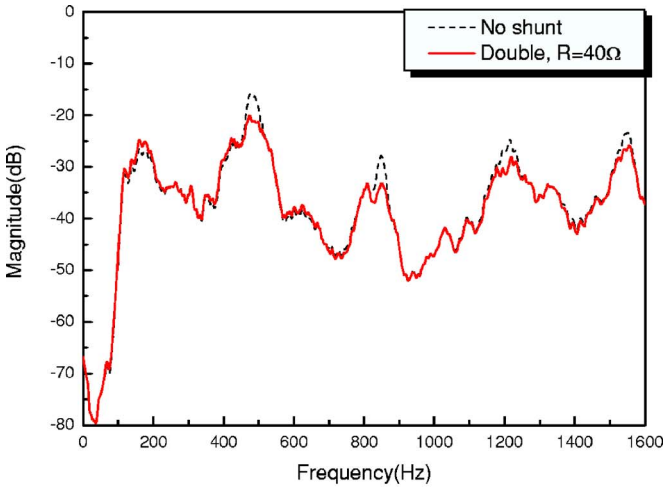


a) Conductance

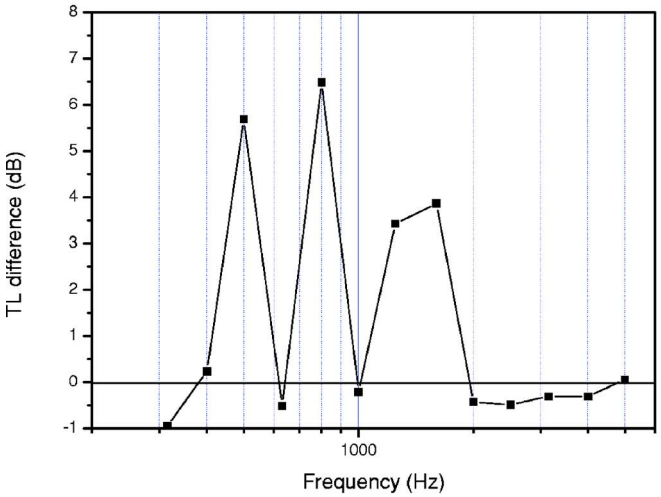


b) Susceptance

FIG. 12. Admittance of dual piezoelectric patches. Admittance curves on the front and back piezoelectric patches were measured by the HP 4192A impedance analyzer. Two curves were almost identical but some deviation was observed, which was caused by material damping and dimension deviations, possibly.



(a) Acceleration



(a) Transmission Loss

FIG. 13. The acceleration and transmission loss of dual-patch NCC shunt damping. The acceleration level was reduced by 10–5 dB at four resonance frequencies with one piezoelectric patch and the NCC shunt circuit. The transmission loss was increased by 7–4 dB at 500–2000 Hz third-octave band.

and ordinary NCC shunt circuit, in terms of acceleration and noise transmission loss. As a tuning method, maximum dissipated energy method in conjunction with the electrical impedance model was adopted. The noise transmission loss was measured according to the SAE J1400 standard.

By using an ordinary NCC shunt circuit, transmission loss of the smart panel was increased by 7–2 dB at 500–2000 Hz third-octave band. However, the performance of the ordinary NCC was reduced as the frequency was increased; it was due to the presence of different intrinsic capacitance of the piezoelectric patch. When dual-patch NCC shunt was used, the transmission loss was increased by 7–4 dB at 500–2000 Hz third-octave band. This noise reduction performance was better than that of the smart panel featuring resonant shunt. The dual-patch NCC shunt is more efficient than ordinary NCC and resonant shunt for achieving broadband noise reduction with smart panels.

In conclusion, piezoelectric smart panels featuring dual-patch NCC shunt is very promising for broadband noise reduction. More investigation on optimizing the location and size of the piezoelectric patch, along with a demonstration on real applications should be followed.

ACKNOWLEDGMENT

This work was supported by the Creative Research Initiative Program of Korea Science and Engineering Foundation, Republic of Korea.

[1]A. D. Belegunde, R. R. Salagame, and G. H. Koopmann, "A general optimization strategy for sound power minimization," Struct. Multidiscip. Optim. **8**, 113–119 (1994).
[2]R. L. St. Pierre, Jr. and G. H. Koopmann, "A design method for minimizing the sound power radiated from plates by adding optimally sized, discrete masses," ASME J. Vibr. Acoust. **117**, 243–251 (1995).
[3]J. S. Bolton, N.-M. Shiau, and Y. J. Kang, "Sound transmission through multipanel structures lined with elastic porous materials," J. Sound Vib. **191**, 317–347 (1996).
[4]C. R. Fuller, "Active control of sound transmission/radiation from elastic

plate by vibration input. I. Analysis," J. Sound Vib. **136**, 1–15 (1990).
[5]D. R. Thomas, P. A. Nelson, and S. J. Elliott, "Experiments on reduction of propeller induced interior noise by active control of cylinder vibration," J. Sound Vib. **112**, 389–395 (1987).
[6]I. Pelinescu and B. Balachandran, "Analytical study of active control of wave transmission through cylindrical struts," Smart Mater. Struct. **10**, 121–136 (2001).
[7]C. A. Gentry, C. Guigo, and C. R. Fuller, "Smart foam for applications in passive-active noise radiation control," J. Acoust. Soc. Am. **101**, 1771–1778 (1997).
[8]J. Kim and J. K. Lee, "Broadband transmission noise reduction of smart panels featuring piezoelectric shunt circuits and sound absorbing material," J. Acoust. Soc. Am. **112**, 990–1008 (2002).
[9]N. W. Hagood and A. von Flotow, "Damping of structural vibrations with piezoelectric materials and passive electrical networks," J. Sound Vib. **146**, 243–268 (1991).
[10]J. Kim, Y.-H. Ryu, and S.-B. Choi, "New shunting parameter tuning method for piezoelectric damping based on measured electrical impedance," Smart Mater. Struct. **9**, 868–877 (2000).
[11]J. J. Hollkamp, "Multimodal passive vibration suppression with piezoelectric materials and resonant shunts," J. Intell. Mater. Syst. Struct. **5**, 49–57 (1994).
[12]S. Y. Wu, "Multiple PZT transducer implemented with multiple-mode piezoelectric shunt for passive vibration damping," in *Proceeding of the SPIE: Smart Structures and Materials 1999: Passive Damping and Isolation* (SPIE, Seattle, 1999), Vol. **3672**, pp. 112–122.
[13]J. Kim and J.-H. Kim, "Multimode shunt damping of smart panel for noise reduction," J. Acoust. Soc. Am. **116**, 942–948 (2004).
[14]A. I. Larky, "Negative impedance converters," IRE Trans. Circuit Theory **4**, 124–131 (1957).
[15]S. Behrens, A. J. Fleming, and S. O. R. Moheimani, "New method for multiple-mode shunt damping of structural vibration using a single piezoelectric transducer," in Proceedings of the SPIE Smart Structures and Materials 2001: Passive Damping and Isolation (SPIE, Seattle, 1999), Vol. 4331, pp. 239–250.
[16]SAE Standards, Document Number J1400: Laboratory measurements of the airborne sound barrier performance of automotive materials and assemblies (SAE 1990).
[17]K. Uchino, *Piezoelectric Actuators and Ultrasonic Motors* (Kluwer Academic, Boston, 1997).
[18]IEEE Standard on Piezoelectricity: ANSI/IEEE Standard 176–1987 (IEEE 1987).

Hybrid feedforward-feedback active noise reduction for hearing protection and communication

Laura R. Ray, Jason A. Solbeck, Alexander D. Streeter, and Robert D. Collier
Thayer School of Engineering, Dartmouth College, 8000 Cummings Hall, Hanover, New Hampshire 03755

(Received 8 August 2005; revised 7 July 2006; accepted 8 July 2006)

A hybrid active noise reduction (ANR) architecture is presented and validated for a circumaural earcup and a communication earplug. The hybrid system combines source-independent feedback ANR with a Lyapunov-tuned leaky LMS filter (LyLMS) improving gain stability margins over feedforward ANR alone. In flat plate testing, the earcup demonstrates an overall C-weighted total noise reduction of 40 dB and 30–32 dB, respectively, for 50–800 Hz sum-of-tones noise and for aircraft or helicopter cockpit noise, improving low frequency (<100 Hz) performance by up to 15 dB over either control component acting individually. For the earplug, a filtered-*X* implementation of the LyLMS accommodates its nonconstant cancellation path gain. A fast time-domain identification method provides a high-fidelity, computationally efficient, infinite impulse response cancellation path model, which is used for both the filtered-*X* implementation and communication feedthrough. Insertion loss measurements made with a manikin show overall C-weighted total noise reduction provided by the ANR earplug of 46–48 dB for sum-of-tones 80–2000 Hz and 40–41 dB from 63 to 3000 Hz for UH-60 helicopter noise, with negligible degradation in attenuation during speech communication. For both hearing protectors, a stability metric improves by a factor of 2 to several orders of magnitude through hybrid ANR. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2259790]

PACS number(s): 43.50.Hg, 43.50.Ki, 43.72.Dv [KAC] Pages: 2026–2036

I. INTRODUCTION

Active noise reduction (ANR) has received considerable attention in the literature as a means of improving attenuation provided by personal hearing protection devices (Akelsson, 1996; McKinley and Nixon, 1993). In commercial feedback ANR devices, the cancellation path transfer function, which is a combination of the ANR speaker characteristics, cavity resonant behavior, and error microphone placement, limits the feedback gain in order to retain stability, and thus the level of active attenuation is limited. ANR measurements for commercial feedback headsets show active performance bands of 10–20 dB from 50 to 400 Hz for stationary broadband white noise (Akelsson, 1996; McKinley and Nixon, 1993; Collier *et al.*, 2003). Feedback ANR often adds noise in the midfrequency speech communication band due to this tradeoff between stability and sensitivity (McKinley and Nixon, 1993; Ward, 1997).

These limitations have spawned research on feedforward ANR using least-mean-squared (LMS) adaptive filters (Pan *et al.*, 1994; Pan *et al.*, 1997; Brammer and Pan, 1998; Cartes *et al.*, 2002a; 2003; Ray *et al.*, 2002). The large dynamic range of noise fields under which LMS filters must operate results in a time-varying signal to noise ratio (SNR) on the measured inputs and associated stability issues (Cartes *et al.*, 2003). Stability is generally retained by using a leaky LMS filter, improving stability under worst-case SNR, but degrading performance in high SNR conditions (Gitlin, 1982). Moreover, tuning the leakage parameter is a highly empirical process. A Lyapunov tuning method, reported and validated in Cartes *et al.* (2002a, 2003) provides an adaptive leakage factor and step size that reduces the tradeoff between LMS filter stability and noise reduction performance in response to time-varying SNR and minimizes empirical tuning.

The advantage of feedforward ANR indicated by the authors' prior research is an increase in low frequency performance over feedback ANR and improved response to tonal noise and nonstationary noise (Collier *et al.*, 2003). However, as with all feedforward systems, the ability to tolerate cancellation path uncertainty is limited, and such uncertainty also impacts performance and can cause instability. Filtered-*X* LMS (FXLMS) variants accommodate the cancellation path gain by prefiltering the reference signal (Kuo and Morgan, 1996). A FXLMS variant has been successfully used for ANR in circumaural earcups in Pan *et al.*, (1997) and Pan *et al.* (1994). However, if the cancellation path transfer function changes during use, with aging, or due to variable sealing conditions of the hearing protector, gain and phase errors are introduced within the FXLMS. Thus, good stability margins are necessary, even with the FXLMS.

A traditional approach to improving stability margins while minimizing performance degradation is to combine feedforward and feedback compensation (Ogata, 1990). When a disturbance can be measured, feedforward control reduces its effect, and feedback control accommodates the residual disturbance with lower gain, which in turn enhances stability margins. In acoustic noise control, this hybrid architecture retains a feedback ANR feature of insensitivity to noise source characteristics, while the feedforward compensator remains effective in canceling tonal noise. Hybrid ANR for hearing protection is studied in Winberg *et al.* (1999) and Rafaely and Jones (2002). In Winberg *et al.* (1999), a complex filtered-*X* LMS filter of length 40 is combined with a commercial analog feedback controller to attenuate 17.7 Hz

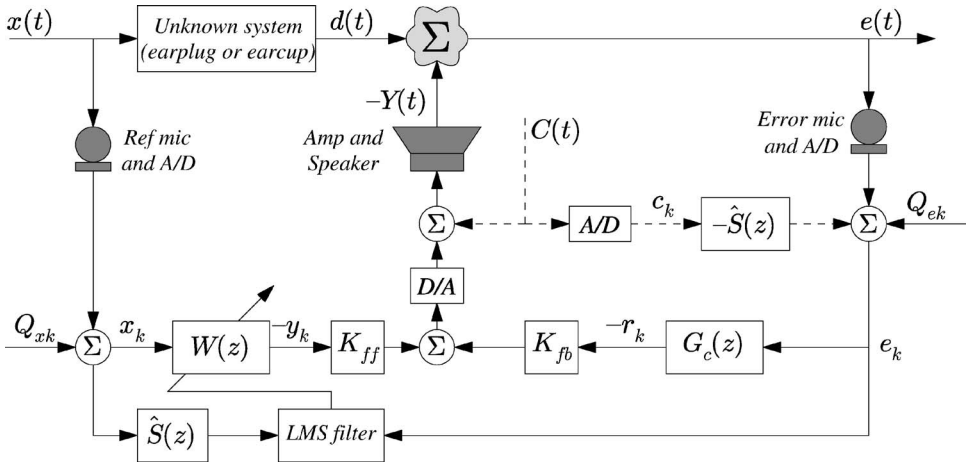


FIG. 1. Hybrid feedforward-feedback ANR topology.

infrasonic helicopter blade passage noise and its harmonics. Feedforward ANR based on a tachometer signal is evaluated by computer simulation. While the feedback system combined with passive earcup attenuation provides 20 dB broadband performance, reduction of the blade frequency is minimal. Computer simulation shows that adding the FXLMS filter reduces the fundamental blade frequency by an additional 20 dB. Rafaely and Jones (2002) study hybrid ANR performance in both reverberant and directional sound fields, and explore the relation between ANR performance and forward path delay. The results show improved performance of the hybrid system in a reverberant field, as compared to the original analog feedback system, while a directional field, which affects the acoustic delay, degrades performance when the noise source does not directly face the reference microphone.

In this paper, we focus on stability and performance aspects of hybrid feedforward-feedback control. We implement a feedback compensator digitally (rather than use an existing commercial control unit) to measure the effect of feedback loop gain on stability and performance of the hybrid system. We pair the feedback controller with a LyLMS filter variant in order to achieve good performance for a large dynamic range and a variety of noise sources. ANR stability and performance metrics are measured experimentally using an earcup from a commercial circumaural headset and a commercial communication earplug, both modified for hybrid ANR. For the earcup, the feedback compensator is paired with the LyLMS algorithm to study hybrid ANR, and for the earplug, which exhibits a more variable cancellation path response than the earcup, the feedback compensator is paired with a filtered-*X* LyLMS (FXLyLMS) algorithm. In the context of introducing the filtered-*X* for the earplug, we consider computationally efficient infinite-impulse response (IIR) cancellation path models. Earplug results are presented both with and without a communication signal present. For both devices, the hybrid architecture improves gain stability margin and reduces sensitivity of overall performance on the temporal characteristics of the noise source.

II. HYBRID FEEDFORWARD-FEEDBACK ANR

Figure 1 shows a block diagram of the hybrid system. The incoming noise $x(t)$ is measured by an electret microphone on the exterior of the hearing protector and is digitized as x_k. The past L samples of x_k constitute the reference input X_k, where L is the filter length. Electronic and quantization noise enters as Q_{xk}. As $x(t)$ passes through the hearing protector to become noise signal $d(t)$, the LMS filter finds a weight vector, $W(z)$, which is applied to x_k to produce a cancellation signal $-y_k = W^T X_k$. An error microphone inside the hearing protector registers the error signal, which is digitized subject to noise Q_{ek}. e_k, along with X_k filtered through $\hat{S}(z)$, adjusts the LMS filter, and e_k also passes through feedback compensator, $G_c(z)$, which creates its own cancellation signal $-r_k$. The two cancellation signals are scaled by gains K_{fb} and K_{ff}, summed, and digitized. The cancellation signal is amplified and broadcast by a speaker as $-Y(t)$ to sum with $d(t)$ within the earcup or earplug cavity. $\hat{S}(z)$ models the cancellation path response from the input voltage to the speaker to output voltage of the error microphone, as in a standard filtered-*X* LMS (FXLMS) algorithm (Kuo and Morgan, 1996). Without the shaping filter, feedforward ANR can be unstable, depending on variation in the cancellation path response.

Figure 1 assumes that a single speaker delivers both cancellation and communication signals, though the architecture is easily modified for a dual speaker system. When a communication signal $C(t)$ is injected, it is sampled and filtered through $\hat{S}(z)$. The result is subtracted from the measured error signal prior to ANR computations so that the residual e_k entering the LMS filter and compensator is due to acoustic noise. $C(t)$, is also passed through to the speaker. This process minimizes cancellation of the communication signal along with the external noise and corruption the LyLMS weight vector due to communication. Note that $C(t)$ could serve as a reference input to the feedback loop; however, this requires a closed-loop response with sufficient bandwidth to pass the signal.

Figure 2 shows the sample cancellation path responses for the earcup on a flat plate and for several configurations of the earphone modified for hybrid ANR. Data were recorded at an input level of approximately 105 dB. Earplug data were collected with the earplug inserted in a flat plate (enclosing a volume comparable to that enclosed within the human ear); in a manikin conforming with standards of an artificial ear at

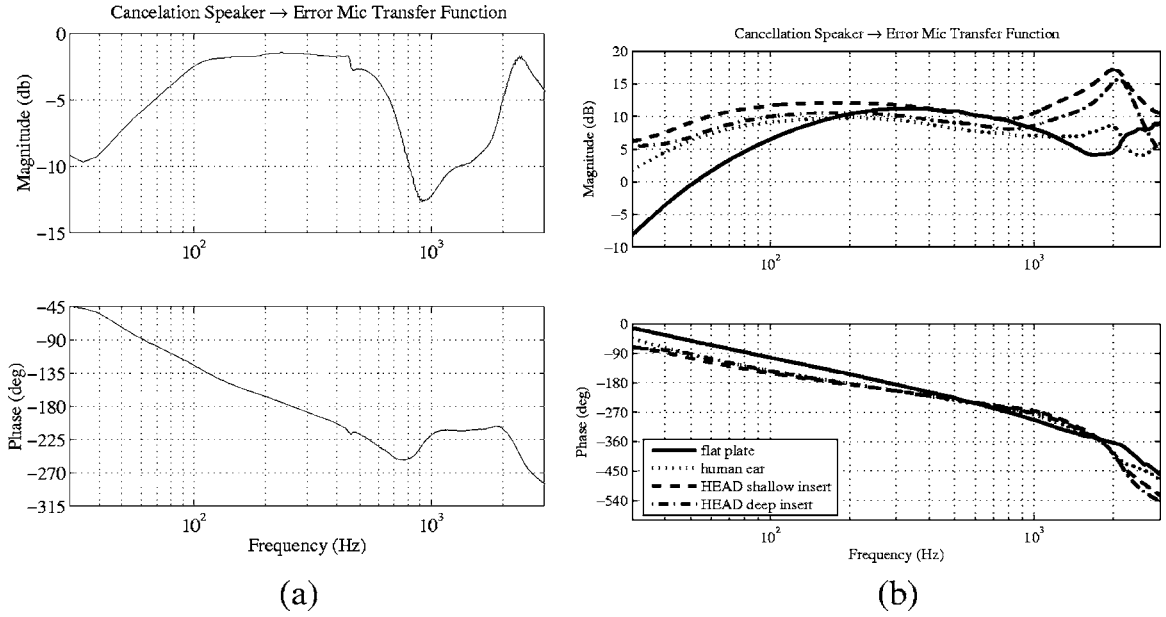


FIG. 2. Open loop transfer function from an internally generated signal, through the cancellation speaker, to the error microphone. (a) Circumaural earcup on flat plate; (b) ANR earplug in flat plate, human ear, and two insertion depths in HEAD Acoustics manikin.

two insertion depths; and within a human ear, to compare responses that can be elicited using common test configurations. The general characteristics of the cancellation path for these devices are similar, indicating that a common ANR architecture is viable. However, variability in the cancellation path response necessitates a system with good stability margins, which poses a challenge for feedback and feedforward ANR individually.

A frequency-dependent cancellation path gain is accommodated using a FXLMS filter, as in Fig. 1, in which a filter shapes the reference input prior to the LMS filter update (Kuo and Morgan, 1996); however, to the extent that the cancellation path varies from user to user, the shaping filter $\hat{S}(z)$ needs either to be adaptive or robust to such variations. Similarly, the feedback system must also be robust to such variations. An adaptive cancellation path filter adds substantial computation—up to double that of the system without a cancellation path model, while a fixed cancellation path filter does not avoid gain and phase errors. The hybrid architecture provides a means to minimize performance degradation while building in adequate stability margins in the face of these variations. The feedback compensator $G_c(z)$ provides a relatively low (5–10 dB) attenuation and effectively "flattens" the cancellation path response, such that the feedback compensated cancellation path gain is less variable than the open-loop gain. Feedforward ANR is based on the Lyapunov-tuned LMS (LyLMS) feedforward algorithm. Tuning laws are detailed in Cartes *et al.* (2002a and 2003).

III. CANCELLATION PATH MODELING

The cancellation path $\hat{S}(z)$ can be represented by either a finite-impulse response (FIR) or infinite-impulse response (IIR). An FIR filter introduces on the order of $2N$ multiplies; N multiplies each for filtering the sampled communication signal c_k, and reference input x_k, where N is the cancellation path filter length. In support of computational efficiency, a "blackbox" IIR cancellation path modeling approach was developed. The automated identification method provides a short white noise burst of moderate volume to the cancellation speaker. The time-domain input and error microphone output data are processed using a fast linear identification technique (Phan *et al.*, 2004) referred to here as *fastid*. This approach, which is intended as an initialization routine, can provide high-fidelity, low-order IIR models for communication feedthrough and filtered-X implementation, using as little as 0.1 s of input-output data.

The computation and memory requirements for *fastid* are high as the algorithm requires inversion of a $p(q+r)+r$ square matrix, where p is the order of the IIR filter, q is the number of outputs, and r is the number of inputs. For example, identification of a tenth order, single-input, single-output model requires a 21×21 matrix inversion. A common approach for IIR filter identification is the recursive least-squares (RLS) algorithm (Juang, 1994). The RLS algorithm begins with a set of IIR coefficients and updates them based on each new sample of input-output data until convergence. For a single-input, single-output system, the only nonscalar operations are 2×2 matrix inversions. The RLS model should be equivalent to that identified using *fastid*. However, the RLS algorithm requires significantly more time-series data to converge to a model of similar fidelity to the *fastid* method, as the *fastid* method benefits from having the entire time series of input-output data available for identification. Experimental validation of *fastid* compared with a RLS and a FIR cancellation path model is given in Sec. V.

IV. EXPERIMENTAL CONFIGURATION AND PROCEDURES

A. Experimental facilities

Experimental evaluation of the hybrid ANR architecture was conducted in a Low Frequency Acoustic Test Cell

(LFATC) for the circumaural earcup, and for the earplug, a combination of LFATC and manikin testing was performed. The LFATC, described in Cartes *et al.* (2002b) has a flat (to ±1 dB) acoustic frequency response from 10 to 200 Hz. Digital equalization extends this range to approximately 1000 Hz. A 100 W speaker mounted in the top plate of the cell provides the noise signal and two Brüel & Kjær (B&K) 4190 Type I microphones mounted through the sidewall and mounted axially in the base plate (through a $\frac{1}{2}$ in. diam, 3 in. long hole) provide calibrated measurements of source and error signals, respectively. Noise floors of these microphones average 50 dB in the measurement range 40–1250 Hz. For earcup testing, a single earcup is secured over the base plate of the test cell. The B&K microphone in the aluminum base under the earcup represents the location of the external opening to the ear canal. Details regarding microphone locations are provided in Cartes *et al.* (2002b). For earplug testing within the LFATC, an ear canal adaptor is inserted into the B&K microphone hole, and the prototype earplug is placed in this adaptor. In this manner, the diaphragm of the B&K microphone is approximately $\frac{1}{2}-\frac{3}{4}$ in. from the tip of the earplug, depending on the earplug insertion depth. The ear canal adaptor is a compliant device intended for use with a HEAD Acoustics manikin, described below, and not specifically designed for use with the test cell.

The HEAD Acoustics Artificial Head Measurement System HMS II.3 reproduces a head related transfer function, conforms to standards ITU-P.57, P.58, and IEC 711, and has left and right ITU-T P.57 Type 3.3 Pinna (ITU, 1998, ITU, 2002, and IEC, 1981). It provides an intermediate facility between the LFATC and human subject testing for hearing protector evaluation. All manikin testing is performed in a 450 ft.^{2} room with gypsum board walls, carpeted floor, and 7.5 ft. high ceiling. The T20 reverberation times measured 0.32 s (125 Hz), 0.31 s (250 Hz), 0.2 s (500 Hz), 0.43 s (1000 Hz), and 0.74 s (2000 Hz). The noise floor of the HEAD microphones averaged between 50 and 55 dB in this room for the measurement range used in these experiments. Source speakers included two, 100 W continuous, 400 W peak JBL floorstanding speakers, −3 dB frequency response 38 Hz–20 kHz, driven by a Halfer P4000 amplifier. Time-domain recordings of microphone signals are made at the ANR sample frequency (reported below) and processed through MathWorks signal processing software to obtain either narrowband or one-third octave frequency-domain results. In all results reported here, the manikin is placed in the near field of the speakers.

In this work, the FXLyLMS was not required to maintain stability during earcup ANR in the LFATC frequency band of interest, but it was required for earplug ANR, which is evaluated over a wider frequency band on the manikin. Thus, the hybrid architecture for the earcup incorporated the LyLMS filter [i.e., $\hat{S}(z)=1$], while for the earplug, the FXLyLMS filter was used with $\hat{S}(z)$ identified using *fastid*. For the earcup, noise sources selected for testing include (1) individual tones at 1/3-octave intervals from 40 Hz through 1250 Hz, (2) a sum-of-tones signal comprised of 1/3-octave tones 50–800 Hz, (3) an F-16 cockpit noise recording (Steeneken and Geirtsen 1988), and (4) a Huey helicopter noise recording (Steeneken and F. W. Geirtsen, 1988). Spectra for the F-16 and Huey sources are provided in Cartes (2000). Within the LFATC, all composite signals are band limited 50–800 Hz. For earplug manikin testing, sources include (1) sum-of-tones signal at one-third octave band center frequencies at 80–2000 Hz and (2) band-limited UH-60 Blackhawk cockpit noise, described below. For earcup testing within the LFATC, sources are set to levels between 105 and 110 dB to avoid cancellation speaker distortion and reference microphone saturation. For earplug testing on the manikin, source levels are set at 110–113 dB. All noise levels are reported in dB relative to a 20 μPa reference pressure. In addition, both manikin software and ANR software provide a digital dB meter with selectable response times to provide average C-weighted dB levels during ANR experiments.

The F-16 noise sample is similar to band-limited pink noise, but with significant temporal variation over a 2 min recording (Cartes *et al.*, 2002a). Huey helicopter noise also resembles pink noise, but with a 55 Hz tonal and associated harmonics, and impulsive staccato components in the time domain, which are harmonics of the 10.7 Hz blade frequency. The UH-60 noise sample available for testing was created from cockpit noise recorded on a Sony 4-channel digital tape recorder using a $\frac{1}{2}$ in. B&K free-field microphone. This 10 min recording was played back within a large reverberant chamber with Altec-Lansing amps and four loudspeakers, and rerecorded onto a CD, thus the low frequency spectrum is room dependent (Houtsma, 2005). The CD recording is used for work presented here. The CD recording was played back through our experimental facility and the one-third octave spectrum measured at the manikin ear was compared with that of the same segment from the original cockpit recording. This showed a low frequency roll-off at 100 Hz, and a +5 dB difference between 250 and 800 Hz. From 800 to 3150 Hz, the playback spectrum exceeded the cockpit spectrum by an average of 16 dB, indicating room-dependent amplification, which is confirmed by reverberation time measurements. The band-limited UH-60 noise exhibits a broadband spectrum, without the infrasonic blade passage fundamental, but with gear whine noise at 1000 Hz and 2000 Hz.

The hybrid controller is implemented on a dSPACE DS1103 controller board with 16-bit A/D and 14-bit D/A converters, at an update frequency of 10 kHz for the earcup and 15 kHz for the earplug. Bandpass antialiasing filters with a first-order roll-off below 10 Hz and a second-order roll-off above 3 kHz are present on the microphone channels. A bandpass anti-imaging filter with a first-order roll-off below 2.4 Hz and a third-order roll-off above 4.8 kHz is used for the speaker driver. In the LFATC, performance data are given as the difference between the B&K microphone outside of the hearing protector and the one inside in the base of the test cell, or as the difference between the reference and error ANR microphones, which are calibrated to B&K microphones. In manikin testing, both free-field and *in situ* calibration of the ANR microphones was performed, with no significant difference noted when the electret microphones

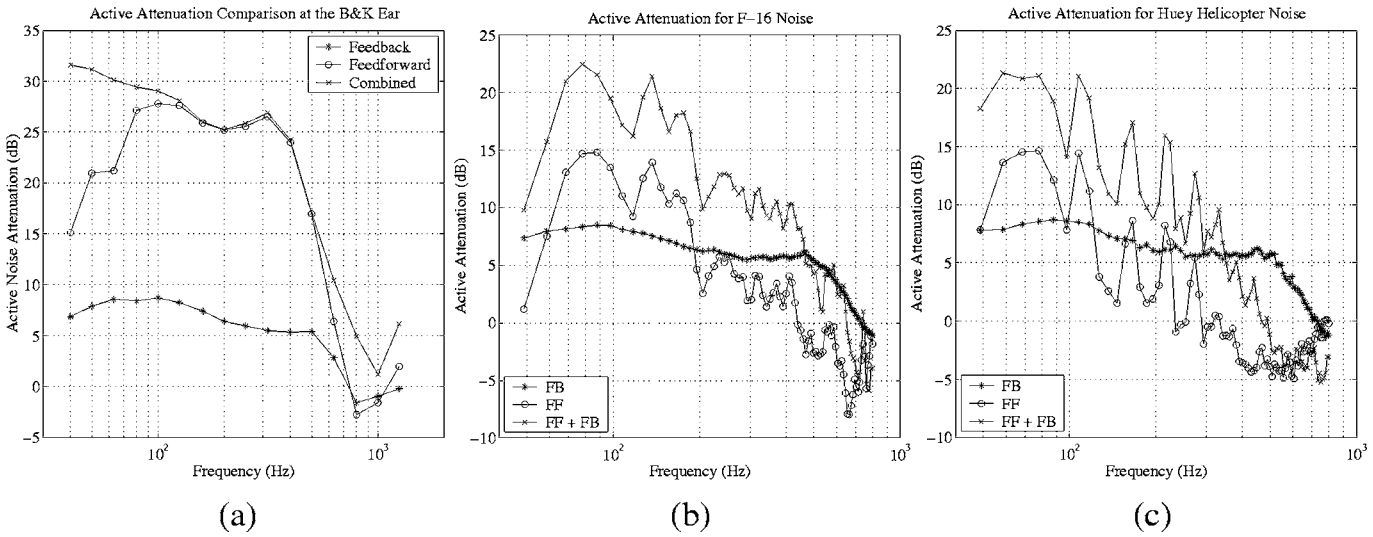


FIG. 3. Active attenuation performance of the ANR systems for circumaural earcup and (a) pure-tone noise, (b) F-16 cockpit noise, and (c) Huey helicopter noise.

were collocated with the standard. Manikin results are presented as insertion loss at the manikin ear microphone, or as the difference between the reference and error ANR microphones.

B. Earcup and earplug configurations

The hybrid ANR earcup was developed in-house from a commercial ANR earcup following the method described in Ray *et al.* (2002). During modification, the existing commercial ANR module housing the error microphone and cancellation speaker was retained, and the analog feedback system was bypassed. A reference microphone was added externally through a rubber grommet in the earcup shell. The passive attenuation of the earcup is 5 dB at 50 Hz, increasing to 30 dB at 800 Hz, matching the passive attenuation of the original, unmodified earcup (Ray *et al.*, 2002).

The earplug consists of a commercial, military communication earphone (CEP) modified for ANR. A Knowles FG-3329 microphone is inserted through a hole (formed with a hot wire) within the foam earplug tip, and a reference microphone, Panasonic WM-64PNT, is secured to the back of the earplug with plastic wire ties. The foam tip provides an average passive attenuation of 30 dB from 80 to 2000 Hz in manikin testing. Since earphone performance depends on the trapped volume of air, the cancellation path transfer function of the earplug was measured in the test cell, on the manikin, and on a human subject, as shown in Fig. 2(b). While neither the flat plate nor the manikin represent the precise characteristics of a human ear, the cancellation path transfer function shown in Fig. 2(b) for manikin insertion is qualitatively similar to that of the earplug when inserted into a human ear.

C. Cancellation path modeling and communication signal

Cancellation path modeling and communication feedthrough during ANR are evaluated using the earplug. For model identification, a set of identification data consisting of random excitation input to the earplug speaker and the resulting output from the internal error microphone is measured with no external sound field present. These data are used to find FIR, *fastid* IIR, and RLS IIR cancellation path models $\hat{S}(z)$ that fit the measured input-output relationship. The model orders and random excitation identification data length used for each candidate method are adjusted to achieve the same model fidelity, as measured by the communication signal prediction error. This metric is defined as follows. A sampled communication signal is sent to the cancellation speaker and is also filtered by each candidate identified model to provide an estimated error microphone response. This communication input-output data set is measured with no external sound field present. The communication signal prediction error is then computed as the standard deviation of the difference between the measured error microphone response and the estimated error microphone response for each identified model. Using this metric, the parameters for each identification method (e.g., model order and identification data length) are adjusted to achieve comparable model fidelity. Because the measured error microphone response is the same in each case, the prediction error between different modeling techniques can be directly compared using this metric.

Communication signals used in these experiments are pre-recorded male voices played back through a digital audio device at 44.1 kHz. The signal from the audio device is sampled by the dSPACE system at the ANR frequency. The sampled communication signal is filtered with $\hat{S}(z)$ in order to predict the error microphone response to the communication signal. The filtered communication signal is then subtracted from the sampled error microphone measurement, and the residual e_k is used for ANR computations in both the feedforward and feedback systems.

V. EXPERIMENTAL RESULTS

A. Evaluation of hybrid ANR for the circumaural earcup

Figure 3 shows the narrowband active attenuation for each noise source, measured by the transmission loss between the two LFATC B&K microphones. For individual

TABLE I. Overall C-weighted passive, active, and total noise reduction performance for feedback, feedforward, and hybrid ANR in earcup testing within a low frequency acoustic test cell.

	Noise level (dB)					Total attenuation (dB)			Active attenuation (dB)		
Noise source	Sourc	Passive	Feedback	Feedforward	Hybrid	Feedback	Feedforward	Hybrid	Feedback	Feedforward	Hybrid
Sum-of-tones 50–800 Hz	110–111	98	90	81.2	70–71	20	29	40	8	17	27
F-16 cockpit 50–800 Hz	110–111	95	88	85.6	78	23	25	32	8	10	17
Huey cockpit 50–800 Hz	105–106	94	86	83.7	75–76	19	22	30	8	11	18

tones, Fig. 3(a) shows that the feedback system provides low level (5–10 dB) noise reduction extending to 600 Hz. The feedforward LyLMS filter performs exceptionally well in the range 80–400 Hz, with diminished performance above and below that range. Whereas both the feedforward and feedback systems add noise above 700 Hz, the hybrid system provides positive active attenuation from 40 to 1250 Hz. Total (active plus passive) noise reduction is 36–51 dB within the 40–1250 Hz band, causing the error microphone signal to approach its noise floor, thus the noise reduction performance approaches its physical limit.

Figure 3(b) presents the narrowband active attenuation for F-16 noise. This noise source contains minimal tonal content and, over long time periods, has a fairly uniform spectral component. However, during short periods of time its spectral content shifts considerably, which presents problems for traditional LMS filters, as coherence may not be high during such periods. Here, the feedback gain K_{fb} is set first to provide at least 6 dB of gain margin, and then the feedforward gain is adjusted to maximize noise reduction performance. Despite the difficulties associated with this nonstationary noise source, the hybrid system provides an average active attenuation of 17 dB (32 dB total attenuation), reducing the 110 dB source level to 78 dB in the 50–800 Hz band. Below 500 Hz, the hybrid system has greater performance than either of the independent systems acting alone. Additionally, whereas the feedforward system adds noise above 500 Hz, the hybrid system largely avoids adding noise from 50 to 800 Hz.

Lastly, the system was subjected to Huey helicopter noise. This noise source contains broadband nonstationary components like the F-16 noise, tonal components following a 55 Hz fundamental attributed to the tail rotor, and a temporal component—the *thwt-thwt-thwt* of the blade passage. When band-limited 50–800 Hz, the temporal component sounds to the listener like a periodic broadband impulse, rather than a series of low-frequency blade passage harmonics (as in Winberg *et al.*, 1999). In order to keep this periodic impulse from forcing the ANR system to overdrive the cancellation speaker, the source level is reduced to 105 dB. The active attenuation results are shown in Fig. 3(c). Once again, the combination of feedback and feedforward ANR significantly improves low-frequency attenuation, in this case by 10 dB on average. The tonal components are attenuated by both the feedforward and hybrid system, but are largely untouched by the feedback system. Qualitative listener comparison of the source and attenuated noise shows that the feedback system is unsuccessful in removing the temporal component of the helicopter noise; the feedforward system could not completely remove it, either. In contrast, the hybrid system is able to almost completely remove the periodic *thwt*, leaving a 75–76 dB broadband residual. Table I summarizes these results, showing overall C-weighted source levels, passive, active, and total noise reduction performance for each noise source, including a sum-of-tones source.

The benefit of the hybrid ANR system for each source is that combining the two independent feedback and feedforward systems has resulted in low frequency performance that is greater than the sum of its parts. Without an FXLMS, the constant cancellation path gain in the feedforward model is erroneous, albeit within the 50–800 Hz band the gain error is 5–7 dB [Fig. 2(a)]. The synergistic effect is attributed to the fact that feedback control has the effect of "flattening" the cancellation path gain, i.e., raising the low frequency ($<$100 Hz) gain, and when both feedforward and feedback ANR are applied, the constant feedforward gain acting on the flatter cancellation path improves active performance. The effect is pronounced below 100 Hz, because the cancellation path gain rolls off at that frequency, as shown in Fig. 2(a), while at the mid-frequencies (125–500 Hz), the open-loop cancellation path gain is already relatively flat.

B. Comparison of cancellation path modeling approaches

In this section, the computational performance and model fidelity of three cancellation path modeling approaches for the earplug are evaluated—the *fastid* IIR method, a RLS IIR method, and a FIR filter. Figure 4 shows the experimentally determined frequency response of the input-output data and the identified models. Above 100 Hz, all models accurately reproduce the magnitude and phase characteristics of the cancellation path. For the *fastid* IIR filter, a model order of $p=10$ was adequate for excellent output prediction, with only 6.4% degradation in the prediction error metric from a much higher-order model ($p=60$). The model orders and identification data length used for the RLS and FIR candidates were adjusted to achieve the same model fidelity as the $p=10$ *fastid* model, as described in Sec. IV C. When matched for model fidelity in this manner, the resulting model orders for *fastid*, RLS, and FIR were 10, 10, and 45, respectively, requiring 21, 21, and 46 multiplies per sample, respectively for filter implementation. Data record lengths of 1500, 6000, and 1500 samples were required for conversion for the for *fastid*, RLS, and FIR methods, respectively. These results show that the *fastid* approach provides

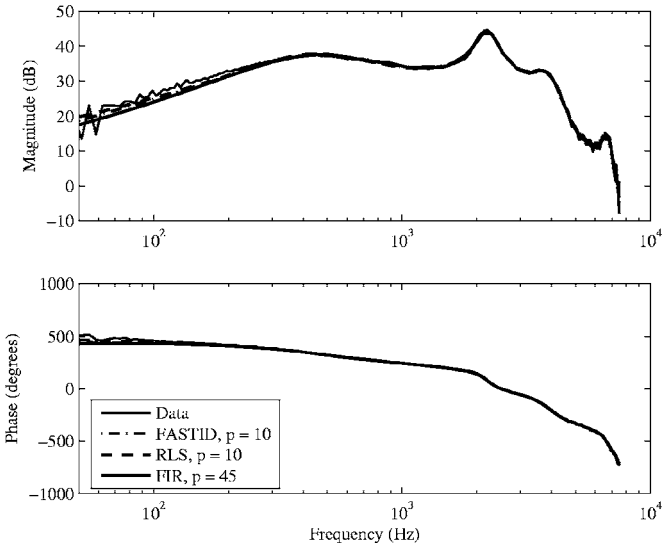


FIG. 4. Comparison of the frequency response function of identified models with transfer function estimate from identification data.

model fidelity that matches the more common RLS approach, using less data for convergence, and presenting computational improvements over the FIR filter model. The *fastid* method is used in all subsequent results.

In order to demonstrate the benefit of automated identification of the cancellation path response through an initialization routine, Fig. 5 compares the frequency responses of identified cancellation path models for two different deep insertions of the earplug and one shallow insertion of the earplug in the LFATC. These results demonstrate the possibility of variation in the cancellation path solely due to repeated insertions. Person-to-person variability could also introduce significant variation in the cancellation path response, e.g., due to individual ear canal geometry and compliance, as can component degradation.

C. Evaluation of hybrid ANR in the earplug

LFATC and manikin testing of the ANR earplug included evaluation of stability aspects of hybrid ANR and

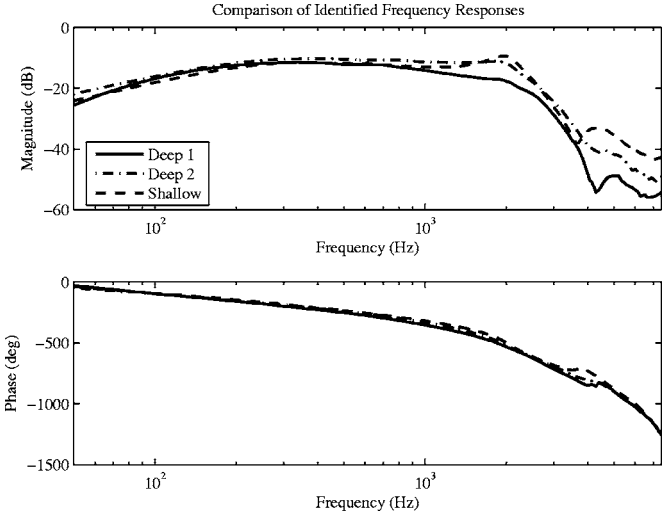


FIG. 5. Comparison of identified cancellation path model frequency responses for two different deep insertions and one shallow insertion of the ANR earplug in the LFATC.

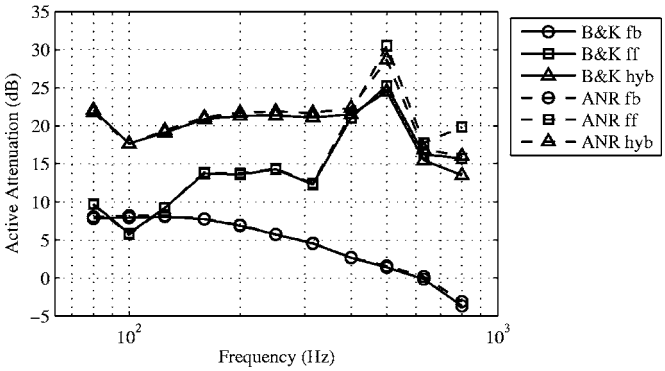


FIG. 6. ANR performance measured using the B&K microphones (solid) and the ANR microphones (dashed) for sum-of-tones 80–800 Hz in the LFATC evaluation.

ANR performance evaluation with and without a communication signal superimposed on the cancellation signal. The LFATC was used to measure the effect of a communication signal on ANR performance, as it allows for simultaneous measurement of source and error signals with precision B&K microphones, to which the ANR microphones can be calibrated *in situ*. It is necessary to first validate the use of the ANR microphones (instead of the LFATC B&K microphones) for ANR performance evaluation, because the B&K microphone measurements respond to both the external noise entering the earplug and the communication signal. All results are presented in one-third octave bands.

Figure 6 shows the performance of the ANR-modified CEP for sum-of-tones noise 80–800 Hz at 110 dB as measured with the B&K microphones in the LFATC and as measured with the ANR microphones. Figure 6 shows the minimal difference between performance as measured using the ANR reference and error microphones and the B&K microphones below 600 Hz. Above 600 Hz, the separation distance between the ANR and B&K error microphones affects the measured attenuation, resulting in an acceptable 2–5 dB difference. Now that it has been established that the calibrated ANR microphones are suitable for evaluation of ANR performance, the effect of the communication signal on ANR performance is measured. Figure 7 shows the performance of the ANR-modified CEP earplug for the same external sound field without and with a communication signal, demonstrating that the removal of the communication signal from the

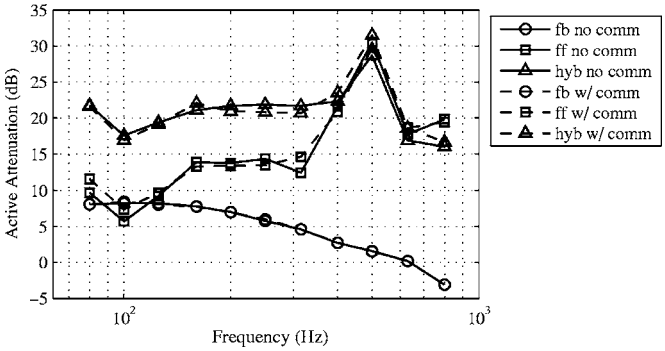


FIG. 7. ANR performance with no communications signal present (solid) and with a communications signal present (dashed) for sum-of-tones 80–800 Hz in the LFATC evaluation.

error signal has little effect on ANR performance within the LFATC. Total (hybrid+passive) performance, shown in Fig. 8, exceeds 40 dB at all frequencies from 80 to 800 Hz, and the hybrid ANR system drives the error signal and also the B&K microphone in the base of the test cell to within a few dB of their noise floors.

For manikin testing, insertion loss measurements are presented for sum-of-tones 80–2000 Hz and UH-60 cockpit noise. The external noise, after passive attenuation by the earplug, remains at least 5 dB above the noise floor at the manikin ear at all frequencies below 3000 Hz. ANR microphones are calibrated to the manikin ear and B&K microphones, respectively, to provide performance as measured at both the ANR microphone and manikin ear.

Figures 9(a) and 10(a) provide performance results for sum-of-tones 80–2000 Hz at 111 dB, measured for the unmodified CEP and for the ANR-modified CEP. These results provide the source level of the unoccluded ear, and of the occluded ear. The passive performance of the unmodified CEP ranges from 20 to 40 dB in the 80–2000 Hz band [Fig. 9(a)]. For the ANR-modified CEP, passive performance is within ±8 dB of the unmodified CEP, and with hybrid ANR, the 111 dB sum-of-tones source is reduced to 65–66 dB at the manikin ear and 59 dB at the ANR error microphone. ANR performance gains over passive attenuation of 2–20 dB are measured below 1000 Hz at the ear microphone, and the ear microphone is driven to its noise floor by 2000 Hz.

Figures 9(b) and 10(b) show performance results for UH-60 cockpit noise measured for the unmodified CEP and for the ANR-modified CEP. Passive attenuation for the unmodified CEP is consistent with Fig. 10(a). For the ANR-modified CEP, passive results are similar to those reported for sum-of-tones and thus are omitted. With hybrid ANR active, the 110–112 dB source level is reduced to 68–70 dB as measured at the ear microphone [Fig. 10(b)]. Performance is 2–5 dB better at the error microphone, with a modest 5 dB difference between the two microphones at 2000 Hz.

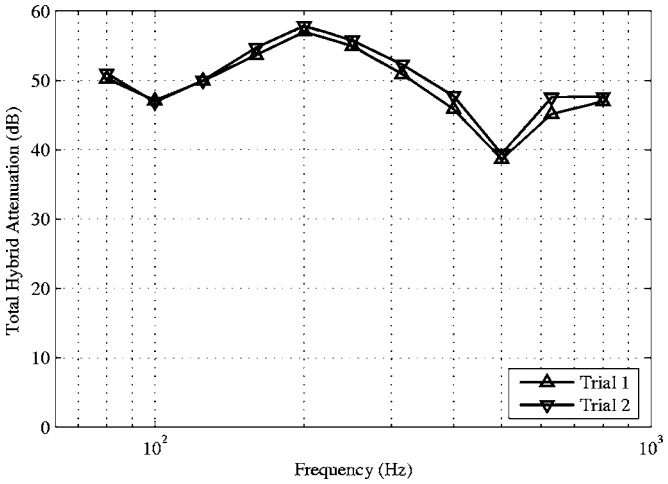


FIG. 8. Total attenuation (passive+active) of sum-of-tones 80–800 Hz for hybrid ANR as measured by transmission loss between the B&K microphones in the LFATC.

The net result is a total attenuation of UH-60 noise of 40–41 dB by combined active and passive means, as measured at the manikin ear microphone.

Figure 11 shows total attenuation measured at the ear and error microphones for each source. Figure 11 also shows ANR performance measured at the error microphone when the communication signal is injected at a 75 dB average level, as measured by a digital dB meter with no external noise source present. This level provides audible speech within the background noise both for the passive earplug and with hybrid ANR active. A summary of total attenuation for sum-of-tones 80–2000 Hz and for the band-limited UH-60 noise is provided in Table II as measured at the manikin ear microphone, and as measured at the error microphone, with and without the communication signal injected. Average attenuation provided by the unmodified CEP in the same frequency band is provided for comparison.

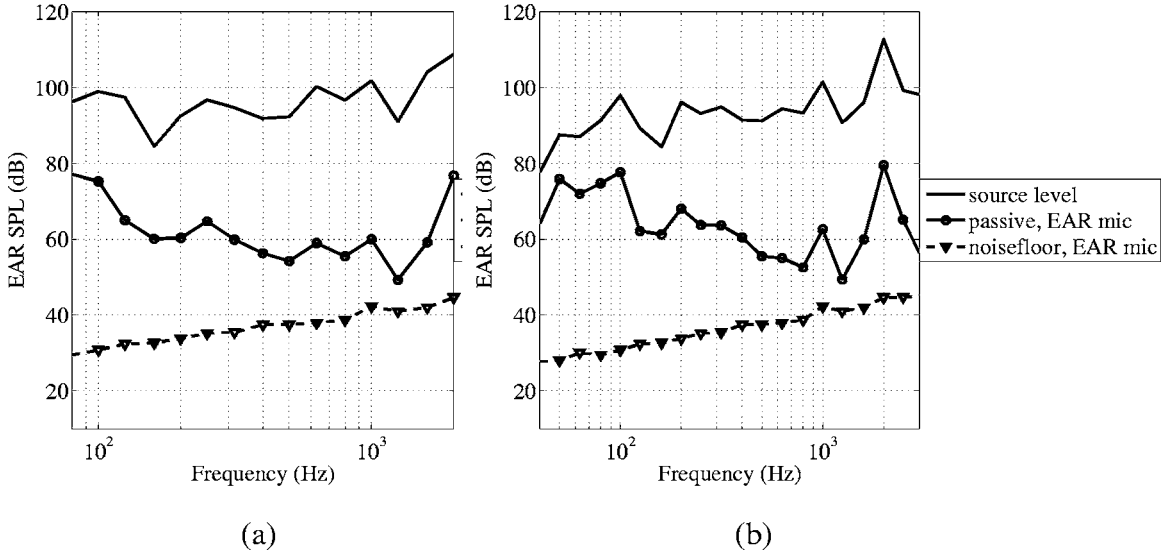


FIG. 9. Sound level measured at unoccluded manikin ear and unmodified CEP occluded manikin ear for (a) sum of tones noise 80–2000 Hz, (b) UH-60 cockpit noise.

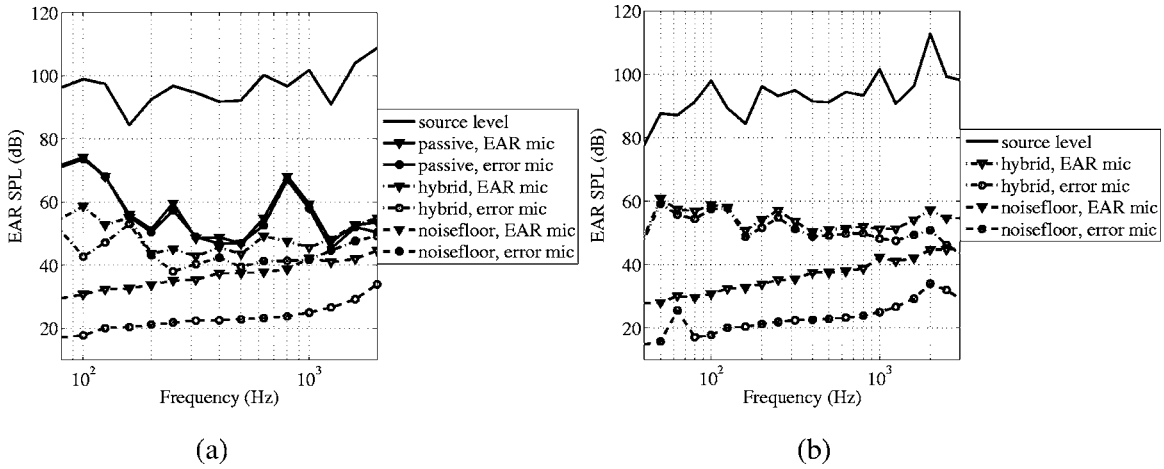


FIG. 10. Sound level at unoccluded manikin ear and after hybrid ANR for ANR-modified CEP and (a) sum-of-tones 80–2000 Hz, and (b) UH-60 cockpit noise. Performance is measured at both the manikin microphone and ANR microphones. Passive levels are provided for sum-of-tones.

D. Hybrid system stability

Hybrid ANR not only influences active performance, but it also improves the gain margin of the individual systems. In the feedback system, increasing the loop gain, K_{fb}, generally increases the feedback attenuation. However, the gain that maximizes noise reduction is close to the threshold of instability, forcing a stability-performance tradeoff. In all hybrid ANR results, the feedback gain is first set to provide a minimum 6 dB gain margin, and then feedforward ANR is enabled (without filtered-X for the earcup and with filtered-X for the earplug). The feedforward gain is then set to maximize performance. The dSPACE implementation offers an environment in which thresholds of instability can be evaluated quantitatively and experimentally *in situ*. For example, just as there is a maximum value of K_{fb} which if exceeded causes instability, there also is a maximum feedforward gain, K_{ff}, above which the weight vector $W(z)$ can grow without bound, or overexcite certain frequencies. While such behavior is diminished with the FXLyLMS, hybrid ANR has a profound stability effect in that when the two systems are combined, both gains can be increased to levels that otherwise would cause instability. When this happens, the increased gain allows for higher overall active attenuation. Alternatively, the hybrid system can fix the attenuation at the same gain values as before, but with larger gain stability margins.

For the feedback system, adding the feedforward system allows K_{fb} to increase by approximately 20% before instability reoccurs, i.e., it provides a 20% increase in gain margin. However, the increased stability is most notable in the feedforward system. Figure 12 shows the ratio between the maximum stable (not necessarily optimal) feedforward gain K_{ff} of the hybrid system and of the feedforward system alone, as determined experimentally as a function of frequency. A system is deemed stable if the weight vector converges and the ANR system does not overexcite certain frequencies so as to cause the controlled system to make more noise than the uncontrolled system, i.e., both asymptotic and marginal stability is considered. For the hybrid results, the feedback gain is set to provide a 6 dB gain margin when acting alone, and then the feedforward gain is set and incrementally increased until the point of instability is noted. Results are presented

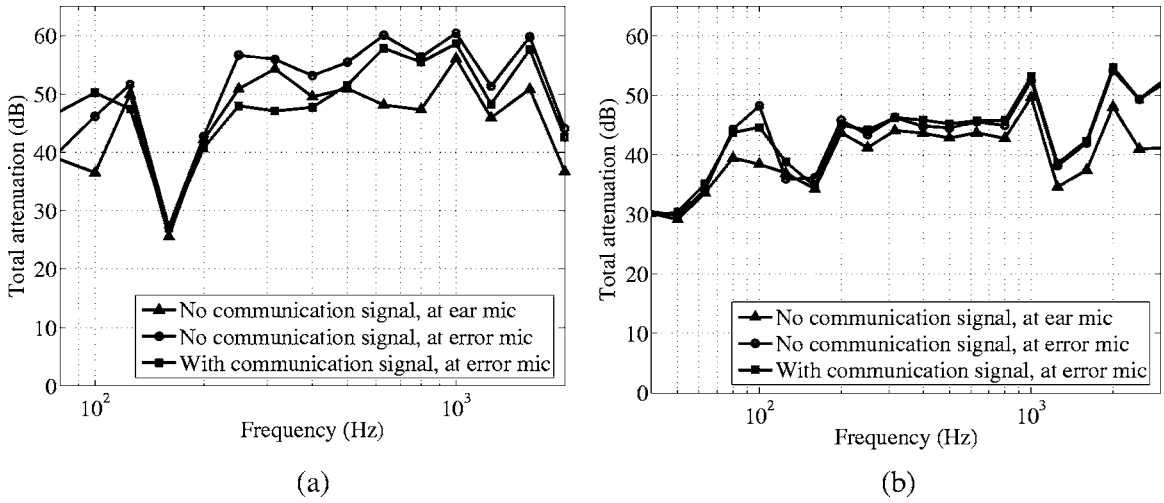


FIG. 11. Total attenuation at error microphone with no communication signal present, and with a communication signal present. Total attenuation also reported at manikin ear microphone. (a) Sum-of-tones 80–2000 Hz, (b) UH-60 cockpit noise.

TABLE II. Overall C-weighted total attenuation of sum-of-tones 80–2000 Hz and UH-60 cockpit noise in the 63–3000 Hz band, both at 110–112 dB source levels.

Noise source	Location	Total attenuation (dB) ANR-modified CEP	Total attenuation (dB) Unmodified CEP
Sum-of-tones 80–2000 Hz	Manikin ear microphone	46–48	34
	Error microphone	50–52	···
	Error microphone (with comm. signal)	48–49	···
Band-limited UH-60 cockpit noise 63–3000 Hz	Manikin ear microphone	40–41	31–32
	Error microphone	45	···
	Error microphone (with comm. signal)	45	···

for both the earcup and ANR earplug; a larger ratio indicates a larger gain stability margin. As Fig. 12 shows, augmenting the hybrid system allows the maximum stable K_{ff} to be increased by, at some frequencies, orders of magnitude. The effect is more profound for the earcup, in which the filtered-X implementation is not used, than for the earplug, which incorporates a reasonable, albeit fixed cancellation path model $\hat{S}(z)$. The gain stability margin depicted in Fig. 12 provides a measure of stability robustness to variations in $\hat{S}(z)$ and the actual cancellation path characteristics.

VI. CONCLUSION

A hybrid ANR architecture, comprised of a broadband digital feedback compensator and a Lyapunov-tuned leaky LMS filter is presented and evaluated for two hearing protection devices—a commercial circumaural earcup and a commercial communication earplug, both modified for feedforward ANR. Flat plate evaluation of the earcup from 50 to 800 Hz shows total performance ranging from 30 dB for Huey helicopter noise to 40 dB for sum-of-tone noise. For the ANR earplug, total performance is 46–48 dB for sum-of-tone noise 80–2000 Hz and 40–41 dB for band-limited UH-60 cockpit noise. A fast identification method is presented for automatic identification of an IIR cancellation path model. The identified model is used for both a filtered-X implementation of the LyLMS filter and for communication feedthrough. Communication feedthrough in the presence of ANR is shown to cause minimal degradation of ANR performance. For both the earcup and earplug, gain stability margins increase through hybrid ANR, providing a measure of robustness to variations in the cancellation path transfer function.

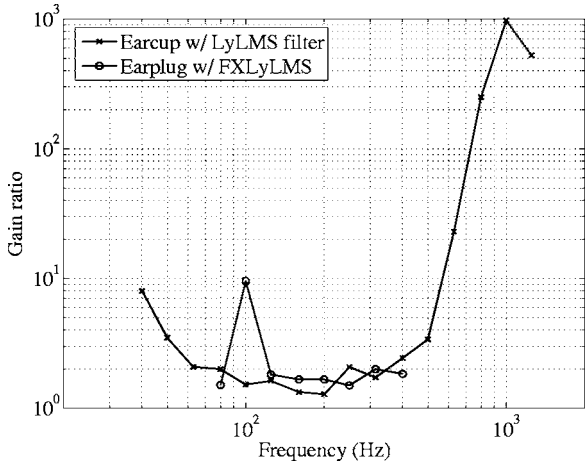


FIG. 12. Ratio of maximum stable gain K_{ff} of the hybrid system and of the feedforward system alone when the feedback gain K_{fb} is set to provide a 6 dB gain margin.

ACKNOWLEDGMENTS

This research was supported by the AFOSR Defense University Research Instrumentation Program (Award No. F49620-03-1-0248), the David Clark Company, Inc., and the U.S. Army Medical Research and Materiel Command Contract No. W81XWH-05-C-0031.

Akelsson, A. (**1996**). "Scientific Basis of Noise-Induced Hearing Loss," in Proceedings of the 5th International Symposium on the Effects of Noise on Hearing, May 1994 (Thiem Medical).

Brammer, A. J., and Pan, G. J. (**1998**). "Opportunities for active noise control in communication headsets," Can. Acoust. **26**, 32–33.

Cartes, D. (**2000**). "Lyapunov tuning and optimization of feedforward noise reduction for single-point, single-source cancellation," Doctoral thesis, Thayer School of Engineering, Dartmouth College.

Cartes, D., Ray, L. R., and Collier, R. D. (**2002a**). "Experimental evaluation of leaky least-mean-square algorithms for active noise reduction in communication headsets," J. Acoust. Soc. Am. **111**, 1758–1772.

Cartes, D., Ray, L. R., and Collier, R. D. (**2002b**). "Low frequency acoustic test cell for the evaluation of circumaural headsets and hearing protection," Can. Acoust., **30**, 13–20.

Cartes, D., Ray, L. R., and Collier, R. D. (**2003**). "Lyapunov tuning of the leaky LMS algorithm for single-source, single-point noise cancellation," Mech. Syst. Signal Process. **17**(5), 925–944.

Collier, R. D., Kaliski, K. H., and Ray, L. R. (**2003**). "Experimental techniques for evaluation of active noise reduction communication headsets with the thayer low frequency acoustic," in Proceedings of NOISECON03, on CD, Cleveland, Ohio.

Houtsma, A. (**2005**). U.S. Army Medical Research and Materiel Command (private communication).

International Telecommunications Union (ITU) Recommendation P.57 Artificial Ears (**2002**).

International Telecommunications Union (ITU) Recommendation P.58, Head and Torso Simulator for Telephonometry (**1998**).

International Electrotechnical Commission, IEC 711, Occluded-ear simulator for the measurement of earphones coupled to the ear by ear inserts (**1981**).

Gitlin, R. P., Meadows, H. C., and Weinstein, S. B. (**1982**). "The tap-leakage algorithm: An algorithm for the stable operation of a digitally implemented fractionally spaced adaptive equalizer," Bell Syst. Tech. J. **61**, 1817–1839.

Juang, J.-N. (**1994**). *Applied System Identification* (PTR Prentice-Hall, Englewood Cliffs).

Kuo, S. M., and Morgan, D. R. (**1996**). *Active Noise Control Systems: Al-*

gorithms and DSP implementations (Wiley, New York).
McKinley, R. L., and Nixon, C. W. (**1993**). "Active noise reduction headsets," in Proceedings of the 6th International Conference on Noise as a Public Health Hazard, Nice, Vol. **2**, pp. 83–86.
Ogata, K. (**1990**). *Modern Control Engineering* (Prentice-Hall, Englewood Cliffs).
Pan, G. J., Brammer, A. J., and Crabtree, R. B. (**1997**). "Adaptive feedforward active noise reduction headset for low-frequency noise," in Proceedings of the Symposium on Active Control of Sound and Vibration, edited by Elliott and Horvath.
Pan, G. J., Brammer, A. J., Goubran, R., Ryan, J. G., and Zera, J. (**1994**). "Broad-band active noise reduction in communication headsets," Can. Acoust. **22**, 113–114.
Phan, M. Q., Solbeck, J. A., and Ray, L. R. (**2004**). "A direct method for state-space model and observer/Kalman filter gain identification," AIAA Guidance, Navigation, and Control Conference, Providence RI.
Rafaely, B., and Jones, M. (**2002**). "Combined feedback-feedforward active noise-reducing headset—The effect of the acoustics on broadband performance," J. Acoust. Soc. Am. **112**, 981–989.
Ray, L. R., Collier, R. D., and Kaliski, K. H. (**2002**). "Optimization of stability and performance of LMS filters for feedforward active noise reduction in communication headsets," ACTIVE02-Symposium on Active Control of Sound and Vibration, pp. 705–715.
Steeneken, H. J., and Geirtsen, F. W. (**1988**). Description of the RSG-10 Noise Database, TNO Human Factors Institute, Soesterberg, The Netherlands.
Ward, W. D. (**1997**). "Effects of high intensity sound," *Encyclopedia of Acoustics*, edited by Malcolm J. Crocker (Wiley, New York), Chap. 119, pp. 1495–1507.
Winberg, M., Johansson, S., Lagö, T. L., and Claesson, I. (**1999**). "A new passive/active hybrid headset for a helicopter application," Int. J. Acoust. Vib. **4**, 51–58.

The relationship between railway noise and community annoyance in Korea

Changwoo Lim[a)]
Center for Environmental Noise and Vibration Research, School of Mechanical and Aerospace Engineering, Seoul National University, Rm 205 Bldg 44, Seoul, 151-744, South Korea

Jaehwan Kim[b)]
School of Mechanical and Aerospace Engineering, Seoul National University, Rm 205 Bldg 44, Seoul, 151-744, South Korea

Jiyoung Hong[c)]
School of Mechanical and Aerospace Engineering, Seoul National University, Rm 205 Bldg 44, Seoul, 151-744, South Korea

Soogab Lee[d)]
School of Mechanical and Aerospace Engineering, Seoul National University, Rm 1303 Bldg 301, Seoul National University, Seoul, 151-744, South Korea

(Received 21 September 2005; revised 10 July 2006; accepted 12 July 2006)

A study of community annoyance caused by exposures to railway noise was carried out in 18 areas along railway lines to accumulate social survey data and assess the relationship between railway noise levels and annoyance responses in Korea. Railway noise levels were measured with portable sound-level meters. Social surveys were administered to people living within 50 m of noise measurement sites. A questionnaire contained demographic factors, degree of noise annoyance, interference with daily activities, and health-related symptoms. The question relating to noise annoyance was answered on an 11-point numerical scale. The randomly selected respondents, who were aged between 18 to 70 years of age, completed the questionnaire independently. In total, 726 respondents participated in social surveys. Taking into consideration the urban structure and layout of the residential areas of Korea, Japan, and Europe, one can assume that the annoyance responses caused by the railway noise in this study will be similar to those found in Japan, which are considerably more severe than those found in European countries. This study showed that one of the most important factors contributing to the difference in the annoyance responses between Korea and Europe is the distance between railways and houses.
[DOI: 10.1121/1.2266539]

PACS number(s): 43.50.Qp, 43.50.Sr, 43.50.Rq [DKW] Pages: 2037–2042

I. INTRODUCTION

Environmental noise pollution due to transportation noise continues to grow and has become a serious problem in many countries.[1] This problem is difficult to regulate because it involves direct and cumulative adverse effects of noise on health. In recent years, therefore, the percentage of respondents who felt "highly annoyed" has become a critical component of environmental impact analyses to support environmental decisions regarding transportation noise.

Noise annoyance produced many responses characteristic of psychological stress.[2–4] Annoyance reactions are sensitive not only to acoustical characteristics (source, noise level), but also to many nonacoustical factors such as social, psychological or economic nature.[5,6] There are considerable differences in individual reactions to the same noise.[7] Therefore, social surveys on transportation noise have been performed in many countries over the past 40 years, from which dosage-response relationships for transportation noise have been evaluated.[7–12]

A majority of studies in European countries reported that railway noise causes less annoyance than other transportation noise sources.[8,13–16] This is a so-called "railway bonus" in noise regulations of some European countries. However, recent Japanese studies have produced very different results.[17–20] They have shown that no railway bonus existed and that railway noise annoyance was nearly the same as or even a little higher than road traffic noise annoyance.

Although many social surveys on the effects of railway noise have been performed throughout the world, they have been carried out mainly in developed countries. Even with similar noise levels and sources, the results of the annoyance responses differ from country to country, because annoyance responses to railway noise are affected by several external factors including cultural differences, languages differences, variations in survey questions, and differences in climatic conditions.[2,16] Therefore, the objective of this paper is to

[a)]Electronic mail: glider20@snu.ac.kr
[b)]Electronic mail: kjh03@snu.ac.kr
[c)]Electronic mail: hongjy@snu.ac.kr
[d)]Author to whom correspondence should be addressed. Electronic mail: solee@plaza.snu.ac.kr

TABLE I. Distance data.

Distance (m)	$d<20$	$20<d<40$	$40<d<100$	$100<d<200$	$d>200$
Number of sites	3	4	7	2	2
Percentile (%)	16.7	22.2	38.9	11.1	11.1

accumulate the social survey data to assess the relationship between railway noise levels and annoyance responses in Korea, and to estimate the applicability of a railway bonus in Korea.

II. METHOD

The most common method of assessment for human response to railway noise is the combination of a field survey that consists of physical measurements and social surveys using a questionnaire. Noise measurements and social surveys were carried out simultaneously.

A. Noise Measurement

1. Site selection

Due to the high population density in Korea, a number of houses are situated close to railway lines and railways pass through the middle of several cities.

Field surveys were performed in 18 areas along Gyungbu and Honam railway lines in Korea. These sites were chosen based on the fact that they have high volumes of train operations that consist of heavy freights and passenger trains that use a diesel engine. The two lines are responsible for more than 60% of the passenger and freight transports in the whole railroad industry.

Most of the houses in the field survey areas are apartment buildings built out of ferroconcretes. Table I shows the distances between the railway lines and the survey sites in this study. The average distance was 90 m, but about 80% of the sites were situated within 100 m from the railway lines. Figure 1 shows an example of some of the selected sites in which field surveys were carried out. As shown in this figure, distances of points 1 through 5 are about 20, 60, 27, 80, and 85 m, respectively. Measurement sites that were chosen were also flat and free of obstacles.

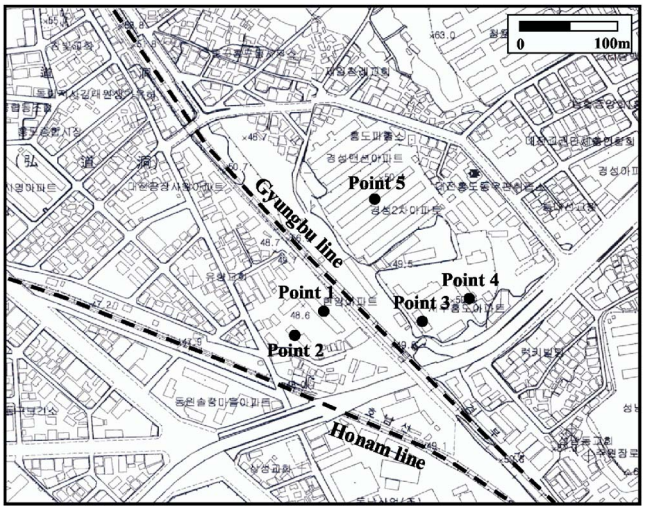


FIG. 1. Some selected sites in which field surveys were carried out.

2. Noise measurement

Table II shows information regarding the details of the train operations of Gyungbu and Honam railway lines. Noise levels were calculated around the two railway lines with different volumes of train operations. The average number of train operations of Gyungbu line is about 250 a day and that of Honam line is about 51 a day.

Railway noise levels were measured with portable sound-level meters (B&K type 2238 and LD 812) at 18 sites. The equipment was mounted on a tripod on the rooftop of houses to avoid obstacles between the railway and the receiver. The microphone was positioned at a height of 1.5 m above the flat, and at least 1 m from any other reflecting surfaces.

It was necessary to carry out extensive measurements, in order to calculate the railway noise levels. So, measurements were taken for three successive days in June 2004. The relative humidity and temperature of the sites varied from 62.5% to 64.6% and 18 °C to 23 °C, respectively, at the time of measurements.

To analyze the relationship between railway noise levels and annoyance responses, day-night noise level L_{dn} was calculated. The day-night noise level L_{dn} was calculated from the formula.[21]

$$L_{\mathrm{dn}} = 10 \log\left[\frac{15}{24} \times 10^{0.1\times L_{\mathrm{day}}} + \frac{9}{24} \times 10^{0.1\times (L_{\mathrm{night}}+10)}\right], \tag{1}$$

where L_{day} and L_{night} represent the day and night-time average sound levels which were calculated from A-weighted sound exposure levels L_{AE} observed for every passing train. The day-time period was defined as 07:00 to 22:00 and the night-time period was defined as 22:00 to 07:00.

B. Social survey

Subjective responses to railway noise were measured by means of a social survey using a questionnaire. The survey

TABLE II. Information on the train operations.

	Type of trains (diesel)		Number of trains per day	
	Passenger	Freight	Day time	Night time
Gyungbu line	152	98	178	72
Honam line	32	19	41	10

was performed in order to investigate the individual's attitude and opinion in regard to different aspects of the railway noise, and it was administered to residents within about 50 m of field survey sites. Therefore, one can assumed that all of the respondents were exposed to similar railway noise levels.

Questionnaires were comprised of questions relating to the assessments of railway noise as well as some general questions about the residents, even if they do not relate to noise. Questions were arranged in three basic sections. The first section sought to obtain demographic data, the second asked questions about the nuisance perceptions of railway noise and vibration, and the third dealt with questions regarding health-related symptoms. Therefore, the questionnaire contained demographic questions, degree of noise annoyance, interferences with daily activities, perception of vibration, psychological and physiological health-related symptoms, and reaction to railway noise. In order to assess the annoyance responses to railway noise, specifically, people were asked questions like "how much were you bothered or annoyed by the railway noise, while staying at home, in the last 12 months,"[22] by selecting one of 11 categories ranging from 0 (not at all annoyed) to 10 (extremely annoyed). The 11-point numerical scale was chosen based on the assumption that respondents are more cognitively familiar with a 0 to 10 scaling than the shorter 7- or 9-point numeric scales.[23]

To avoid any bias in opinion, the surveys were not introduced to the interviewees in advance and respondents were randomly selected from residents near the survey sites based on simple random sampling method. Questionnaires were distributed in person and respondents completed the questionnaire independently while researchers waited. Each questionnaire took about 20 min to complete. They were administered concurrently with the noise measurements at each site. 61.7% of the randomly selected respondents participated in the surveys, resulting in a total of 724 respondents for the analysis of exposure-effect relationships between railway noise levels and annoyance responses.

III. RESULTS

The ages of respondents exhibit a wide range: younger than 20 years (4%), 20–40 (52%), 40–60 (32%), and older than 60 years (12%). Most of the respondents were female (76%) and were married (85%). These results were due to the nature of the Korean culture where most females become housewives after marriage. The duration of residency of the respondents was as follows: less than 1 year (8%), 1–3 years (24%), 3–10 years (41%), and 10–30 years (27%). Only the responses of respondents who had resided in the area for more than 1 year were analyzed for the purposes of this study.

Annoyance responses to railway noise were elicited by means of an 11-point numerical scale. Under the definition of the annoyance scale, the term "highly annoyed" was defined as the upper 27–28 % of the annoyance scale. Therefore, the "highly annoyed" variable of annoyance responses was calculated as a binary datum. This means, the equal variance assumption and the assumption that responses vary about the mean according to a normal distribution are not valid due to the binary nature of the data. In cases where the variable is binary, logistic regression analysis is more reliable.[24,25] The data, therefore, had to be dichotomized to conduct a logistic model. The numerical scale of annoyance response was dichotomized with the responses in the top three out of 11 categories (3/11) being defined as "highly annoyed" and the remaining are not. The responses were bounded between zero and one. The logistic model can be expressed as follows:

$$E(Y_i/X_i) = \frac{e^{(\beta_0+\beta_1 X_i)}}{1+e^{(\beta_0+\beta_1 X_i)}}, \tag{2}$$

where, β_0 and β_1 are the intercept and the slope of the logistic response function.

In the case of this study, maximum likelihood estimation (MLE) was used to dispose of the assumption problems mentioned earlier, and to estimate the parameters of a logistic model.[25,26]

When assessing the effects of noise on health, the percentage of respondents who felt highly annoyed (%HA) are recommended as the indicator of noise annoyance and the day-night average sound level (L_{dn}) is selected as the uniform metric for the description of noise in many countries, such as the European Union, North America, and Australia. Therefore, %HA and L_{dn} have been used to assess the effects of railway noise on health in terms of dose-response relationship between railway noise levels and annoyance responses.

Figure 2 shows the %HA prediction curve of railway noise in this study. Square spots show the percentage of respondents who felt "highly annoyed" as a function of L_{dn}. Bars represent the 95% confidence intervals at each data point. The 95% confidence intervals were calculated to estimate the distribution of "highly annoyed" respondents at each field survey site. The levels of railway noise exposures

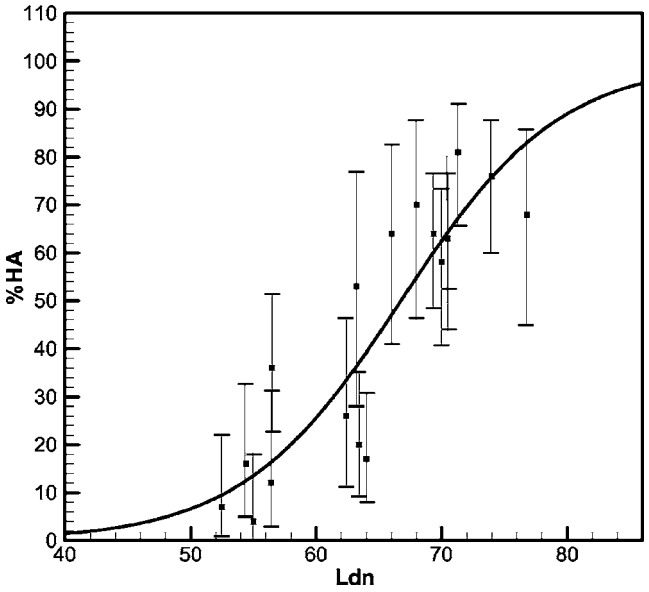


FIG. 2. Prediction curve for the percentage of highly annoyed respondents (%HA) based on noise exposure at the dwelling. Solid line is %HA prediction curve. Points are field survey data in 18 areas. Bars are 95% confidence intervals for the data point. N=613.

TABLE III. Estimated coefficients for the logistic equation using L_{dn} as the noise exposure metric.

Parameter	Estimate	Std. error	P Value
β_0	−10.547	1.028	<0.0001
β_1	0.158	.015	<0.0001

range from 52 to 76. The solid line is the %HA prediction curve that was determined by logistic fit procedure based on field survey data. The estimates of coefficients β_0, β_1 are presented in Table III with their estimated standard errors and significance levels. As shown in this table, the significance of p value is less than 0.01, meaning the parameters of this model are significantly effective.

A next step estimates the measure of fit of the established logistic model. As the criterion of an explanatory power, the coefficient of determination R^2 is used in linear regression models. Then again, the correct classification rate (CCR) is generally considered to estimate the measure of fit of logistic models. In this model, total CCR is 72.3. It shows a good relationship between railway noise levels and the percentages of respondents feeling "highly annoyed." As shown in Fig. 2, it is found that with an increase of L_{dn}, the percentage of respondents who felt highly annoyed also increased.

IV. DISCUSSION

In order to investigate the community response to railway noise, 18 areas were chosen and field surveys were carried out at each area. Then, the dose responses conducted in this study were compared with those of other countries to examine whether or not the annoyance responses to railway noise were equivalent among countries.

Figure 3 indicates comparison between the noise annoyance curve in this study and the one in the European survey.[8] Square spots are each field survey data showing %HA with respect to L_{dn}. The solid line is the %HA prediction curve of this study by logistic fit procedure. The dashed line is the one of Miedema and Vos by polynominal fit.[8] As shown in this figure, the results are different. Of course, we know that surveys can differ from one to another due to many factors, such as cultural differences, languages differences, different phrasings of the annoyance questions, and differences in climatic conditions.[16] The results of this study, however, are much more severe than those of European countries even after taking those external factors into account. A number of studies in foreign countries showed that noise annoyance from railway noise causes less annoyance than other transportation noise.[8,13–16] This is called a "railway bonus" in European countries. While there is no scientific evidence to suggest why respondents feel that railway noise is less annoying than other transportation noise, some researchers believe that some sentimental feelings about railways are influential.[17,27] Railways are often considered socially more acceptable than other types of transportation because of safety, economy, and convenience.[16] However, recent Japanese studies have produced different results.[17–20] Railway noise annoyance in Japan was much higher than in European countries. Figure 4 shows the comparisons of annoyance curves with respect to $L_{Aeq,24h}$ which was calculated from an average of A-weighted sound exposure level (L_{AE}) observed for every passing train in a Japanese study[28] and three European surveys.[29] The solid line is the %HA prediction curve of this study based on field survey data. Dash-dotted line (a) is the %HA prediction curve of conventional railways in Japanese survey. Dashed lines (b), (c), and (d) are the %HA prediction curves of French, Danish, and British surveys, respectively. Figure 4 shows that the result of this study is not only similar to that of the survey in Japan, but more severe when compared to those in European countries. The distance between the railway and the house may be an important

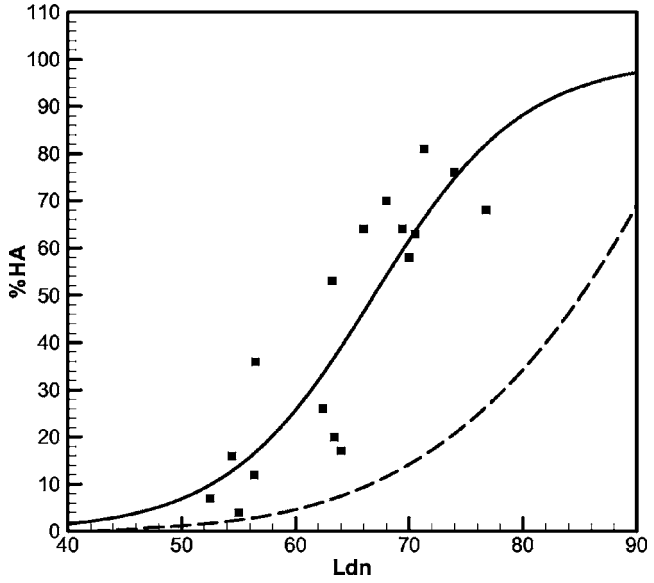


FIG. 3. Comparison between the %HA prediction curve of railway noise in this study and that in a European country. [■, field survey data with respect to L_{dn} in this study; ——, %HA prediction curve in this study; - -, the Miedema and Vos %HA prediction curve (Ref. 8).]

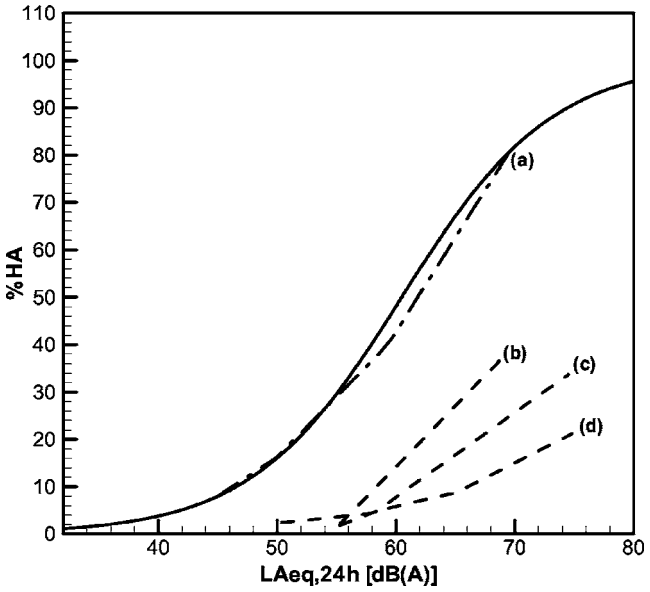


FIG. 4. Comparison between the %HA prediction curve of railway noise in this study and those in other country surveys. [——, %HA prediction curve in this study; (a) Japan 1992 (Ref. 28); (b) France 1988 (Ref. 29); (c) Denmark 1988 (Ref. 29); (d) U.K. 1984 (Ref. 29).]

TABLE IV. Measurement data at sites within 50 m of railway lines when a train passes by.

Measurement point	$L_{A\max}$	Measurement point	$L_{A\max}$
Point 1	93.1	Point 6	100.5
Point 2	93.7	Point 7	99.9
Point 3	97.6	Point 8	92.8
Point 4	94.1	Point 9	96.1
Point 5	91.6		

cause of the difference in annoyance responses between Korea and European countries. A number of houses in Korea are situated closer to railway lines than those in European countries due to high population density. Therefore, noise and vibration levels caused by train passages are usually higher than those of European countries. For example, Table IV shows the maximum noise level ($L_{A\max}$) at sites within about 50 m of railway lines when trains pass by. As shown in this table, all of them exceed 90 dB(A) and Point 6 even exceeds 100 dB(A). These values are too high to lead a peaceful life, especially when nighttime exposure to heavy freight trains occurs.

The position of the balcony also has a significant effect on general annoyance responses. Respondents who lived in apartments with balcony windows facing a railway may be more annoyed by the noise than those having balcony windows not facing in the railway direction.[20] In Korea, many apartments have balconies that face the railway.

Attitudes toward railway noise could also have an influence on annoyance responses since the noise could cause social and economic costs, such as property value depreciation. In Korea, the price of apartments in areas close to railway lines is lower than those in noise free areas. The increase of 1 dB noise affects the price of an apartment by about 0.3%.[30]

For these reasons, Korean people living close to railways in this study may feel more annoyed than European people. There is a similarity between the results of the Japanese survey and those of this study because conditions, such as high population density in metropolitan areas and distances from railways to houses, in Japan are similar to those in Korea.

Another factor contributing to our findings may be the inconsistent nature of railway noise. About 15% of the respondents said that they have been surprised by very loud and unexpected railway noise every day. They also considered the railway noise as having negative effects on their health, and complained that the exposure to railway noise caused insomnia, nervousness, and indigestion.

V. CONCLUSIONS

Environmental noise pollution due to transportation noise continues to grow and has become a serious problem in many countries. This problem is difficult to regulate because it involves direct and cumulative adverse effects of noise on health. In recent years, therefore, the percentage of respondents who felt "highly annoyed" has become a critical component of environmental impact analyses to support environmental decisions regarding transportation noise. The World Health Organization (WHO) has recommended annoyance as one of the environmental health indicators to support environmental noise policy-making activity in many countries. However, WHO does not recommend an international consensus on how to predict annoyance from transportation noise sources. Therefore, this study of community annoyance caused by railway noise exposures was carried out to accumulate social survey data and to assess the relationship between railway noise levels and annoyance responses in Korea. Noise measurements were carried out in 18 areas along Gyungbu and Honam railway lines in Korea. Social surveys were administered to residents living within 50 m of the noise measurement sites. The total number of respondents for the social surveys was 726.

It can be concluded that the community annoyance of railway noise in this study is similar to that found in Japan but more severe than that found in European countries. The cause of the difference can be ascribed to the distance between railways and houses, the position of the balcony, and attitudes of the residents toward the source of the noise. Based on these results, we claim that a railway bonus should not be applied to railway noise guidelines in Korea.

ACKNOWLEDGMENT

This work was sponsored by Core Environmental Technology Development Project for Next Generation in Korea Institute of Environmental Science and Technology.

[1]WHO, Guidelines for Community Noise, World Health Organization (2000), Geneva, Switzerland.
[2]R. Guski and U. Felscher-Suhr, "The concept of noise annoyance: How international experts see it," J. Sound Vib. **223**, 513–527 (1999).
[3]H. S. Koelega, "Environmental Annoyance: Characterization, Measurement and Control" (Elsevier, Amsterdam, 1987).
[4]R. Guski, "Conceptual, methodological and dose-response problems related to annoyance and disturbance," Inter-Noise **97**, 1077–1082 (1997).
[5]J. M. Fields, "Effects of personal and situational variables on noise annoyance in residential areas," J. Acoust. Soc. Am. **93**, 2753–2763 (1993).
[6]H. M. E. Miedema and H. Vos, "Demographic and attitudinal factors that modify annoyance from transportation noise," J. Acoust. Soc. Am. **105**, 3336–3344 (1999).
[7]T. H. J. Schultz, "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405 (1978).
[8]H. M. E. Miedema and H. Vos, "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445 (1998).
[9]K. D. Kryter, "Community annoyance from aircraft and ground vehicle noise," J. Acoust. Soc. Am. **72**, 1212–1242 (1982).
[10]K. D. Kryter, "Community annoyance from aircraft and ground vehicle noise (Response of K. D. Kryter to modified comments by T. H. J. Schultz of K. D. Kryter's paper)," J. Acoust. Soc. Am. **73**, 1066–1068 (1983).
[11]S. Fidell, D. S. Barber, and T. H. J. Schultz, "Updating a dosage-effect relationship for the prevalence of annoyance due to general transportation noise," J. Acoust. Soc. Am. **89**, 221–233 (1991).
[12]L. S. Finegold, C. S. Harris, and H. E. von Gierke, "Community annoyance and sleep disturbance: Updated criteria for assessing the impacts of general transportation noise on people," Noise Control Eng. J. **42**, 25–30 (1994).
[13]U. Moehler, "Community response to railway noise: a review of social surveys," J. Sound Vib. **120**, 321–332 (1988).
[14]C. G. Rice, "Subjective assessment of transportation noise," J. Sound Vib. **43**, 407–417 (1975).
[15]V. Knall and R. Schuemer, "The differing annoyance levels of rail and road traffic noise," J. Sound Vib. **87**, 321–326 (1983).
[16]J. M. Fields and J. G. Walker, "Comparing the relationships between noise

level and annoyance in different surveys: A railway noise vs. aircraft and road traffic comparison," J. Sound Vib. **81**, 51–80 (1982).
[17]T. Yano, T. Yamashita, and K. Izumi, "Comparison of community annoyance from railway noise evaluated by different category scales," J. Sound Vib. **205**, 505–511 (1997).
[18]J. Igarashi, "Comparison of community response to transportation noise: Japanese results and annoyance scale," J. Acoust. Soc. Jpn. **13**, 301–309 (1992).
[19]J. Kaku and I. Yamada, "The possibility of a bonus for evaluating railway noise in Japan," J. Sound Vib. **193**, 445–450 (1996).
[20]T. Morihara, T. Sato, and T. Yano, "Comparison of dose-response relationships between railway and road traffic noises: The moderating effect of distance," J. Sound Vib. **277**, 559–565 (2004).
[21]H. M. E. Miedema and C. G. M. Oudshoorn, "Annoyance from transportation noise: Relationships with exposure metrics Ldn and Lden and their confidence intervals," Environ. Health Perspect. **109**, 409–416 (2001).
[22]C. Lim and S. Lee, "Questionnaire on environmental noise: The core set," Center for Environmental Noise and Vibration Research, 2003.
[23]International Standard Organization (ISO), Acoustics – Assessment of noise annoyance by means of social and socio-acoustic surveys, ISO/TS 15666, 2003.
[24]S. Lee, S. Min, J. Park, and S. Yun, *The Practice on Logit and Probit Model* (Pakyoungsa, Seoul, 2005).
[25]A. J. Dobson, *An Introduction to Generalized Linear Models* (Chapman and Hall, London, 1990).
[26]P. D. Allison, *Logistic Regression Using the SAS System: Theory and Application*, Cary, NC: SAS Institute, Inc., 1991.
[27]M. M. Hawkins, "Subjective evaluation of noise in areas with low ambient levels," *Proceedings of the Institute of Acoustics Spring Conference*, 20.E6.1–4, Southampton, 1979.
[28]Environment Agency of Japan, Report of the research on the guideline for conventional railway noise (1994).
[29]J. Kaku, "Community response to railway noise-comparison of social survey results between Japan and other countries," *Proceedings of Inter-Noise 94*, Vol. I, p. 121–124, Yokohama, Japan, 1994.
[30]Y. T. Lim, "Estimating the value of traffic noise through the analysis of housing prices," Ph.D. Thesis, University of Seoul, 2000.

On the use of a diffusion model for acoustically coupled rooms[a)]

Alexis Billon, Vincent Valeau,[b)] and Anas Sakout
LEPTAB, Pôle Sciences et Technologie, Université de La Rochelle, Avenue Michel Crépeau, 17042 La Rochelle cedex 01, France

Judicaël Picaut[c)]
Laboratoire Central des Ponts et Chaussées, Section Acoustique Routière et Urbaine, Route de Bouaye, Boîte Postale 4129, 44341 Bouguenais cedex, France

(Received 31 January 2006; revised 29 June 2006; accepted 28 July 2006)

A numerical model is proposed to predict the reverberant sound field in a system of two coupled volumes that are connected through an open aperture. The model is based on the numerical implementation of a diffusion model that has already been applied to predict the sound-energy distribution and the sound decay in single rooms. In comparison with the statistical theory, the proposed approach permits the prediction of the sound field by taking into account the sound source location and the receiver locations as well as the transition from one room to the other at the coupling aperture. Moreover, the diffusion model results match satisfactorily the experimental data in terms of sound-pressure level and reverberation times, both in the room containing the source and in the receiving room. Simulations with a ray-based model are also carried out, leading to results similar to those of the diffusion model, but at a cost of larger computation times. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2338814]

PACS number(s): 43.55.Br, 43.55.Ka, 43.55.Cs [NX] Pages: 2043–2054

I. INTRODUCTION

Coupled volumes systems, composed of two or more spaces that are connected through acoustically transparent openings (i.e., a coupling aperture), have attracted considerable attention in architectural acoustics. This configuration can be found in various buildings or constructions such as concert halls fitted with reverberation chambers, industrial halls, or office spaces. In workspaces the user's acoustical comfort is of particular interest. In concert halls, a very high acoustic quality is required. The prediction of the different acoustical parameters (sound-pressure levels, reverberation times, speech intelligibility, etc.) is then needed. Since Davis's initial work,[1] several models have been proposed such as statistical theory,[1–8] statistical energy analysis,[9,10] modal theory,[11–13] finite-element methods,[14] and ray-based models.[8,15–22] Despite being based on the diffuse-sound-field theory assumption, the statistical theory has been compared satisfactory with experimental[2,15,20,21,23] and numerical results.[8,21,22] Nevertheless, different authors question the ability of the statistical theory to deal with room modes, geometric and acoustic details, as well as nondiffuse sound field.[8,11,13] The modal theory and the finite-elements methods are limited to the low-frequency range, due to increasing computation times at higher frequencies. The ray-based model has shown quite good agreements with experimental data for churches[16] and concert halls,[23] as well as with statistical models,[8] provided that a large number of rays is emitted for small coupling apertures, which implies long computation times. This quick review shows that a model allowing spatial variations of the reverberant sound field—i.e., not based on the diffuse-sound-field assumption—both in terms of sound level and sound decay, for acceptable computation times is needed.

In a recent paper, Valeau *et al.*[24] have proposed a generalization and a numerical implementation of the diffusion model[25] for three-dimensional enclosures with perfectly diffusing walls. The numerical model has been validated both in stationary and time-varying states. The main interest of the model is its ability to give satisfactory results in only a few seconds or minutes, whatever the number of sound sources. Moreover, the sound decay and the sound level are calculated at any location in the considered volume: the diffusion model does not require definition of sound receivers at specific locations. In this paper, this numerical model is extended to the coupled-rooms configuration and is compared with experimental data as well as with the statistical theory and a ray-based model, in the case of a system of two coupled rooms. The three models used in this study are presented in Sec. II. Numerical applications of the diffusion model for coupled rooms are then proposed and compared to the statistical theory in Sec. III. In Sec. IV, both models as well as a ray-based model are compared to experimental data.

II. BACKGROUND: MODEL PRESENTATION

The practical configuration under interest consists of two rooms acoustically coupled through an acoustically transparent aperture; the one containing the source is called *source room* and the second one, *receiving room*. The three different

[a)]Part of this work was presented at the 148th and 149th meetings of the Acoustical Society of America [J. Acoust. Soc. Am. **116**, 2554 (2004); **117**, 2581 (2005)].

[b)]Present address: LEA, UMR CNRS 6609, Bâtiment K, 40 Av. Recteur du Pineau, 86022 Poitiers cedex, France.
[c)]Author to whom correspondence should be addressed. Electronic mail: judicael.picaut@lcpc.fr

models used in this paper are detailed in the present section. The first model (Sec. II A) is an application of the well-known statistical theory based on the diffuse-sound-field assumption. The second one, which is the main subject of this paper, is based on the concept of diffusion. The general features of this model are presented in Sec. II B. The third model is based on the ray-tracing theory (Sec. II C).

A. Statistical theory for two coupled rooms

1. Steady state

The energy decay of a system of coupled rooms can be evaluated by using the statistical theory based on Sabine's model.[26] The model was first developed by Davis[1] and is detailed in several references.[3–7] Following the diffuse-sound-field assumption, the sound-energy density of the reverberant sound field is supposed to be uniform in each room—with values denoted E_S and E_R for the source room and the receiving room, respectively. The energy transition between the rooms is then stepwise at the coupling aperture. Although a modified model of coupled rooms can be found in a recent paper,[8] only the original one[5–7] is considered in the following.

By writing the stationary energy balance of the system,[5–7] the relationship between the sound-pressure level (SPL) in each room (L_S and L_R for the source and receiving rooms, respectively) is given by[5]

$$L_S - L_R = -10\log(k_R), \tag{1}$$

where k_R $(0 \leq k_R \leq 1)$ is the coupling factor of the receiving room. Here, k_R depends on the coupling area S_c and on the absorption area of the receiving room

$$A_R = \sum_j \alpha_j S_j - S_c, \tag{2}$$

where j denotes the jth element of the room surface, as[5,23]

$$k_R = \frac{E_R}{E_S} = \frac{S_c}{S_c + A_R}, \tag{3}$$

where $k_R \approx 1$ denotes a strong coupling (the two rooms behave as a single larger one) and $k_R \approx 0$ a weak coupling (weak energy transfer from the source room to the receiving room). One can also define the coupling factor of the source room by

$$k_S = \frac{S_c}{S_c + A_S}. \tag{4}$$

2. Sound decay

The time-varying energy balance of the system allows one to calculate the temporal sound-energy decay in both rooms. Let us denote δ_S and δ_R the damping constants of the source and receiving rooms, respectively, as if they were uncoupled. The variation of the energy density with time as a function of the coupling parameters can be written[5–7]

$$E_S(t) = E_{S1}\exp(-2\delta_I t) + E_{R2}\frac{k_S}{1 - \delta_{II}/\delta_S}\exp(-2\delta_{II} t), \tag{5}$$

and

$$E_R(t) = E_{S1}\frac{k_S}{1 - \delta_I/\delta_R}\exp(-2\delta_I t) + E_{R2}\exp(-2\delta_{II} t), \tag{6}$$

where $\delta_{I,II}$ are the corresponding eigenvalues for the coupled room, defined by

$$\delta_{I,II} = \frac{1}{2}(\delta_S + \delta_R) \mp \sqrt{\frac{1}{4}(\delta_S - \delta_R)^2 - (1-\kappa)^2 \delta_S \delta_R}, \tag{7}$$

and where κ is the mean coupling factor,

$$\kappa = \sqrt{k_S k_R} = \sqrt{\frac{S_c^2}{(S_c + A_R)(S_c + A_S)}}. \tag{8}$$

E_{S1} and E_{R2} refer to the initial value for the different exponential decays given by[7]

$$E_{S1} = \frac{E_{0S} - E_{0R} k_S/(1 - \delta_{II}/\delta_S)}{1 - \kappa^2/(1 - \delta_I/\delta_S)(1 - \delta_{II}/\delta_S)}, \tag{9}$$

$$E_{R2} = \frac{E_{0R} - E_{0S} k_R/(1 - \delta_I/\delta_S)}{1 - \kappa^2/(1 - \delta_I/\delta_S)(1 - \delta_{II}/\delta_S)}, \tag{10}$$

where E_{0S} and E_{0R} are the initial conditions of the source room and the receiving room according to

$$E_{0S} = \frac{4}{c}\frac{P}{(A_S + S_c)(A_R + S_c) - S_c^2}, \tag{11}$$

and

$$E_{0R} = k_R E_{0S}, \tag{12}$$

with P the acoustic power of the omnidirectional source. The decay properties are unaltered by the coupling other than the deviation of the decay constants δ_I and δ_{II} from the decay constants of each uncoupled room δ_S and δ_R as predicted by the statistical theory.

This model is most accurate when applied to systems that are not strongly coupled[1,3–6] but its validity limits are not well defined.[8]

B. Diffusion model

Picaut *et al.*[25] have proposed a model, first derived by Ollendorff,[27] to describe the local acoustic energy density in rooms with perfectly diffuse reflecting walls. By using a physical analogy with the diffusion of particles in a medium containing spherical scattering objects, as presented by Morse and Feschbach,[28] they showed that the acoustic energy density may be the solution of a diffusion equation. The diffusion model has been established as a natural extension of the classical theory for diffuse sound field.[24] This model has been successfully applied to one-dimensional room-acoustic applications.[29] In a more recent paper,[24] Valeau *et al.* have generalized this model to three-dimensional enclosures and proposed a numerical implementation of this model for room-acoustic predictions. The main features of the diffusion model are now presented.

For a room of volume V and surface area S, the mean-free path of the room λ can be evaluated by the simple analytical relation

$$\lambda = 4V/S. \tag{13}$$

In particular, this expression is valid in the case of a room with diffusively reflecting boundaries—i.e., with directional reflection properties that are described by Lambert's law,[6] but can also be applied to quasicubic room with specular reflections.[30]

Following the physical analogy with the diffusion of particles in a scattering medium, the local acoustic energy-density flux $\mathbf{J}(\mathbf{r},t)$ can be approximated as the gradient of the energy density,

$$\mathbf{J}(\mathbf{r},t) = -D\,\nabla\, w(\mathbf{r},t), \tag{14}$$

where the variables $\mathbf{r}$ and t denote the position and time, respectively. D is the so-called diffusion coefficient, and its analytical expression is directly taken from the theory of diffusion for particles in a scattering medium,

$$D = \frac{\lambda c}{3}, \tag{15}$$

$$= \frac{4Vc}{3S}, \tag{16}$$

where c is the speed of sound. This term takes the room morphology into account through its mean-free path.

Let us consider the case of a room containing an acoustic omnidirectional point source located at position $\mathbf{r}_s$ and with an output acoustic power $P(t)$. It can be shown from Eq. (14) that the acoustic energy density, denoted $w(\mathbf{r},t)$, is the solution of the following diffusion equation [Eq. (17)], associated with mixed-type boundary conditions [Eq. (18)]:[24,25,28]

$$\frac{\partial w(\mathbf{r},t)}{\partial t} - D\nabla^2 w(\mathbf{r},t) = P(t)\,\delta(\mathbf{r}-\mathbf{r}_s) \quad \text{in } \mathcal{D}, \tag{17}$$

$$D\frac{\partial w(\mathbf{r},t)}{\partial n} + \frac{c\alpha}{4} w(\mathbf{r},t) = 0 \quad \text{on } \mathcal{S}. \tag{18}$$

In these equations, ∇^2 is the Laplace operator, $\mathcal{D}$ denotes the domain delimited by the room surfaces, $\mathcal{S}$ denotes the room boundaries, and α is the local wall absorption coefficient. In Eq. (17), the right-hand term is a source term which models the omnidirectional acoustic source in terms of power output and location.[24] The boundary condition defined by Eq. (18) models the sound-energy absorption by the room walls using Sabine's absorption coefficient. This approach has been applied analytically[25,29] to the calculation of the reverberant sound field in long rooms to predict both the stationary field and the sound decay, and numerically[24] for several room configurations (rooms with homogeneous dimensions, long rooms, flat rooms, etc.). In this paper, this concept is applied to two coupled rooms.

C. Ray tracing

The CATT-Acoustic ray-tracing software (V.8.0c), including diffuse reflection, was used to obtain spatial variations of the sound pressure and the sound decay in each room. Diffuse reflections are modeled by using Lambert's law:[6] a scattering coefficient is then defined as the fraction of the energy which is not specularly reflected. Moreover, an algorithm is proposed in this software to evaluate the late-part ray tracing of the impulse response, specifically to achieve better predictions in coupled-rooms configurations.[20,31]

III. NUMERICAL RESULTS OF THE DIFFUSION MODEL

A. Numerical solving of a diffusion equation for coupled rooms

For simulating the acoustics of two coupled rooms connected by an open aperture, the system of equations (17) and (18) is solved numerically in the case where $\mathcal{D}$ is the domain

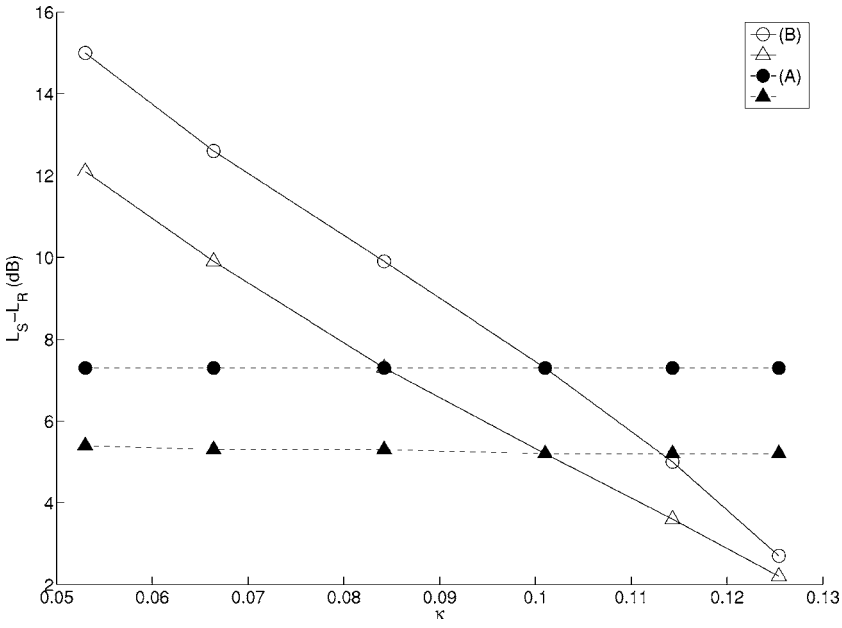


FIG. 1. Sound-pressure level difference between rooms as a function of κ (reverberant sound field only, i.e., without the direct field). (A) The absorption coefficient of the receiving room is constant (α_R=0.1) while the absorption coefficient of the source room α_S varies from 0.02 to 0.7: (•) statistical theory; (▲) diffusion model. (B) The absorption coefficient of the source room is constant (α_S=0.1) while the absorption coefficient of the receiving room α_R varies from 0.02 to 0.7: (◦) statistical theory; (△) diffusion model.

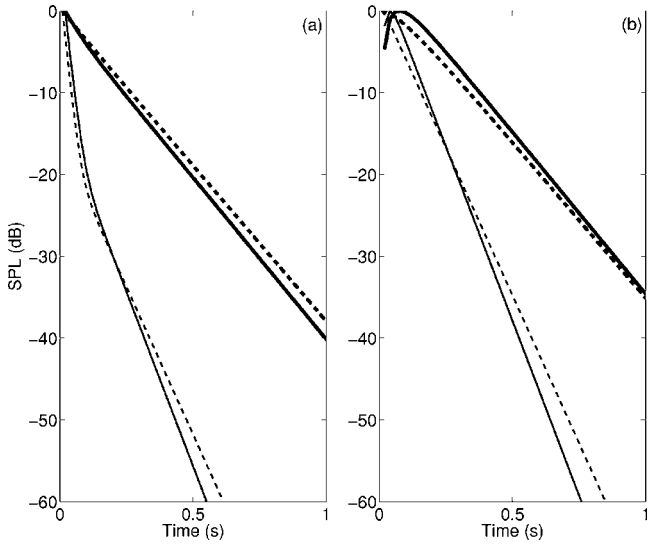


FIG. 2. Sound decays in the source room (a) and the receiving room (b) for varying source room's absorption α_S=0.05 (bold lines) and α_S=0.5 (thin lines) with a constant receiving room's absorption α_R=0.1. (—) diffusion model; (- - -) statistical theory.

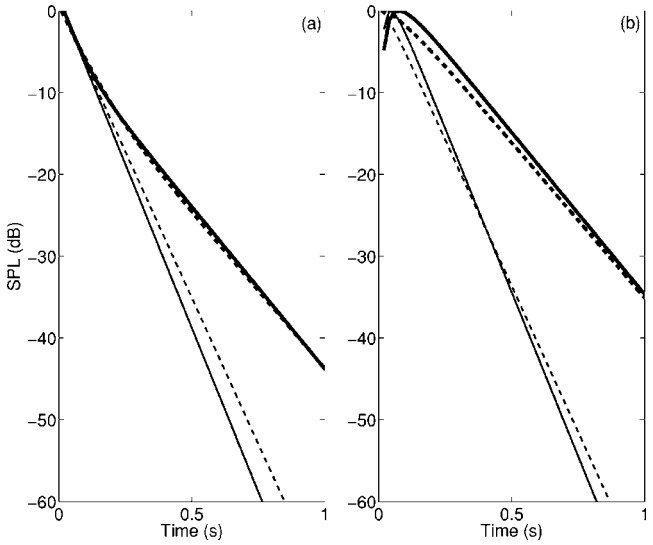


FIG. 3. Sound decays in the source room (a) and the receiving room (b) for varying receiving room's absorption α_R=0.05 (bold lines) and α_R=0.3 (thin lines) with a constant source room's absorption α_S=0.1. (—) diffusion model; (- - -) statistical theory.

defined by the two coupled rooms.[32] It is decomposed into three subvolumes, $\mathcal{D}_s$, $\mathcal{D}_r$, and $\mathcal{D}_{ss}$, the union of which makes the total calculation domain $\mathcal{D}$. $\mathcal{D}_{ss}$ is a small sphere with a radius of the order of 0.17 m and a volume v, modeling the space occupied by an omnidirectional acoustic source with an output acoustic power P_s. $\mathcal{D}_s$ is the domain delimited by the walls of the source room and the coupling aperture, minus $\mathcal{D}_{ss}$. $\mathcal{D}_r$ is the domain delimited by the walls of the receiving room and the coupling aperture.

In this study the mean-free path of each coupled room is set to the value that it would have if the rooms were uncoupled—i.e., the mean-free path that can be calculated for each room by using Eq. (13), as the open aperture is replaced with a wall surface. As mentioned in Sec. II B, the mean-free path of each room can be however dissimilar. This approximation means that the coupling aperture area is small compared to the area of the wall surfaces for each room, so that the mean-free path is not affected much by the open aperture. The approximated values for the mean-free path are denoted λ_s and λ_r for the source and receiving room, respectively.

The reverberant field in each room can be calculated by using the diffusion model, with diffusion coefficients $D_s = \lambda_s c/3$ and $D_r = \lambda_r c/3$ [see Eq. (15)]. The physical problem which has to be solved is then nonhomogeneous diffusion—i.e., the diffusion coefficient exhibits spatial variations over the calculation domain.

For simulating the response of the coupled rooms to a stationary sound source located in the source room, the following set of equations is solved numerically to obtain the stationary energy density $w(\mathbf{r})$ in $\mathcal{D}$:

$$-D_s \nabla^2 w(\mathbf{r}) = 0 \quad \text{in } \mathcal{D}_s, \tag{19}$$

$$-D_s \nabla^2 w(\mathbf{r}) = \frac{P_s}{v} \quad \text{in } \mathcal{D}_{ss}, \tag{20}$$

$$-D_r \nabla^2 w(\mathbf{r}) = 0 \quad \text{in } \mathcal{D}_r. \tag{21}$$

where v is the sound source volume. These equations are the stationary forms of the general diffusion equation (17). The right-hand term in Eq. (20) models an acoustic power supply P_s from the source to the rooms.[24] The solution for $w(\mathbf{r})$ represents the reverberant field. This set of equations is associated with mixed-type boundary conditions such as in Eq. (18) in order to take the local variations of the wall absorp-

TABLE I. Reverberation times in the source room (a) and in the receiving room (b), for configuration A (α_R is constant while α_S is varying).

α_S	0.02	0.05	0.1	0.3	0.5	0.7
(a) Source room						
Diffusion model (RT_1)	2.23	1.49	0.99	0.29	0.15	0.10
Statistical model (RT_1)	2.26	1.54	1.08	0.30	0.17	0.11
Diffusion model (RT_2)	⋯	⋯	⋯	0.73	0.70	0.69
Statistical model (RT_2)	⋯	⋯	⋯	0.80	0.81	0.83
(b) Coupled room (α_R=0.1)						
Diffusion model (RT_1)	2.26	1.55	1.09	0.76	0.71	0.70
Statistical model (RT_1)	2.75	1.60	1.07	0.83	0.83	0.83

TABLE II. Reverberation times in the source room (a) and in the receiving room (b), for configuration B (α_S is constant while α_R is varying).

(α_S=0.1) α_R	0.02	0.05	0.1	0.3	0.5	0.7
(a) Source room						
Diffusion model (RT_1)	1.02	0.98	0.99	0.75	0.71	0.69
Statistical model (RT_1)	1.06	1.00	0.94	0.84	0.83	0.83
Diffusion model (RT_2)	2.21	1.50	⋯	⋯	⋯	⋯
Statistical model (RT_2)	2.73	1.50	⋯	⋯	⋯	⋯
(b) Coupled room						
Diffusion model (RT_1)	2.25	1.54	1.09	0.76	0.72	0.71
Statistical model (RT_1)	2.75	1.60	1.07	0.84	0.83	0.83

tion coefficient into account. The sound-pressure level of the reverberant sound field is then given by

$$L'_p(\mathbf{r}) = 10\log(\rho c^2 w(\mathbf{r})/P_{\text{ref}}^2). \tag{22}$$

Finally, the local sound-pressure level, including the contribution of the direct field, can be computed by using the following relation:

$$L_p(\mathbf{r}) = 10\log(\rho c[P_s/(4\pi r^2) + w(\mathbf{r})c]/P_{\text{ref}}^2), \tag{23}$$

ρ being the air density, r the source-receiver distance, and with $P_{\text{ref}}=2\times10^{-5}$ Pa. For later use, and in order to make the comparison with statistical-theory results easier (as this theory provides a single value of the SPL per room), a "mean" SPL of the reverberant field can also be obtained from the diffusion model, by averaging the stationary energy density $w(\mathbf{r})$ over each room.

To model the sound decay in the rooms, the following system of equations is solved:

$$\frac{\partial w(\mathbf{r},t)}{\partial t} - D_s\nabla^2 w(\mathbf{r},t) = 0 \quad \text{in } \mathcal{D}_s, \tag{24}$$

$$\frac{\partial w(\mathbf{r},t)}{\partial t} - D_s\nabla^2 w(\mathbf{r},t) = 0 \quad \text{in } \mathcal{D}_{ss}, \tag{25}$$

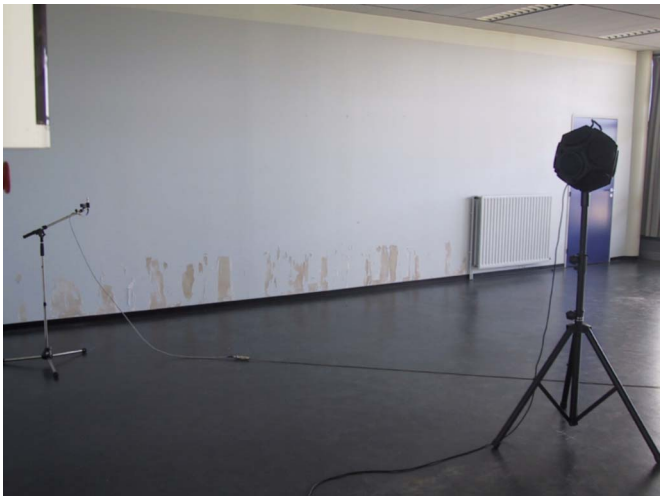


FIG. 4. Photography of the source room (the receiving room is similar). Note that the door between the rooms was removed for the S1 and S2 measurements configurations.

$$\frac{\partial w(\mathbf{r},t)}{\partial t} - D_r\nabla^2 w(\mathbf{r},t) = 0 \quad \text{in } \mathcal{D}_r, \tag{26}$$

together with the following initial conditions:

$$w(\mathbf{r},0) = 0 \quad \text{in } \mathcal{D}_s \cup \mathcal{D}_r, \tag{27}$$

$$w(\mathbf{r},0) = w_0 \quad \text{in } \mathcal{D}_{ss}, \tag{28}$$

where w_0 is the initial value for the energy density contained in the source domain $\mathcal{D}_{ss}$. The obtained solution for $w(\mathbf{r},t)$ allows one to estimate the sound decay at any location in the rooms, and subsequently, the reverberation time (RT20, based on the −5 to −25-dB parts of the temporal sound-decay curve).

In this paper, numerical simulations for the diffusion model are performed using a finite-element method-based software (FEM).

B. Comparison with statistical-theory results

In this section, the behavior of the diffusion model as a function of the coupling factor is first compared to results of the statistical theory, both in terms of difference between the mean sound-pressure levels in each room and in terms of reverberation times for each room. The studied geometry is composed of two rooms ($5\times5\times2.5$ m^3) coupled through a 0.9×2.5 m^2 aperture located on the side of the separating wall.

In the following, the total volume of the coupled rooms is meshed by using 3200 linear Lagrange-type elements. The calculation time for obtaining the pressure level with the diffusion model at any point in the room is about 1 min for stationary calculations, and about 15 min for time-dependent calculations.

The absorption is homogeneous for each room. Two cases are presented: for the first one (A), the absorption coefficient of the receiving room α_R remains constant and equal to 0.1, while the absorption of the source room α_S varies over the range 0.02 to 0.7. For the second one (B), the source room absorption coefficient is constant (0.1) and the absorption of the receiving room changes from 0.02 to 0.7. The results are presented as a function of the mean coupling factor κ of the system [Eq. (8)].

Figure 1 plots the difference between the mean reverberant sound-pressure levels of the source room and the receiving room (L_S and L_R, respectively), for several values of the

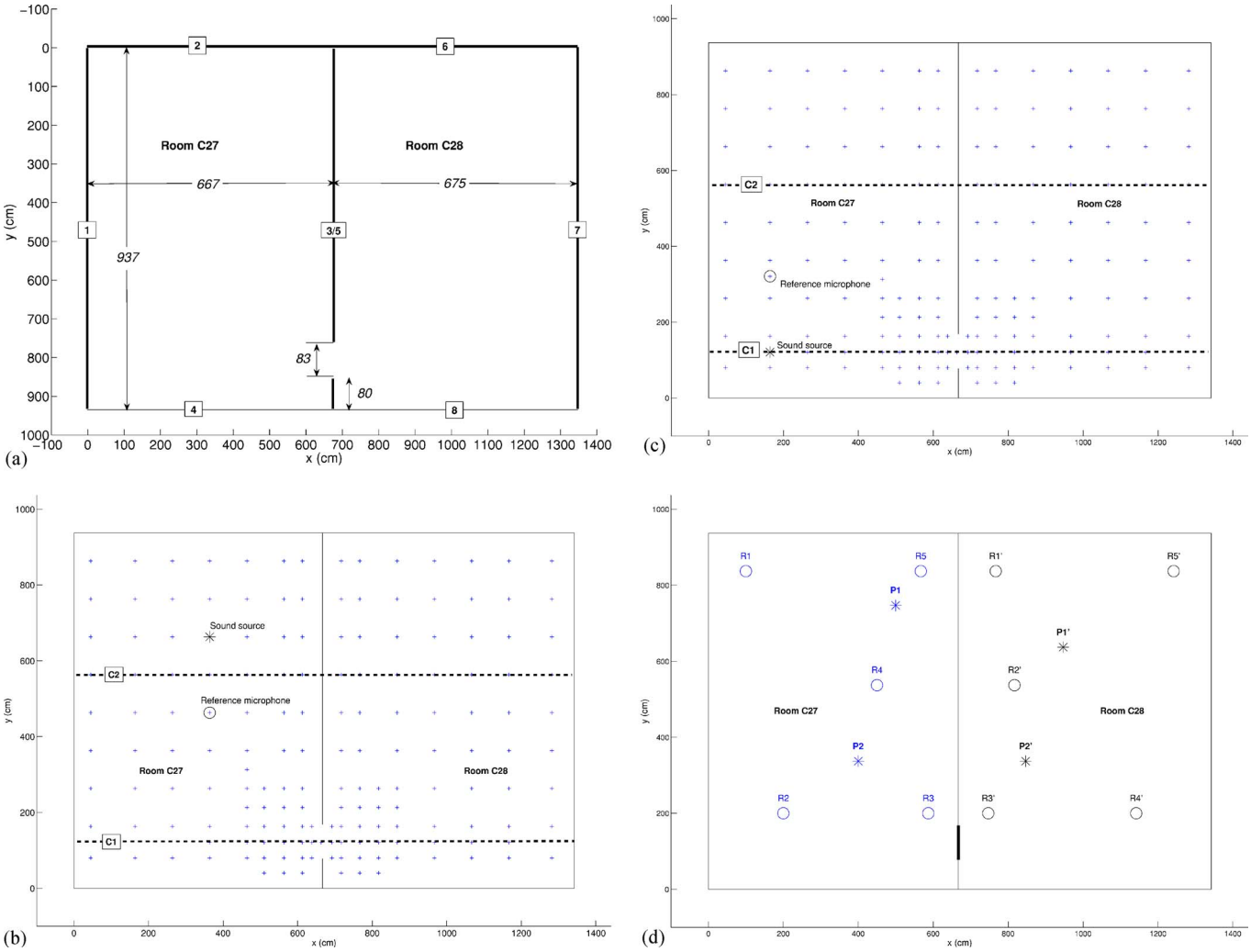


FIG. 5. Measurement configurations; (a) room geometry; (b) configuration S1: the omnidirectional sound source is located in room C27 with a reference microphone at 2 m from the source; 165 measurement points (symbols +) are distributed in both rooms (82 in C27 and 83 in C28). The aperture is opened; (c) configuration S2: the sound source is located in room C27 with a reference microphone at 2 m from the source; 168 measurement points (symbols +) are distributed in both rooms (85 in C27 and 83 in C28). The aperture is opened; (d) configurations S3/S4: 2 sound source locations (P1/P2 and P1'/P2') are considered in each room with 5 receivers (R1 to R5 and R1' to R5'). The aperture is closed.

coupling factor κ. For the statistical theory, the pressure level difference between the rooms depends only on k_R [see Eq. (1)]. When the absorption coefficient of the source room increases (configuration A), which implies that the mean coupling factor decreases whereas k_R remains constant, the difference of sound-pressure level between both rooms does not change. The diffusion model describes the same behavior with, nevertheless, a bias of 1.9 dB. When the absorption coefficient of the receiving room increases (configuration B), which implies that the mean coupling factor and k_R decrease, the models exhibit similar behaviors—i.e., the SPL difference decreases with κ—with a maximum difference of 2.9 dB for the most reverberant receiving room configuration.

Sound decay examples are presented in Figs. 2 and 3 for several room absorption conditions. Sound decays given by the diffusion model match satisfactorily with the statistical theory, in the source room as well as in the receiving room. When the receiving room is more reverberant than the source room [Fig. 2(a)], the diffusion model gives the typical double-sloped decay in the source room. Moreover, the change of slope occurs at the same position. To account for the nonlinear decay, two reverberation times RT_1 and RT_2 are calculated by extrapolating the linear sound decay of each slope, and presented in Tables I and II, for configurations A and B, respectively. The diffusion model and the statistical theory are in good agreement. When the source room is more reverberant than the receiving room (α_S=0.7 for configuration A or α_R=0.02 for configuration B), the first decay (RT_1) in the source room is due to the source room's absorption. On the other hand, the second decay (RT_2) in the source room is similar to the receiving room one. Note that the double-sloped decay is only visible in the source room and not in the receiving room. When the receiving room is very reverberant (α_S=0.7 in configuration A or α_R=0.02 in configuration B), several differences between both models occur. As in this configuration the rooms cannot be considered as weakly coupled; the results given by the statistical theory may be questionable. Otherwise, when the absorption becomes greater than 0.1, reverberation times predicted by the statistical theory remain almost constant, whereas the dif-

TABLE III. Reverberation times $RT20_{exp,C27}$ and $RT20_{exp,C28}$ measured in rooms C27 (a) and C28 (b) respectively, for each octave band. The equivalent absorption area $A_{exp,C27}$ and $A_{exp,C28}$ have been obtained from the Sabine formula. The absorption coefficients have been estimated for each material by adjusting the absorption coefficients in order to match the experimental values. The resulting equivalent absorption area $A_{emp,C27}$ and $A_{emp,C28}$, as well as the deviation σ from the experimental value, are also given.

(a) Room C27			125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	Full band
$RT20_{exp,C27}$			0.71	0.77	0.82	1.02	1.25	0.92	0.92
$A_{exp,C27}$			42.8	39.5	36.9	29.7	28.8	32.9	32.9
$A_{emp,C27}$			43.0	39.7	37.7	28.6	28.5	33.9	32.9
σ			0.38%	0.38%	2.13%	3.54%	1.08%	3.19%	2.1%
Wall	Surface	Material	Absorption coefficient						
1	28.4 m^2	Concrete	0.40	0.25	0.28	0.01	0.04	0.05	0.02
2	20.2 m^2	Plaster	0.03	0.08	0.09	0.14	0.06	0.09	0.06
3	26.7 m^2	Plaster	0.03	0.08	0.09	0.14	0.06	0.09	0.06
4	20.2 m^2	Glass	0.35	0.25	0.18	0.12	0.07	0.04	0.16
Floor	62.5 m^2	Linoleum	0.02	0.03	0.03	0.03	0.03	0.03	0.03
Ceiling	62.5 m^2	Fiber	0.35	0.35	0.32	0.28	0.34	0.41	0.39
(b) Room C28			125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	Full-band
$RT20_{exp,C28}$			0.86	0.84	0.77	0.69	0.80	0.88	0.92
$A_{exp,C28}$			35.7	36.6	39.6	44.4	38.1	34.9	37.9
$A_{emp,C28}$			32.8	35.2	38.9	44.7	37.6	35.4	38.1
σ			8.15%	3.95%	1.75%	0.49%	1.42%	1.49%	5.2%
Wall	Surface	Material	Absorption coefficient						
5	26.7 m^2	Plaster	0.03	0.08	0.09	0.14	0.06	0.09	0.06
6	20.4 m^2	Plaster	0.03	0.08	0.09	0.14	0.06	0.09	0.06
7	28.4 m^2	Plaster	0.03	0.08	0.09	0.14	0.06	0.09	0.06
8	20.4 m^2	Glass	0.35	0.25	0.18	0.12	0.07	0.04	0.16
Floor	63.2 m^2	Linoleum	0.02	0.03	0.03	0.03	0.03	0.03	0.03
Ceiling	63.2 m^2	Fiber	0.35	0.35	0.42	0.28	0.47	0.41	0.45

fusion model gives a slight decrease. The diffusion model seems to be more accurate in these cases because the sound decay increases with the absorption.

As a conclusion to this preliminary parametric study, the diffusion model and the statistical theory are in agreement for the tested parameters. For a given aperture size, in terms of mean SPL difference between the rooms, both models are only sensitive to the receiving room absorption area. The influence of the mean coupling factor on the reverberation time is also very similar for both models. At this step, the main difference between the two theories is that the diffusion model allow one to model the spatial variation of the reverberant sound field, while the statistical model gives only one constant value for each room. In the next section, this study is extended by comparing the diffusion model with experimental data, as well as with the ray-based model and the statistical theory.

IV. EXPERIMENTAL COMPARISONS

A. Measurements

Measurements were carried out in the Orbigny teaching building of La Rochelle University. Two empty coupled rooms with nearly identical geometries were considered [Fig. 4]: the source room C27 containing the sound source ($9.37 \times 6.67 \times 3.03$ m) and the receiving room C28 ($9.37 \times 6.75 \times 3.03$ m). The aperture size between the two rooms is 0.83 m wide and 2.06 m high. Both walls 4 and 8 [Fig. 5(a)] are entirely made of window glasses. The walls between the two rooms (wall 3/5) and along the corridor (walls 2 and 6) have a 10.5-cm thickness, and are made of a multilayer material, plaster/wood/glass wool/wood/plaster. Wall 7 of the room C28 is also made with the same material. Wall 1 of the room C27 is a concrete wall. Ceilings are made with 2-cm-thick wood fiber plates set up on a 20-cm plenum.

TABLE IV. Computation times of the numerical simulations.

Computation time (by octave band)	Steady state	Time-varying state
Diffusion model	30 s	8 min
Ray-based model	6 h	45 min

Four measurement configurations were considered, S1 to S4, respectively: two with the open aperture between the rooms (S1 and S2, i.e., the door is opened), and two with the closed aperture (S3 and S4, i.e., the door is closed). The last configurations were used to evaluate the noncoupled reverberation time, and then to estimate the wall absorption for each room, while the first configurations allow one to study the coupling between the rooms in terms of reverberation time and sound-pressure level.

For configurations S1 and S2 [Figs. 5(b) and 5(c)], the sound source was located in the C27 room at height 1.56 m. Impulse response measurements were carried out in both rooms C27 and C28, with a specific attention around the aperture: 165 and 168 receiver locations at 1.2 m from the floor were considered for configurations S1 and S2, respectively. For all measurements, a reference receiver was located 2 m from the sound source at height 1.2 m.

For configurations S3 and S4, two source locations and five receivers were considered for each room [Fig. 5(d)].

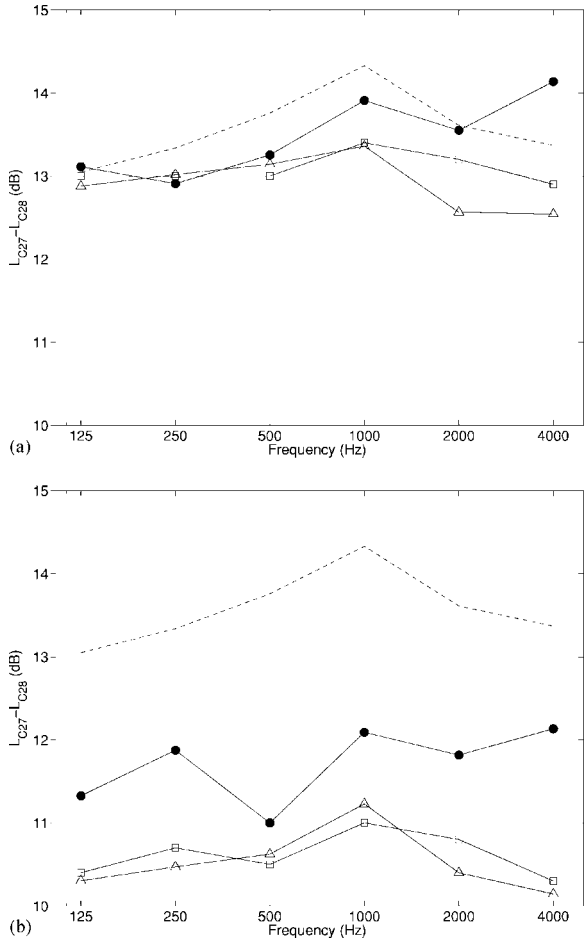


FIG. 6. Mean SPL difference between rooms as a function of frequency. (a) configuration S1, (b) configuration S2. (•) Experimental data; (△) diffusion model; (- - -) statistical theory; (□) ray-based model.

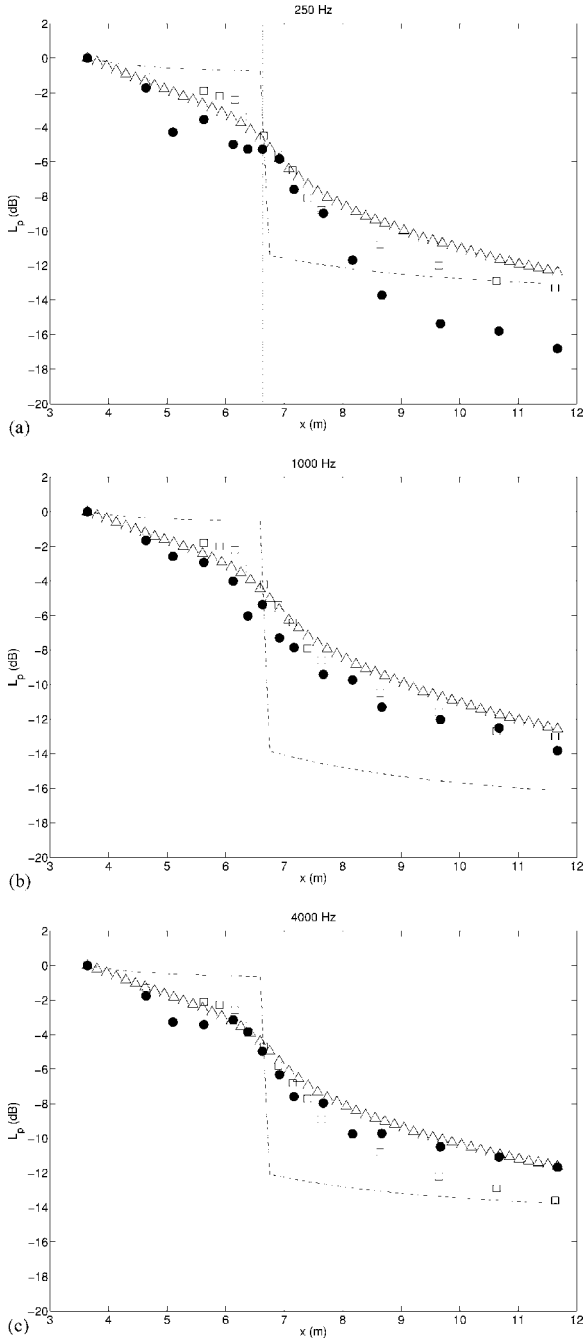


FIG. 7. Sound attenuation along line C1 for configuration S2 (with direct field): (a) 250 Hz; (b) 1000 Hz; (c) 4000 Hz. (•) Experimental data; (△) diffusion model; (- - -) statistical theory; (□) ray-based model.

Reverberation times for each source and receiver location were calculated from the impulse response. Due to the low signal-to-noise ratio (SNR), the RT20 was used instead of the conventional RT30. Then, the mean reverberation time of each room was obtained by averaging the reverberation time for both source and receiver locations in the same room.

Measurements were carried out with an omnidirectional loudspeaker type B&K 4296 connected to a power amplifier type B&K 2716, and using two 1/4-in. microphones type B&K 4135 connected to 2619 preamplifiers, all manufactured by Brüel & Kjær. Both microphones are connected to a NEXUS conditioning amplifier B&K 2690. The NEXUS and the source power amplifier are connected to a personal computer using a high-quality sound card.

Impulse response measurements were realized with the DSSF3 acoustic analysis software, by using the time-stretched pulse (TSP) method. A TSP signal is sent to the sound source while data acquisition is carried out on both microphones with a sample frequency of 48 kHz and a measuring time of 2.731 s. In order to avoid the small fluctuations due to the background noise, and to increase the SNR, 5 impulse responses were averaged at each receiver location for the S1/S2 configurations, and 20 impulse responses for the S3/S4 configurations. The sound source level is automatically adjusted by the DSSF3 software in order to achieve an optimal SNR at each receiver location. Sound pressure at each receiver location is then normalized by considering the sound pressure at the reference microphone.

B. Numerical data and parameters

In order to achieve accurate simulations with the diffusion and the ray-based models, the absorption coefficients of each material of both rooms C27 and C28 have to be determined with accuracy. From the S3 and S4 measurement configurations, one can determine the equivalent absorption area for each room, by octave band, from the measured reverberation time and the Sabine formula [Table III]. Moreover, one

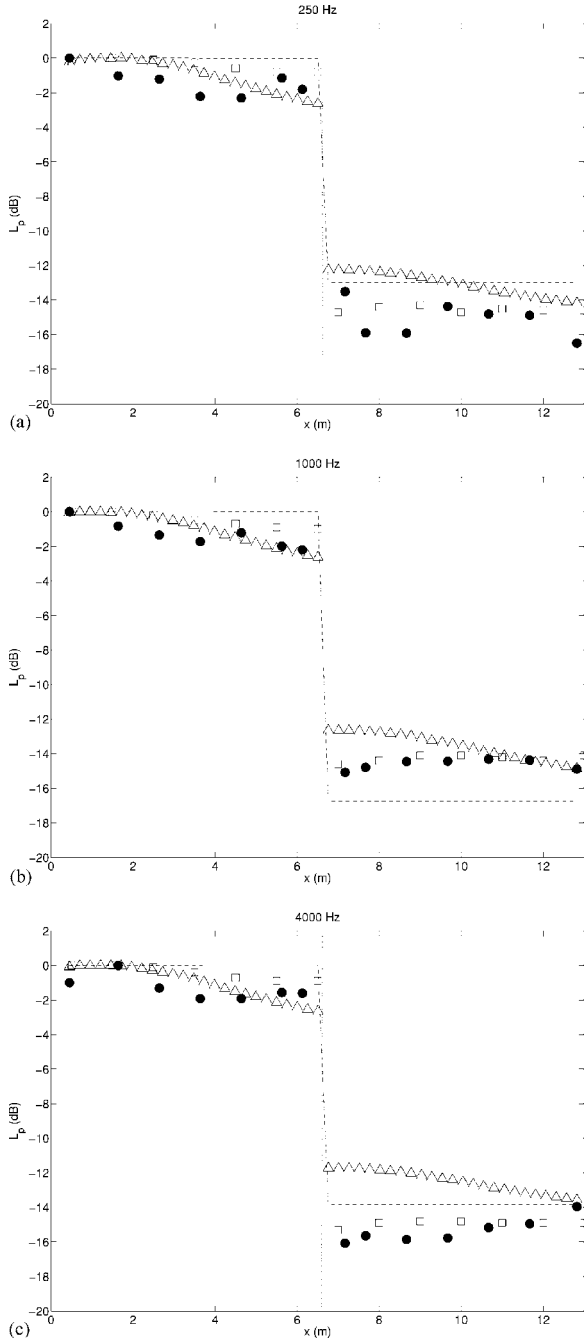


FIG. 8. Sound attenuation along the line C2 for configuration S2 (with direct field): (a) 250 Hz, (b) 1000 Hz; (c) 4000 Hz. (•) Experimental data; (△) diffusion model; (- - -) statistical theory; (□) ray-based model.

can also estimate the absorption coefficient of each material by using an absorption coefficient database. The coefficients are then slightly adjusted in order to match the experimental equivalent absorption area. Table III gives the resulting absorption coefficients for each wall, the resulting equivalent absorption area, and the deviation from the experimental value for each room. One can remark that room C27 is more reverberant than the C28 room in the middle frequency range (due to the concrete wall).

For the diffusion model, the simulated geometry is discretized into 5400 elements and the impulse responses are calculated over a 1.2-s interval. For ray-tracing simulations, preliminary numerical simulations have shown that 20

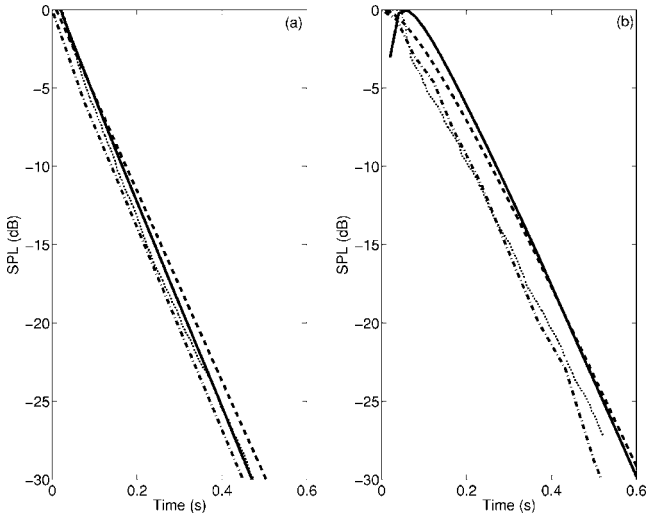


FIG. 9. Sound decays in (a) the source room at (x=4.64 m, y=4.63 m) and (b) the receiving room at (x=9.00 m, y=1.21 m): (⋯) experimental data; (—) diffusion model; (- - -) statistical theory; (- · -) ray-based model.

$\times 10^6$ and 200×10^3 sound rays are required to evaluate the steady-state sound levels and the reverberation times, respectively. Scattering coefficient was set to 1 (100% diffusely reflecting walls). Echogram length has been fixed to 1.5 s, and the late-part ray-tracing method has been used to estimate the sound decay of the impulse response. Direct field is included in the diffusion model and the statistical theory.

Wall absorption are extracted from Table III. All numerical simulations have been conducted on the same personal computer. Computation times are presented in Table IV. One can remark that the diffusion model clearly requires much less computation time than the sound-ray-based software for similar calculations. However, in the same time, ray-based models give also more information (D50, C50 for example) than the diffusion model.

C. Steady-state comparisons

1. Difference between mean sound-pressure level

Figure 6 plots the sound-pressure level difference between the source room (L_{C27}) and the receiving room (L_{C28}), for all octave bands between 125 Hz and 4 kHz, for both configurations S1 and S2. For the statistical theory, the sound level difference is given by Eq. (1). As explained in Sec. III A, for the other data (measurements, diffusion, ray-based model), the local energy density is averaged in each room, leading to one single value of the SPL (including the direct field when existing). The SPL in the receiving room is then subtracted from the mean SPL of the source room.

For the configuration S1 [Fig. 6(a)], all the models are close to the experimental results with a mean difference of about 0.5 dB for the diffusion model, 0.4 dB for the statistical theory, and 0.3 dB for the ray-based model, compared to experimental data. In this configuration, the source is located far away (roughly 6 m) from the coupling area: the energy provided by the direct field to the receiving room is then negligible in comparison with the reverberant field. In this case, the statistical theory assumption (i.e., only diffuse en-

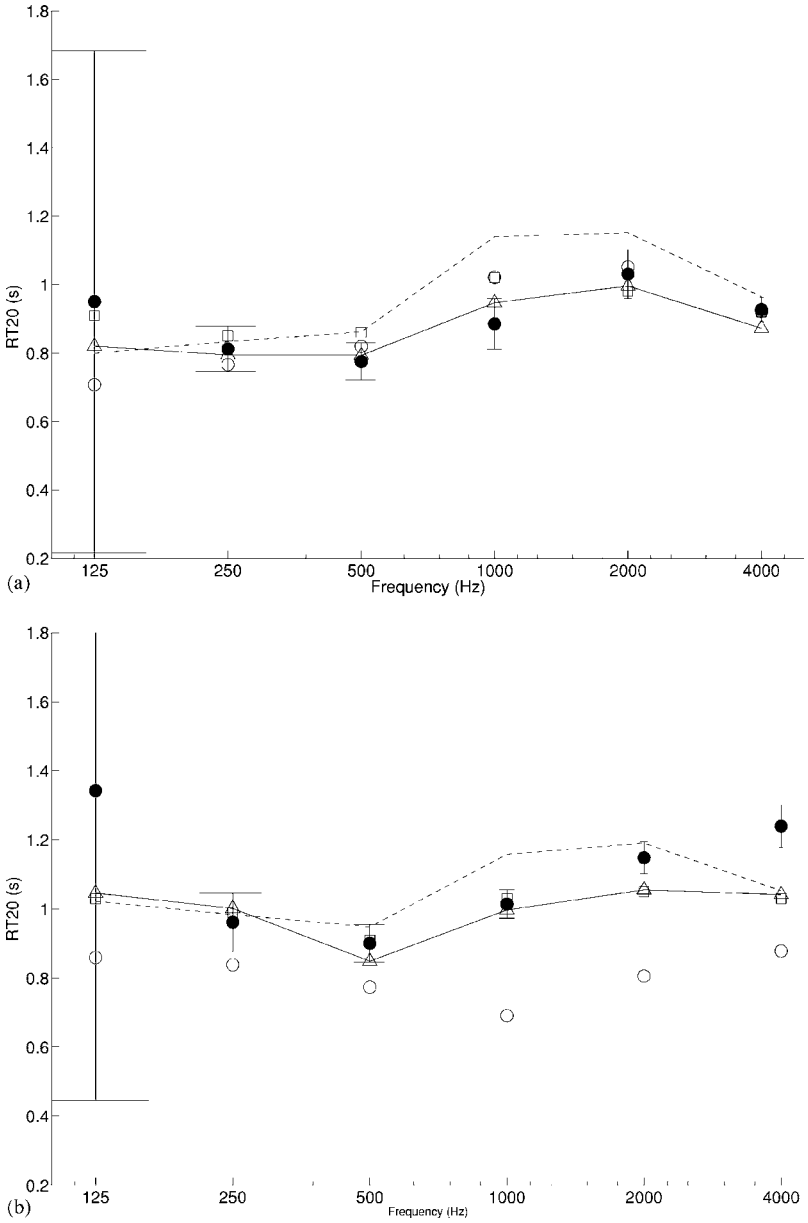


FIG. 10. Reverberation time (RT20) for configuration S2. (a) Source room; (b) receiving room. Coupled rooms: (•) experimental data; (△) diffusion model; (□) ray-based model; (- - -) statistical theory; Uncoupled rooms: (○) experimental data. The vertical bars indicate the dispersion of the RT measurements at all locations.

ergy is exchanged) is well respected, which explains the good agreement of the statistical theory with measurement.

In configuration S2 [Fig. 6(b)], the SPL difference between the two rooms is about 2 dB below the one obtained for configuration S1, due to the direct sound field radiated into the receiving room. This behavior is not predicted by the statistical theory since it does not take the effect of the source location into account: both configurations S1 and S2 give exactly the same results. Conversely, the diffusion and the ray-based models predict this drop between both configurations, although it tends to slightly underestimate the pressure level difference, with a maximum of 2 dB.

2. Sound attenuation through the coupling area

The sound-pressure level (with the direct field) is investigated along the C1 line parallel to the x axis [see Fig. 5(c)] at height 1.20 m and $y=1.21$ m. As the spatial decrease of the SPL is the main parameter in this section, the SPL is normalized with the maximum value along the line. Results for configuration S2 are presented in Fig. 7 for three representative octave bands (250, 1000, and 4000 Hz).

As stated in Sec. II A, the statistical theory does not predict the true gradual transition through the coupling area: the diffuse sound field at the coupling area is stepwise. The slight transition given at Fig 7 for the statistical theory is only due to the direct field. Conversely, the gradual sound attenuation due to the opening is well predicted both by the diffusion and ray-based models.

At 250 Hz [Fig. 7(a)], all models underestimate the sound attenuation within the receiving room. This may be due to the modal behavior of the coupled-rooms system for frequencies close to the Schroeder's frequency of the uncoupled rooms (about 130 Hz). At 1000 Hz [Fig. 7(b)], the diffusion and ray-based models give a sound attenuation close to the experimental data. Only for the highest octave band [Fig. 7(c)], the diffusion model seems to be more accurate than the sound ray-based software. For all octave bands the statistical theory overestimates the sound attenua-

tion in the receiving room [as formerly observed in Fig. 6(b)]. This last result shows that such model is not always adapted to predict the sound attenuation in coupled rooms with accuracy.

3. Sound pressure along the C2 line

The normalized SPL along line C2 is presented in Fig. 8 for configuration S2, for the same octave bands as previously. All models predict a sound attenuation in fair agreement with the experimental data, with 2 and 1 dB of averaged difference for the diffusion model at 250 and 1000 Hz, respectively. At 4000 Hz, as shown in Fig. 8(c), the diffusion model tends to underestimate the sound attenuation in the receiving room. For all octave bands, ray-based model seems to give the best prediction of the sound attenuation.

D. Time-varying state comparisons

1. Sound decays

The full-band sound decays in each room are presented in Fig. 9 for configuration S2. In the source room [Fig. 9(a)], all the models are in good agreement with the experimental results. One can observe that the nonlinear behavior of the experimental decay, usually observed for coupled-room systems,[33] is weak in this case, at least in the early part of the decay presented in Fig. 9. Then, in the following, the sound decay will be characterized using the RT20. The study of longer couples of sound decay would probably require the use of recent techniques especially dedicated to non-linear decay analysis.[33] In the receiving room [Fig. 9(b)], the agreement is not as well, since the models tend to slightly underestimate the sound decay compared to the experimental data. However, all models predict a linear sound decay, as experimentally observed.

2. Mean reverberation time

In order to compare to statistical theory, mean experimental RT20 (i.e., averaged over all receiver locations for each room) are presented in Fig. 10 for configuration S2, both for the coupled and uncoupled rooms. One can see that the high dispersion of results at frequency 125 Hz, as this octave band is just below the Schroeder's frequency of the uncoupled rooms. Since the influence of the sound source location on the reverberation time is weak, only the results for configuration S2 are plotted.

For the source room [Fig. 10(a)], all models are in fair agreement with the experimental data, with a mean difference about 6% in C27 and 11% in C28 for the diffusion model, about 7% and 10% for the ray-based model, and 13% and 14% for the statistical theory. This figure shows also that the reverberation time for C27 is not strongly affected by its coupling with C28. Conversely, the reverberation time in C28 increases due to the coupling with C27, which is more reverberant [Fig. 10(b)]: a part of the sound energy of the source room is transmitted through the coupling area into the receiving room during the temporal sound-energy decay, increasing the reverberation time of the receiving room.

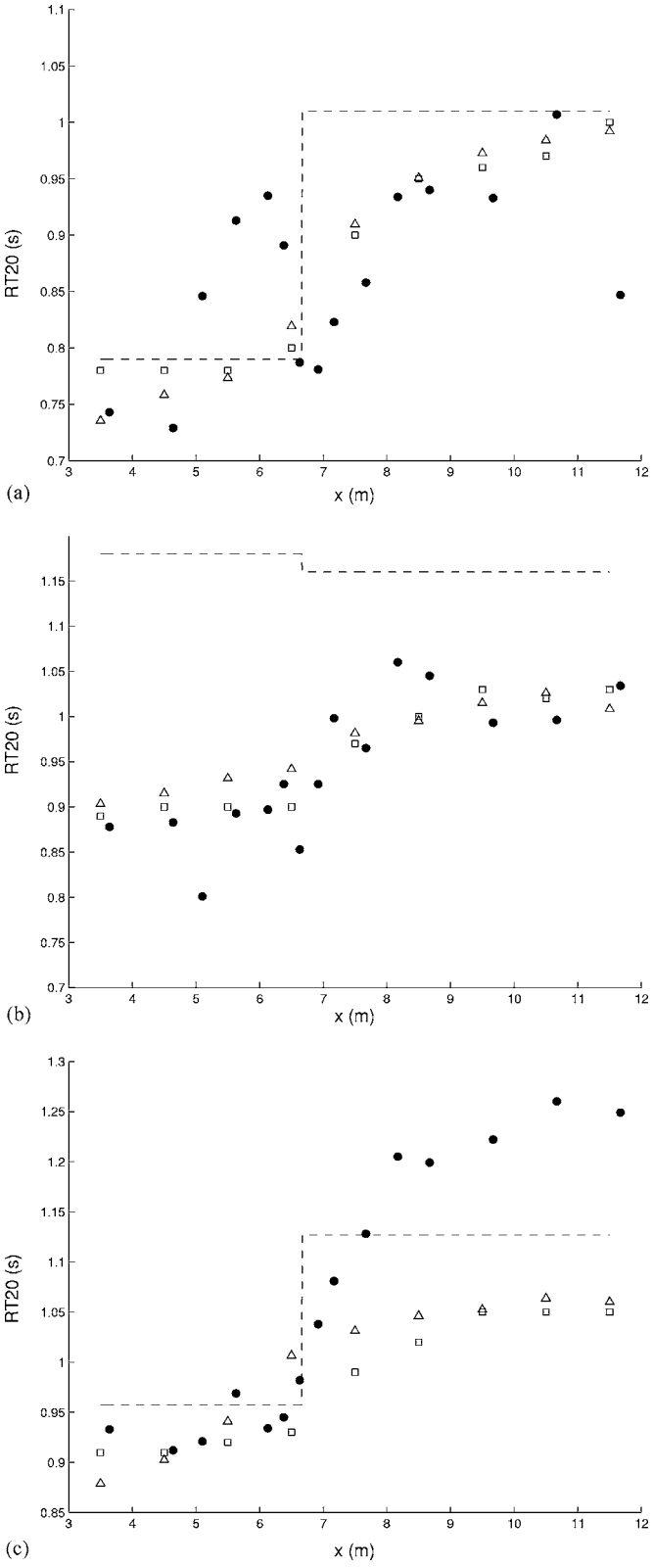


FIG. 11. Variation of the reverberation time (RT20) through the coupling aperture, for configuration S2 at (a) 250 Hz; (b) 1000 Hz; and (c) 4000 Hz. (•) Experimental data; (△) diffusion model; (□) ray-based model; (- - -) statistical theory.

3. Reverberation times through the coupling aperture

The variation of the reverberation time (RT20) through the coupling aperture is presented for the 150, 1000, and 4000-Hz octave bands in Fig. 11. The gradual increase of the

reverberation time, when passing through the source room to the receiving room (coupling aperture at $x=6.70$ m), is clearly observed both with the diffusion and the ray-based models. The agreement for the two first octave bands is good. On the other hand, for the third octave band, as previously shown in Figs. 9(b) and 10(b), all the models underestimate the reverberation time in the receiving room.

V. CONCLUSION

In this paper, the numerical implementation of a diffusion model for room acoustics has been applied to a configuration of two coupled volumes that are connected through an open aperture. In comparison with the statistical theory, for which only one SPL/RT value can be calculated for each room, the numerical solution of the analytical diffusion equation permits the prediction of the spatial variations of the sound-pressure level and reverberation time. Moreover, the diffusion model takes the sound source location into account, while the statistical theory does not. Measurements have been carried out for two coupled rooms, in order to describe with a good spatial resolution the variations of stationary level and sound decay. The statistical theory gives predictions that can describe only coarsely the acoustic coupling, but these predictions are in fair agreement with the main features of the experimental data. Conversely, the diffusion and ray-based models are able to describe the details of the acoustics of the coupled rooms with a satisfactory agreement, and in particular the transition phenomena occurring at the coupling aperture. However, the calculation time for sound-pressure levels and reverberation times is larger for the ray-based model than for the diffusion model, both for the stationary state and the time-varying state, for an accuracy of the same order. The proposed diffusion model should also be easily extended to the case of an arbitrary number of acoustically coupled volumes.

ACKNOWLEDGMENT

The authors would like to thank the *Agence de l'Environnement et de la Maîtrise de l'Énergie* (ADEME) for providing financial support of this work.

[1]A. H. Davis, "Reverberation equations for two adjacent rooms connected by an incompletely sound-proof partition," Philos. Mag. **50**, 75–80 (1925).
[2]C. F. Eyring, "Reverberation time measurements in coupled rooms," J. Acoust. Soc. Am. **3**(2), 181–206 (1931).
[3]C. D. Lyle, "The prediction of the steady state sound levels in naturally coupled enclosures," Acoust. Lett. **5**, 16–21 (1981).
[4]C. D. Lyle, "Recommendation for estimating reverberation time in coupled spaces," Acoust. Lett. **5**, 35–38 (1981).
[5]L. Cremer and H. Muller *Principles and Applications of Room Acoustics* (Applied Science, London, 1978), Vol. **1**.
[6]H. Kuttruff, *Room Acoustics*, 4th ed. (Spon, London, 1999).
[7]J. E. Summers, "Technical note: Remark on the formal identity of two statistical-acoustics models of coupled rooms," Build. Acoust. **12**, 41–50 (2005).
[8]J. E. Summers, R. R. Torres, and Y. Shimizu, "Statistical-acoustics models of energy decay in system of coupled rooms and their relation to geometrical acoustics," J. Acoust. Soc. Am. **116**(2), 958–969 (2005).
[9]C. B. Burroughs, R. W. Fischer, and F. R. Kern, "An introduction to statistical energy analysis," J. Acoust. Soc. Am. **101**(4), 1779–1789 (1997).
[10]M. Ohta, H. Yamada, and H. Iwashige, "A system-theoretical evaluation method for the reverberation time of an acoustically coupled room system," J. Acoust. Soc. Jpn. **16**(3), 137–145 (1995).
[11]C. M. Harris and H. Feshbach, "On the acoustics of coupled rooms," J. Acoust. Soc. Am. **22**(5), 572–578 (1950).
[12]A. Marchioni and O. Oldenbourg, "Description modale du couplage de deux salles," Rev. Acoust. **57**, 115–118 (1981).
[13]C. Thompson, "On the acoustics of a coupled space," J. Acoust. Soc. Am. **75**(3), 707–714 (1984).
[14]Y. Zhao and S. Wu, "Acoustical normal mode analysis for coupled rooms," in *Proceedings of the 21st International Conference of the Audio Engineering Society St. Petersburg, Russia* (Audio Engineering Society, New York, 2002).
[15]J. S. Anderson and M. Bratos-Anderson, "Acoustic coupling effects in St Paul's cathedral London," J. Sound Vib. **236**(2), 209–225 (2000).
[16]U. Ayr, E. Cirillo, and F. Martellotta, "Predicting room acoustical behaviour of coupled rooms with computer simulation techniques: A case study," in *Proceedings of the 17th International Congress on Acoustics* (17th International Congress on Acoustics, Rome, Italy, 2001).
[17]L. Nijs, G. Jansens, G. Vermeir, and M. van der Voorden, "Absorbing surfaces in ray-tracing programs for coupled spaces," Appl. Acoust. **63**, 611–626 (2002).
[18]J. E. Summers, "Comments on 'Absorbing surfaces in ray-tracing programs for coupled spaces'," Appl. Acoust. **64**, 825–831 (2003).
[19]L. Nijs, G. Jansens, and G. Vermeir, "Reply to 'Comments on Absorbing surfaces in ray-tracing programs for coupled spaces'," Appl. Acoust. **64**, 833–844 (2003).
[20]J. E. Summers, R. R. Torres, Y. Shimizu, and B.-I. L. Dalenbäck, "Adapting a randomized beam-axis-tracing algorithm to modeling of coupled rooms via late-part ray tracing," J. Acoust. Soc. Am. **118**(3), 1491–1502 (2005).
[21]M. Ermann, "Coupled volumes: Aperture size and the doublesloped decay of concert halls," Build. Acoust. **12**(1), 1–14 (2005).
[22]D. T. Bradley and L. M. Wang, "The effects of simple coupled volume geometry on the objective and subjective results from nonexponential decay," J. Acoust. Soc. Am. **118**(3), 1480–1490 (2005).
[23]M. Ermann and M. Johnson, "Exposure and materiality of the secondary room and its impact on the impulse response of coupled-volume concert halls," J. Sound Vib. **284**, 915–931 (2005).
[24]V. Valeau, J. Picaut, and M. Hodgson, "On the use of a diffusion equation for room acoustic predictions," J. Acoust. Soc. Am. **119**(3), 1504–1513 (2006).
[25]J. Picaut, L. Simon, and J.-D. Polack, "A mathematical model of diffuse sound field based on a diffusion equation," Acust. Acta Acust. 83, 614–621 (1997).
[26]W. C. Sabine, *Collected Papers on Acoustics* (Dover, New York, 1964).
[27]F. Ollendorff, "Statistische raumakustik als diffusionsproblem," Acustica **21**, 236–245 (1969).
[28]P. Morse and H. Feshbach, *Methods of Theoretical Physics* (Mc Graw-Hill, New-York, 1953).
[29]J. Picaut, L. Simon, and J.-D. Polack, "Sound field in long rooms with diffusely reflecting boundaries," Appl. Acoust. **56**, 217–240 (1999).
[30]J. Pujolle, "Les différentes définitions du libre parcours moyen du son dans une salle," Rev. Acoust. **36**, 44–50 (1976).
[31]J. E. Summers, "Reverberant acoustic energy in auditoria that comprise system of coupled rooms," Ph.D. dissertation, Renselear Polytechnic Institute, Troy, NY, 2003.
[32]V. Valeau, J. Picaut, A. Sakout, and A. Billon, "Simulation of the acoustics of coupled rooms by numerical resolution of a diffusion equation," in *Proceedings of the 8th International Congress on Acoustics, Kyoto* (Science Council of Japan, Japan, 2004).
[33]N. Xiang and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian parameter estimation," J. Acoust. Soc. Am. **110**(3), 1415–1424 (2001).

On the use of poroelastic materials for the control of the sound radiated by a cavity backed plate

François-Xavier Bécot[a] and Franck Sgard
Laboratoire des Sciences de l'Habitat, DGCB URA CNRS 1652,
Ecole Nationale des Travaux Publics de l'Etat, 69518 Vaulx-en-Velin Cedex, France

(Received 13 October 2005; revised 15 May 2006; accepted 19 May 2006)

The purpose of this paper is to examine the potential of poroelastic materials to control the low frequency noise radiated outside a parallellepipedic cavity enclosing a point source. The enclosure consists of five rigid walls and one flexible plate, all of which may be treated with a porous slab. The Biot-Allard theory, three equivalent fluid approaches and a locally reacting assumption are used to model the porous medium. The response of the system is calculated using a finite element model for all the components. The two issues addressed are the modeling of a porous material in a complex structure and the control of the sound radiated outside the cavity. Concerning the first point, calculations confirmed the validity range of the locally reacting assumption and prove the relevance of a limp porous model for unbonded plate treatments. Regarding the second issue, the sound power reduction obtained with the treatment of nonvibrating walls is compared to that achieved when treating the plate. The efficiency of the different mounting conditions of the porous slab to the plate is also discussed. Finally, the calculation of the dissipated powers inside the system provides a crucial information to optimize the sound absorbing treatment.
 [DOI: 10.1121/1.2214134]

PACS number(s): 43.55.Dt, 43.55.Ev, 43.50.Gf [NX] Pages: 2055–2066

I. INTRODUCTION

The two issues considered in the present paper are the modeling of poroelastic media in a complex structure and the control of the sound radiated by a cavity backed plate using poroelastic materials.

Systems consisting of a rectangular flexible panel coupled to a parallelepiped cavity are found in many problems of automotive, aircraft or industrial machinery applications. Vehicle cabin, aircraft fuselage or enclosures of a compressor are typical examples of this problem.

Thus, it is not surprising that, this configuration has been under the scope of numerous research works for many years. A good review of these works can be found for instance in Ref. 1. Among all of them, one should cite the reference works in Refs. 2 and 3, which examine the free vibrations of the coupled system based on the modal description of the plate *in-vacuo* and of the rigid walled cavity. Regarding the forced response, most of the studies assume an external sound source, which may be airborne[4,2,3,5,6] or structure borne[7,8] and are interested in calculating the pressure inside the cavity or the transmission loss factor associated to the vibrating panel.

All these works are based on an analytical modal approach, whose efficiency decreases at high frequencies or when there is a strong coupling between the plate and the cavity. In these cases indeed, the number of modes needed for the result convergence is so high that calculation times become prohibitive. Work in Ref. 8 makes use of pseudo-static corrections to overcome this difficulty but a modal approach would still be unadapted for our purpose because of the reasons explained further in the text.

On the other hand, only a few studies examine situations where the noise source is located inside the cavity. Most of the works found in literature consider plate/cavity systems having one or more apertures.[9–12] These apertures are needed, e.g., for the engine transmission or to allow for the flow of matter through the enclosure. The main purpose is then to control the inner cavity resonance, often called Helmholtz resonance, which can contribute significantly to the noise radiated outside the cavity. These studies are mainly based on pure boundary element methods (BEM) or coupled boundary element/finite element methods (BEM/FEM).

To control the sound field inside the cavity or radiated outside the plate/cavity system, porous materials ranging from mineral or glass wool to polymer foams are used. Four approaches are available to account for the dissipation induced by the porous material, three of which are examined in the present work.

First, dissipation can be introduced via the structural damping factor of the fluid inside the cavity.[7] In this case, there is no information on the exact location of the porous material inside the cavity. The second approach consists in applying an impedance boundary condition on the cavity walls. For example, this simplified boundary has been used in modal approaches[1] or in BEM (Refs. 9 and 12) and in FEM (Ref. 13) methods to account for dissipation induced by an absorptive material. This approach, also used in the present paper (see Sec. III C), proves to be valid in situations where the porous material is attached to a nonvibrating cavity wall.[13]

[a]Author to whom correspondence should be addressed. Electronic mail: becot@entpe.fr

The third and fourth approaches are based on three-dimensional representations of the porous material and are classically implemented in finite element models. The third approach considers that the porous material can be assimilated to an equivalent fluid[14] Within this approach one should distinguish the motionless skeleton assumption[15] and the limp skeleton assumption[15,16] (see Sec. III B). In both representations, the porous skeleton is assumed to be rigid: in the first assumption, it is not allowed to move, and in the second one it moves as a whole, including thus the inertial effects.

The fourth and last approach is based on Biot theory[17,18] and leads to the more accurate representation of the sound propagation inside the porous medium.[14] Models of the first type, referred to as $(\underline{u},\underline{U})$ formulation, use the solid phase displacement vector $\underline{u}$ and the fluid phase displacement vector $\underline{U}$ as field variables.[19] Models of the second type use the interstitial fluid pressure p instead of the fluid displacement vector. Several variations of this formulation have been proposed.[20,21] Referred to as $(\underline{u},p)$ formulations, they were introduced mainly as a more efficient method on the numerical point of view because the number of degrees of freedom goes down from 6 in the $(\underline{u},\underline{U})$ formalism to 4 in the $(\underline{u},p)$ formulation.

Poroelastic finite element formulations associated with $(\underline{u},\underline{U})$ and $(\underline{u},p)$ descriptions have been proposed to solve the problem of a plate/cavity system involving a porous material. For instance, a $(\underline{u},\underline{U})$ formulation was used in Ref. 13. The advantage of using a $(\underline{u},p)$ formulation have been discussed in details in Ref. 20. In the case of the initial formulation as proposed in this latter work, the coupling of the porous medium with an acoustic domain or another poroelastic domain is natural in the sense that no additional coupling matrix is needed and only the continuity of pressure must be ensured.[22] This factor, together with the reduction of the number of degrees of freedom, leads to important saving in memory storage and calculation times. The main difficulty in using linear poroelastic finite elements is that there is no strict criteria ensuring convergence of the result.[15,23] The required mesh size may vary depending on the material, on the excitation and on the vibroacoustic indicator of interest. Therefore, the final mesh should be chosen after a series of test calculations by progressively increasing the number of finite elements used.

Besides this difficulty, the relevance of using a finite element model is justified by the frequency range of interest of the present study. Noise enclosures are inefficient mainly at low frequencies, i.e., where porous materials have low absorption properties and the transparency of the vibrating panel is high. In this context, a finite element model of the complete plate/cavity system including the porous material seems suitable. For the porous medium, a $(\underline{u},p)$ formulation is retained to decrease the computational times. It should be underlined that the external fluid loading on the vibrating plate is neglected in the present study because the plate outer face radiates in air.[2,1] Note finally that the problem of interest can be solved using a modal technique provided that appropriate generalized complex modal basis are used for the porous material.[24,25]

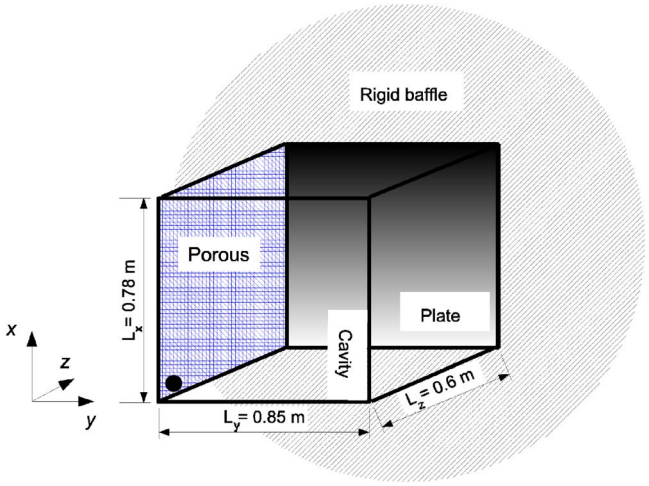


FIG. 1. (Color online) Scheme of the system: plate backed by an air cavity.

The paper is thus organized as follows. The problem objectives are presented in the next section. The model of the porous medium based on the Biot-Allard poroelasticity equations is then presented. In addition, three alternative models of the porous material are introduced with a view to reduce the calculation times. The complete finite element model is then described and different configurations with and without sound absorbing treatments are tested and compared. In the last part of the paper, dissipation mechanisms involved in each configuration are identified in order to optimize the noise control treatment.

II. PROBLEM DESCRIPTION

The geometry of the system studied in this paper is depicted in Fig. 1. It comprises a rectangular plate, a parallelepipedic cavity and a porous slab, altogether coupled.

The plate, also termed panel in the following, is simply supported. Its upper surface is embedded into an infinite rigid baffle and radiates sound into a semi-infinite medium of characteristic impedance $\rho_a c_a$. Its lower surface is coupled to an air-filled cavity. Unless mentioned, the walls of the cavity are acoustically rigid. The plate is excited by the sound field generated by a harmonic acoustic point source located in the corner of the cavity, at position $(x,y,z)=(0,0,0)$. The poroelastic material of interest is a mineral wool used for thermal insulation. It may be placed on the rigid walls of the cavity or attached to the panel. In this latter case, the porous layer is unbonded or directly bonded onto the plate. In all cases, sliding motion boundary condition are imposed on the lateral faces of the porous layer. The modeling of the sound propagation inside the porous material is treated in the next section.

To control the sound radiated by the plate in the semi-infinite space, one strategy is to control the sound which impinges on the plate surface by reducing the sound pressure level inside the cavity. This can be achieved by placing the porous layer on one of the cavity walls. The second strategy consists in controlling the sound radiated by the plate outside the cavity by attaching the porous layer to its inner surface. In this case however, the plate vibrational properties are more or less affected depending on the nature of the porous medium. This may lead to the situation where the noise re-

duction obtained by reducing the sound field inside the cavity is masked by the increase of the vibrational level of the plate in a given frequency range.

Hence, the optimum noise control solution will ideally take advantage of both the sound reduction inside the cavity and the modification of the plate vibrational properties. Therefore, the vibroacoustic indicators of interest here are the averaged quadratic plate velocity, the averaged quadratic cavity pressure and the sound power radiated in the semiinfinite space outside the cavity. The radiation efficiency of the plate is also examined and discussed in the present study but it is not shown here for the sake of shortness.

III. POROELASTIC MODELING: FIVE DIFFERENT APPROACHES

A poroelastic material comprises both a solid phase and a fluid phase, which are able to interact through elastic, inertial, viscous, and thermal effects. In the present paper, five approaches are used to account for the presence of the material inside the cavity. Except for the approach using the locally reacting surface assumption (see Sec. III C), all these approaches are based on a $(\underline{u},p)$ formulation of the Biot-Allard poroelasticity equations presented in Ref. 20. The basis of Biot's theory assumes that the poroelastic material is homogeneous on a macroscopic scale, i.e., for wavelengths very large compared to a representative elementary volume of the material. The problem is expressed from two coupled equations: the equation of motion of the solid phase and of the equation of motion of the fluid phase.[14] To solve this problem, a weak integral form is associated to these equations (see for instance Ref. .20).

A "direct" finite element discretization of the equations obtained gives the first model considered here, which is called the *poroelastic model*. The second and third models consider the poroelastic as a rigid frame medium. The second one assumes the poroelastic material skeleton as being motionless and the third one considers the skeleton as having no stiffness. In the fourth model, the presence of the poroelastic material inside the cavity is accounted for using an impedance boundary condition. Each approach is briefly recalled in the following. In these equations, a $e^{+j\omega t}$ time dependency is used.

A. The Biot-Allard poroelasticity equations

The weak integral form of the Biot-Allard poroelasticity equations can be found in Ref. 20 and writes

$$\int_{\Omega_p} [\underline{\underline{\tilde{\sigma}}}(\underline{u}):\underline{\underline{\varepsilon}}(\delta\underline{u}) - \tilde{\rho}\omega^2\underline{u}\cdot\delta\underline{u}]d\Omega + \int_{\Omega_p}\left[\frac{\phi^2}{\omega^2\tilde{\rho}_{22}}\underline{\nabla}_p\cdot\underline{\nabla}\delta_p - \frac{\phi^2}{\tilde{R}}p\delta_p\right]d\Omega - \int_{\Omega_p}\tilde{\gamma}\delta(\underline{\nabla}_p\cdot\underline{u})d\Omega - \int_{\partial\Omega_p}[\underline{\underline{\tilde{\sigma}}}\cdot\underline{n}]\cdot\delta\underline{u}d\Gamma + \int_{\partial\Omega_p}\left[\tilde{\gamma}\underline{u}\cdot\underline{n} - \frac{\phi^2}{\omega^2\tilde{\rho}_{22}}\frac{\partial p}{\partial n}\right]\delta p\, d\Gamma = 0 \quad \forall\,(\delta\underline{u},\delta p). \tag{1}$$

Ω_p is the poroelastic domain and $\partial\Omega_p$ its boundary surface. $\delta\underline{u}$ and δp represent admissible variations of the solid phase displacement vector $\underline{u}$ and of the interstitial pressure p, respectively. $\underline{\underline{\tilde{\sigma}}}$ and $\underline{\underline{\varepsilon}}$ are the in-vacuo stress and strain tensor of the skeleton of the porous material, $\underline{\underline{\tilde{\sigma}}}$ is related to the total stress tensor of the material $\underline{\underline{\sigma}}^t$ by the relation, $\underline{\underline{\tilde{\sigma}}}=\underline{\underline{\sigma}}^t+\phi(1+\tilde{Q}/\tilde{R})p\underline{\underline{1}}$.

In this equation, the superscript "$\tilde{}$" indicates that the variable is frequency dependent. $\tilde{\rho}$ is a modified density given by $\tilde{\rho}=\tilde{\rho}_{11}-\tilde{\rho}_{12}^2/\tilde{\rho}_{22}$, where $\tilde{\rho}_{11}$, $\tilde{\rho}_{22}$ and $\tilde{\rho}_{12}$ are the modified Biot's densities accounting for viscous dissipation. $\tilde{\gamma}$ is a coupling factor given by $\tilde{\gamma}=\phi(\tilde{\rho}_{12}/\tilde{\rho}_{22}-\tilde{Q}/\tilde{R})$. $\tilde{Q}$ is a factor which couples the skeleton strain to the fluid strain, and $\tilde{R}$ can be interpreted as the bulk modulus of a volume of fluid occupying a fraction ϕ of the porous media, ϕ being the porosity of the medium.

As shown in Ref. 21, Eq. (1) is particularly suited to couple a poroelastic material to a fluid domain provided that $\phi(1+\tilde{Q}/\tilde{R})\approx 1$. In this case indeed, no coupling matrices are needed and only the continuity of pressure must be ensured at the interface porous/fluid. When coupling the porous material to an elastic medium, another formulation is used to avoid the calculation of coupling matrices. This formulation, not given here, can be found in Ref. 21. Using this formulation, only the continuity of the solid displacement must be ensured,[22] which leads to a significant reduction of the size of the system matrices.

This model leads to the most comprehensive description of the phenomena in the sense that structural, thermal, and viscous dissipation are considered in this approach (see Sec. III D). This is also the most demanding in terms of computational effort. To overcome this, a number of approximated models have been developed which are presented in the following.

B. Equivalent fluid approaches

If the porous material is bonded onto a vibrating plate, the porous frame displacement cannot be neglected. However, in some situations, the effect of the solid phase deformation on the material response can be neglected and the problem can be formulated in terms of fluid pressure only. This is, for example, the case for very heavy and very stiff skeleton materials bonded onto a nonvibrating surface and acoustically excited. This greatly reduces the size of the finite element system to be solved and thus leads to a more efficient numerical implementation compared to the poroelastic model.

Motionless skeleton model: For highly stiff and dense materials, $\underline{u}$ and $\underline{\underline{\varepsilon}}$ equal zero. In this case, the fluid pressure inside the porous material satisfies a standard Helmholtz equation with propagation constants denoted as $\tilde{\rho}_e$ and $\tilde{K}_e$. In this case, the associated weak integral is [see Eq. (1)]

$$\int_{\Omega_p}\left[-\frac{\phi}{\omega^2\tilde{\rho}_e}\underline{\nabla}p\cdot\underline{\nabla}\delta p - \frac{\phi}{\tilde{K}_e}p\delta p\right]d\Omega - \int_{\partial\Omega_p}\frac{\phi}{\omega^2\tilde{\rho}_e}\frac{\partial p}{\partial n}\delta p\, d\Gamma = 0 \quad \forall\,\delta p. \tag{2}$$

In the equations above, $\tilde{\rho}_e=\tilde{\rho}_{22}/\phi$ and $\tilde{K}_e=\tilde{R}/\phi$ are the dynamic density and the bulk modulus of the equivalent fluid. This approach is expected to be valid for the treat-

ment of the cavity walls, which do not vibrate. This corresponds to the *motionless skeleton* model. Results in Ref. 15 prove that this approach could be successfully used for the prediction of transmission losses of multilayered systems involving poroelastic materials in the case that the latter are not directly bonded to the vibrating parts of the system.

Rigid body and limp model: Two other equivalent fluid approaches exist which are valid when the inertial forces dominate the elastic ones or when the elastic stresses are weak. The first approach can typically be used when a porous material is placed in front of the vibrating panel without being coupled to it. In this case, the porous slab may move as a whole while still being undeformed. This assumption, referred to as the *rigid body* model, implies that div $u=0$. This model is different from the second approach which assumes that the bulk modulus of the porous skeleton are close to zero. Referred to as the *limp* porous model,[26] this model is better suited for materials having low Young's modulus like glasswool. In both the *rigid body* and the *limp* models, the interstitial pressure satisfies a Helmholtz equation and the weak integral forms for these approaches are similar to Eq. (2). In these models though, different expressions of the dynamic densities need to be substituted. We have, for the *rigid body* model $\tilde{\rho}_{\text{eq}}^{\text{rigid body}} = (1/\tilde{\rho}_e + (\tilde{\gamma}^2 + (1-\phi)\tilde{\gamma})/\phi\tilde{\rho})^{-1}$ and for the *limp* model, $\tilde{\rho}_{\text{eq}}^{\text{limp}} = (1/\tilde{\rho}_e + \tilde{\gamma}^2/\phi\tilde{\rho})^{-1}$. From these equations, it is clear that the two assumptions lead to the same model for materials having a porosity close to one. Therefore, in the following, only the *limp* porous model has been used.

The crucial difference between the *motionless frame* hypothesis and the *limp* model is that inertial and damping effects of the solid phase are included in the latter model. These differences are exemplified in Sec. VIII. Note that these two models are not valid if resonances associated to the fiber motion are present in the frequency range of interest because stiffness effects are not accounted for. If this is the case, work in Ref. 16 proposes a modified limp porous model which includes the stiffness of the corresponding resonant mode; this model will however not be discussed further in the present work.

C. Impedance approach

A last approximation consists in accounting for the dissipation due to the poroelastic material inside the cavity by using a normal impedance boundary condition at the surface of the cavity walls. This approach was proved to be valid for the prediction of the pressure inside a cavity having one rigid wall treated with a porous material.[13] The impedance distribution is taken to be constant on the entire treated wall surface. In this case, the problem is solved in terms of the cavity pressure and of the panel displacements only. Numerically speaking, this approach leads to the most efficient implementation of the present problem.

At the surface of a cavity wall, the well-known boundary condition is $\partial p_a/\partial n = -\mathrm{j}\omega\rho_a p_a/Z_p$ where Z_p denotes the surface impedance of the porous medium and n the surface outward normal.

D. Expressions of the dissipated powers

An interesting feature-of the approach used in the present paper is that the power balance in the porous material can be expressed directly from the modified form of the Biot-Allard poroelasticity equations, as shown in Ref. 27. The calculation of these expressions are quite lengthy and the interested reader is referred to Ref. 28, Appendix A, p. 245, and Ref. 29 for the detailed expressions of the dissipated powers.

As a consequence of using this approach, the different dissipation mechanisms can be identified and quantified which is of great interest for the optimization of the material properties. The mechanisms responsible for the dissipation of energy may be structural damping, viscous or thermal effects in the porous layer and structural damping in the plate.

For a harmonic excitation, the injected power equals the dissipated power. On the one hand, power is injected into the porous layer via the solid phase and the fluid phase. On the other hand, the dissipated power is the superposition of the powers dissipated due to structural damping of skeleton, viscous effects, and thermal effects in interstitial fluid, the expressions of which can be found in Ref. 27.

IV. MODELING OF THE SYSTEM

This section briefly presents the model assembling. In addition to the poroelastic material, the system examined here contains an elastic medium, the plate, and an acoustic medium, the air cavity.

A standard displacement formulation is used for the plate and the pressure inside the cavity satisfies the Helmholtz equation. The weak integral forms associated to the plate and the cavity are classical and can be found for instance in Ref. 30. Note that, as indicated in Table I, structural damping is accounted for in the plate by considering a complex Young's modulus with $\tilde{E}=E(1+j\eta_s)$.

In addition, when there is no porous material inside the cavity, dissipation in the fluid is accounted for using a complex sound speed $c_a(1+j\eta_a)$.

The entire system is finally assembled using the appropriate coupling conditions between the different components. The detailed coupling equations can be found in Ref. 22. As previously mentioned, the selection of an appropriate formulation for the porous medium (see Sec. III A) avoids the computation of coupling matrices and reduces the calculation times.

The weak integral form in each domain is then discretized using finite elements and the problem is solved for the corresponding nodal variables in each domain. For the air inside the cavity and for the poroelastic material, eight noded linear elements are used; for the plate, four noded thin shell elements are used.

In the frequency range of interest, up to 600 Hz, the highest *in-vacuo* mode of the simply supported plate is the mode (8,8). Using six elements per wavelength, the final mesh contains 24 elements in both the x direction and the y direction. The mesh is extruded, which means that the discretization in the x and y direction is the same for all the system components.

TABLE I. Properties or the system components.

Aluminium plate	
Mass density	$\rho_s=2800$ kg/m^3
Young's modulus	$E_s=75.10^9$ Pa
Poisson's ratio	$\nu_s=0.3$
Structural damping	$\eta_s=0.01$
Mineral wool	
Flow resistivity	$\sigma=135$ kNs/m^4
Porosity	$\phi=0.94$
Tortuosity	$\alpha_\infty=2.1$
Viscous characteristic length	$\Lambda=49$ μm
Thermal characteristic length	$\Lambda'=166$ μm
Skeleton mass density	$\rho_1=175$ kg/m^3
Skeleton Young's modulus	$E=4\,400\,000$ Pa
Poisson's ratio of the skeleton	$\nu=0$
Skeleton structural damping	$\eta=0.1$

For the description of the cavity in the z direction, 21 elements are used. This latter mesh size is overestimated but is taken to be coherent with the mesh in the other two directions of space. Up to 600 Hz indeed, the highest rigid walled cavity uncoupled modes is mode (2,2,2).

When the porous material is modeled using the Biot-Allard poroelasticity equations or using one of the equivalent fluid approaches, eight finite elements are taken along the material thickness.

V. TYPICAL RESPONSE OF THE PLATE/CAVITY SYSTEM

The material properties of the plate is given in Table I. The plate thickness is 2 mm and for the cavity, $\rho_a=1.213$ kg/m^{-3} and $c_a=342.2$ m/s^{-1}. In this section, no sound absorbing treatment is present inside the cavity. However, to avoid infinite values of the resonance peaks, dissipation in the cavity is introduced using a complex sound speed as described previously.

It is worth noting that the finite element code has been validated by comparisons with the modal approach presented in Ref. 3. Values of the averaged quadratic -cavity pressure and the averaged quadratic plate velocity, not shown here for sake of concision, are in perfect agreement on the entire frequency range. 16 eigenfrequencies of the plate *in vacuo* are found below the first nonzero cavity uncoupled mode located at 201 Hz. Below 600 Hz, 17 modes of the stand alone cavity can be identified.

The sound power radiated by the cavity backed plate, not shown for sake of shortness, has been computed and is compared to the case where the plate is embedded into an infinite baffle and radiates into two semi-infinite media, i.e., without being coupled to the cavity.

First, for the system examined here, the Schröeder frequency is estimated to be around 2000 Hz. Therefore in the frequency range of interest, one could expect a modal behavior of the system. In fact, strong coupling occurs mainly in the case of a shallow cavity or in the presence of heavy fluids like water or oil.[2,1] This is confirmed by the simulation results because all the cavity uncoupled modes below 600 Hz can be identified. All of them are slightly shifted towards high frequencies by about 2 to 3 Hz compared to the frequencies of the uncoupled cavity modes (an analogous frequency shift has been reported in Ref. 7 for a similar plate/cavity system). As a result, the radiated power is mainly governed by the cavity controlled modes, which mask the contribution from the plate controlled modes, leading to an increase of about 20 dB of the third octave band values compared to when the plate radiates in the free field. For noise control purposes, this means that the treatment of the cavity using the previously described porous material is expected to have a significant effect on the radiated sound in this frequency range.

Below 200 Hz however, i.e., below the first cavity controlled mode, the radiated power is mainly governed by plate controlled modes. In this frequency range, the sound field inside the cavity is almost uniform and the plate modes with even order (symmetric modes) are only weakly excited. Moreover, the frequency of the first plate controlled mode is shifted towards higher frequencies due to a stiffness effect induced by the cavity.

At this frequency, the system behaves as a mass-spring system, the mass being the plate and the spring including the stiffness of the plate and that of the air contained in cavity. As shown in Ref. 2, this resonance frequency can be found quite accurately by using the concept of direct acoustic stiffness. This additional stiffness represents the coupling between the different plate modes which is due to the stiffness of the fluid contained in the cavity. According to this work, the resonance frequency can be written $f=1/2\pi\sqrt{(K_{11}+K_{11}^{(11)})/M_{11}}$ where K_{11} is the generalized stiffness of the plate, $K_{11}^{(11)}$ is the direct acoustic stiffness and M_{11} is the generalized mass of the plate, the expressions of which can be found in Ref. 2. This gives $f=30.4$ Hz, which agrees well with the value found graphically at around 30 Hz. Therefore, in this frequency range, additional dissipation in the cavity fluid, e.g., by means of an absorbing material lined on the cavity walls as proposed in this paper, is expected to have little effect on the radiated power. Actually, this solution influences the power injected to the plate and thus may modify the vibrational properties of the plate. However, the direct treatment of the plate is expected to be more efficient in this frequency range.

VI. ACOUSTICAL ABSORPTION

The poroelastic material used in the present work is a mineral wool of 6 cm thickness used for thermal insulation. To quantify the absorption properties of this material, a measure in a standing wave tube is simulated. In addition, the normal surface impedance, which is needed in the impedance approach, is calculated using the simulated values of the reflection coefficient. For this experiment, the porous substrate is modeled with eight noded poroelastic linear elements using the Biot-Allard description of the material. The convergence of the solution has been checked by performing a series of runs with an increasing number of finite elements.

In the frequency range of interest, the porous material has reasonable absorption properties. The absorption coefficient increases uniformity from 0.3 at 200 Hz to nearly 0.5 at 600 Hz. When dealing with poroelastic material, it is of in-

terest to identify the quarter of wavelength resonance frequency associated with a structural resonance of the skeleton. At this frequency, the skeleton and the interstitial fluid are moving in phase which leads to a decrease of the viscous effects in the porous medium. Therefore, the absorption coefficient drops at this frequency for materials with low structural damping but may exhibit a peak for materials having a high skeleton loss factor.

For the material used in the present study, the absorption coefficient increases at the resonance frequency though it does not occur in the frequency range examined here. For sliding boundary conditions indeed, an estimation of this frequency gives 661 Hz according to Ref 14. This value corresponds well with the one found by numerical results around 670 Hz.

In the following, the material is assumed to be locally reacting. This hypothesis is tested and discussed in the next section. Note that at a given frequency, the value of normal impedance is constant over the entire surface of the treated wall.

VII. TREATMENT OF THE CAVITY WALLS

In this section, the sound absorbing treatment is placed on one of the nonvibrating walls of the cavity and the only dissipation in the system is due to the presence of the porous material.

A. Validity of the impedance approach

As a preamble, the different approaches to model the porous material presented in Sec. III are compared with the aim of reducing computational times.

The limp porous model has not been considered for this test because, since the porous layer is applied on a rigid wall, results are expected to be similar to those obtained using the motionless assumption. It is worth mentioning here that the implementation of the linear poroelastic finite elements have been validated by comparisons with measurement data in previous work.[31] Experimental validations of the complete plate/cavity model including porous materials have been presented in the case of point force excitation acting on the plate.[32] The mesh used in this test has been presented in Sec. II and the porous material is placed on the wall opposite to the plate ($z=0$ see Fig. 1).

As expected,[13] the three approaches gives very similar results on the entire frequency range for the plate velocity and for the radiated power (see Fig. 2). The largest deviations are observed for the cavity quadratic pressure at frequencies around 200 Hz and 300 Hz. The deviations observed, which may locally reach 3 dB, coincide with resonances associated with cavity modes in the direction parallel to the surface of the treated wall [modes (0,1,0), (1,0,0), and (1,1,0)]. Similar results are obtained when the treatment is applied on the other cavity walls except at frequencies corresponding to grazing incidence modes.

Moreover, at these frequencies, the examination of the dissipation due to viscous effects, not shown here for the sake of shortness, is underestimated using the equivalent fluid approach compared to when using the poroelastic model. Hence, a dissipation mechanism is missing in the equivalent fluid approach, and also in the impedance approach, at these frequencies. The mechanism which is missing in these two models is the dissipation related to structural effects in the direction parallel to the porous surface.

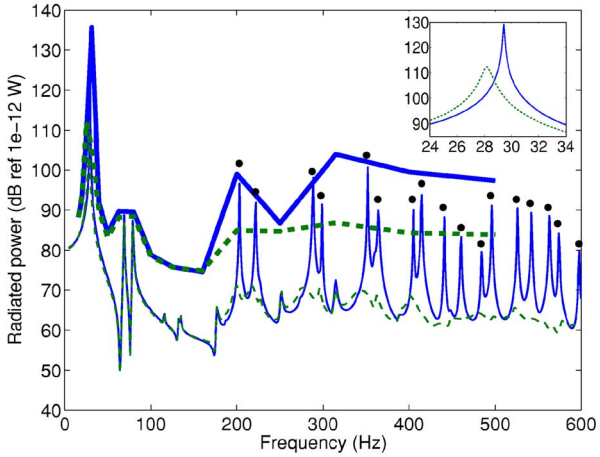


FIG. 2. (Color online) Radiated power when the porous layer is bonded onto the wall $z=0$. No treatment (—), with treatment (- - -).

This means that the sound field, which impinges with a grazing incidence on the treated cavity wall at these frequencies, generates a significant shear motion inside the porous material. In this case, the impedance distribution, which has been obtained by simulating a measure in a standing wave tube, i.e., under a normal excitation, does not account properly for the propagation of sound in the direction parallel to the porous surface.

In total, the impedance approach yields overestimated results, which means that the treatment efficiency is underestimated compared to the Biot-Allard theory. Therefore, as far as the radiated power is concerned, it seems safe to consider only the impedance approach when the poroelastic material is attached to a nonvibrating wall. Note that identical conclusions were obtained in Ref. 13 for the inside quadratic pressure for a similar plate/cavity system, the dimensions of which were such that only one cavity controlled mode was present in the frequency range of interest.

B. Treatment of one cavity wall

In this section, the porous layer is bonded onto the wall $z=0$ and the results are compared to the case when there is no treatment inside the cavity. Results are shown in Fig. 2 for the radiated power.

As observed on this figure (see the zoom-up's in the corresponding plots), the first system resonance is shifted towards low frequencies and significantly damped. At this frequency, the effect of the absorbing material inside the cavity corresponds to a decrease of the apparent stiffness of the aforementioned mass/spring system. The reduction of the peak level is due to the reduction of the injected power to the plate. Therefore, a reduction of the radiated sound could be achieved in a frequency range controlled by in-vacuo plate modes without treating directly the plate surface. However, this phenomenon may be difficult to illustrate experimentally because it occurs at very low frequencies (below 30 Hz) where measurements are usually inaccurate.

At higher frequencies and up to the first cavity controlled mode, the presence of the sound absorbing treatment has no noticeable effects as expected. Beyond 200 Hz however, the resonance peaks observed are significantly damped. The examination of the radiation efficiency, not shown in the present paper, shows the same trend: it is not significantly modified below 200 Hz while it is somewhat reduced at higher frequencies. All these effects lead to a significant reduction of the radiated power in both "continuous" spectrum and in third octave band values (thicker lines in the plot) at frequencies beyond 200 Hz. This reduction is substantial at all frequencies up to 600 Hz and reaches around 15 dB in third octave band values at 500 Hz.

C. Treatment optimization

Influence of the treatment position: In order to optimize the sound absorbing treatment, it is of interest to examine the influence of the treatment position inside the cavity. Figure 3 compares third octave band values of the radiated power for different locations of the porous slab inside the cavity. As previously, the first peak is shifted towards low frequencies compared to when the system is not treated. Hence, the energy of the first resonance peak splitted into two contiguous third octave bands, the values of which appear then lower than the corresponding third octave band values for the untreated case. But in fact, before the first cavity controlled mode appears, none of the treatments give a noticeable noise reduction.

At higher frequencies and up to the 315 Hz band, the reduction of the radiated power is similar for all the treatment configurations. At frequencies above 315 Hz, i.e., when half of the sound wavelength in air becomes comparable to the dimension of the treated surface, the treatment of the surface $z=0$ gives a slightly larger noise reduction than the other two configurations. This could be expected as this configuration corresponds to the largest treated surface area and thus to the largest volume of the porous material inside the cavity. Hence this verifies the statement of the larger the quantity of porous material inside the system, the larger the reduction. The two series of calculations shown in the following give further insights concerning this point.

Treatment of two walls: The first series of calculations shows radiated powers in the case where two walls are treated simultaneously and results are compared to the case where only the wall $z=0$ is treated (see Fig. 4). As expected, all treatment configurations give similar power levels below 315 Hz because the sound absorbing treatment has no significant effect in this frequency range, whatever the treatment configuration. The picture is slightly different at higher frequencies. Inside the plot, a zoom-up on the frequency range from 315 Hz to 500 Hz is displayed. It shows that the reduction is larger in each configuration where two walls are treated simultaneously compared to when only the wall $z=0$ is treated.

However the gain is weak, at most 2 dB at 500 Hz. This proves that the treatment of one well chosen cavity wall may give similar noise reduction as a treatment of two walls. This is illustrated in the 315 Hz third octave frequency band where three modes are in the z direction out of the four modes present in this band. It is observed that the treatments involving walls in the x and in the y direction (see curves for $x=L_x$ and $y=0$, $y=0$ and $y=L_y$, $x=L_x$ and $y=L_y$) give noise reductions which are similar to that achieved by the treatment of the single wall $z=0$. Therefore, the choice of the treatment location should be addressed according to the direction of the mode which contributes the most in the frequency region targeted.

Moreover, the noise reduction achieved by applying the treatment on the walls $y=0$ and $y=L_y$ is lower than that achieved by placing the treatment on walls $x=L_x$ and $y=0$ for instance. This corresponds to the well-known result from architectural acoustics which states that in a room, the treatment of two walls opposite to each other yields a lower sound reduction than applying the treatment on two consecutive walls. In fact, this corresponds to control a larger number of modes because two directions of space are involved instead of one. Note however that the treated surface area for these two configurations is different: the surface area for the treatment of walls $x=L_x$ and $y=0$ is around 4% larger than that of the treatment of walls $y=0$ and $y=L_y$.

Finally, treatments involving the wall $z=0$ provides the largest sound reduction on the entire frequency range. Among these two configurations, the reduction of the radiated power is larger when the second wall involved is the wall $y=0$ than when the second wall is $x=L_x$. In the first configuration, the absorbing slabs are located next to the noise source. Thus, this treatment corresponds to controlling the radiation impedance of the source. This provides a more efficient sound absorbing treatment than the treatment of walls opposite to the source position.

Partial treatment of the wall surface: The second series of calculations investigates situations where only a portion of a nonvibrating wall is covered by the porous layer. For all calculations, the treatment is applied on the wall $y=0$ as an example. Similar conclusions are obtained if another wall is chosen. The porous slab is assumed to be rigidly backed and the outer surface of the porous layer is embedded into the otherwise non vibrating part of the wall. This means that there is no diffraction of sound at the material edges other than that due to the impedance discontinuity.

The partial treatments are characterized by a covering rate which is defined as the ratio of the porous layer surface area to the total wall area. Two covering rates are examined here: 83% and 54%. For a given covering rate, the patch of sound absorbing treatment may be compact and centered on the surface or located in the corner where the source is located. The third possible configuration consists in a number of impedance patches distributed randomly at the wall surface for the same treated surface area. The treatments are assessed in terms of the difference between the radiated power for a treatment covering the entire surface and that for a treatment covering only a portion of the surface. Therefore, negative values correspond to a deterioration of the sound power reduction. Moreover, only third octave band values are displayed here. Results are shown for a 83% covering rate on the top plot of Fig. 5, and for a covering rate of 54% on the bottom plot, for all patch configurations referred to as *center, corner*, and *random*.

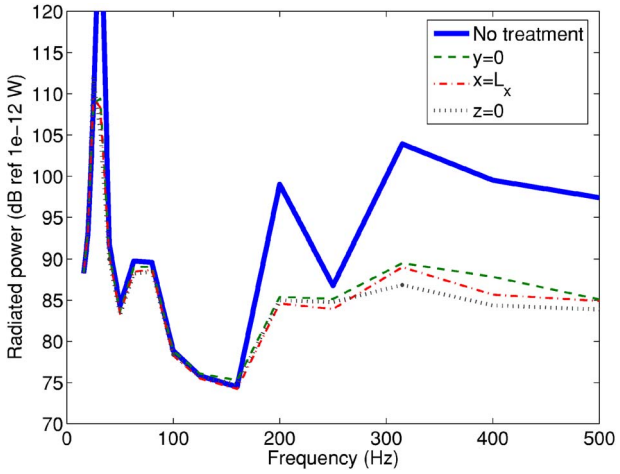


FIG. 3. (Color online) Influence of the treatment position on the radiated power when only one wall is treated.

For the two covering rates examined here, at the first system resonance, there is an improvement followed immediately by a diminution of the noise reduction. These variations are due to a slight frequency shift of the system resonance. On the rest of the frequency range, partial treatments of the surface give higher radiated power than the treatment of the entire surface. However, the deterioration does not exceed 2 dB for a 83% covering rate and for all configurations. In this case, a reduction of the treatment weight may be considered. For a 54% covering rate, the noise reduction may become worse by more than 5 dB for a compact distribution of the impedance patches (see results for *center* and *corner* situations). If a random distribution of the patches is used, the situation is only worse by about 3 dB up to 500 Hz, which may be acceptable regarding the important weight reduction obtained in this case. Therefore, for the two covering rates examined here, the best treatment of the partial surface would consist in distributing the impedance patches randomly on the wall surface. Details on the practical implementation of the solutions proposed here are beyond the scope of this paper.

Finally, these results show that the level of sound power

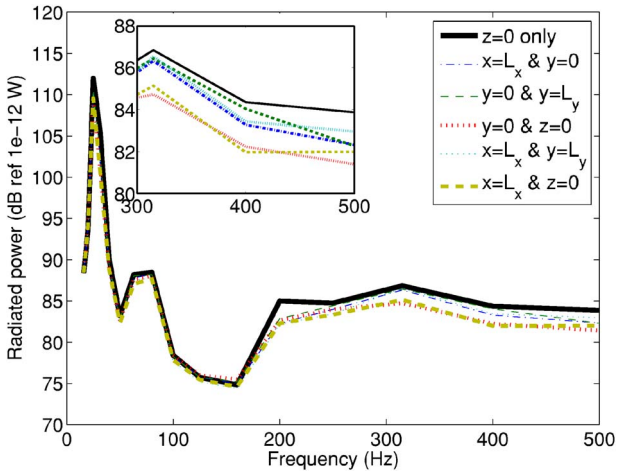
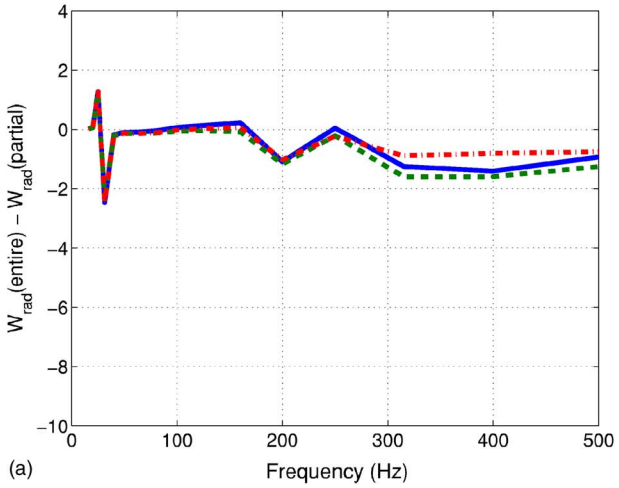


FIG. 4. (Color online) Influence of the treatment position on the radiated power when two walls are treated simultaneously.



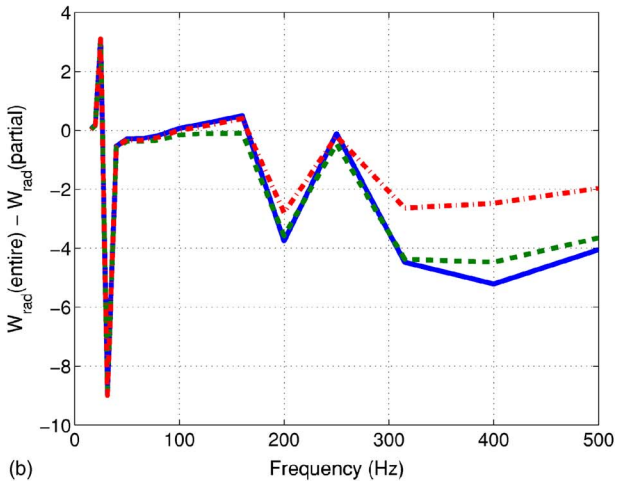


FIG. 5. (Color online) Noise reduction differential due to a partial treatment of the cavity wall $y=0$ compared to when the entire wall surface is treated. Covering rate is 83% (top) or 54% (bottom). Configuration, center (—), corner (- - -), random (---).

reduction is governed by the treated surface area, but also and to the same extent, by the distribution of the absorbing patches on the surface.

VIII. TREATMENT OF THE PLATE

In the light of previous results, besides a significant noise reduction above 200 Hz, the treatment of one or two cavity walls has very little effect at frequencies below 200 Hz, namely in the frequency range where the plate controlled modes dominate. Therefore the present section examines situations where the sound absorbing treatment is directly applied onto the plate. In the first paragraph, bonded and unbonded case conditions are compared with respect to the largest reduction of radiated power provided. In the second paragraph, the modeling approaches available in the unbonded case are assessed with the aim of diminishing the computational costs.

A. Influence of bonding conditions

In this section, either the porous layer is bonded onto the plate (*bonded* case) or a 3 mm air gap is inserted between the porous layer and the plate (*unbonded* case). In the first configuration, the porous solid phase is coupled to the plate,

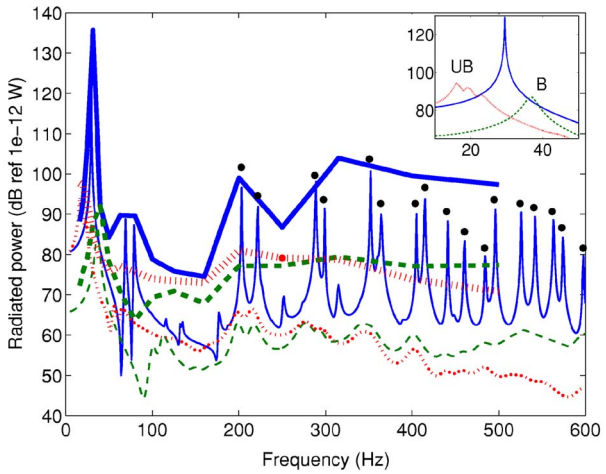


FIG. 6. (Color online) Comparison between the unbonded (UB) and bonded (B) conditions when the treatment is applied on the plate. Radiated power for three configurations: No treatment (—), bonded (- - -), unbonded, (···).

therefore the only relevant model is obtained using the Biot-Allard theory. The vibroacoustic indicators examined here are the same as for the treatment of a nonvibrating cavity wall. but only the radiated power is shown in Fig. 6. To make the observation at low frequencies easier, a zoom-up around the first system resonance, i.e., around 30 Hz, has been displayed inside the plot.

Contrary to the treatment of one cavity wall, the effect of the porous material can be seen for the two bonding conditions on the entire frequency range and for all indicators. As observed on the plot of the plate velocity, not shown here for the sake of conciseness, bonding the chosen material onto the plate corresponds to stiffening the structure whereas the unbonded condition corresponds to an added mass effect. The first system resonance is shifted upwards in the first case and downwards in the latter case. Moreover, the maximum level of the first system resonance is significantly reduced for both bonding conditions. This reduction reaches more than 30 dB in the unbonded case and around 40 dB for the bonded case at the resonance. Moreover, the bonded case gives a larger reduction of the vibrational level than the unbonded case at frequencies below around 300 Hz while the reverse situation is observed at higher frequencies. In both cases, the reduction remains above 30 dB and may reach 40 dB for the unbonded situation.

For the pressure inside the cavity, the reduction achieved is of the same order for the two mounting conditions. Therefore for the material used here, the value of this indicator depends mainly on the absorption of the material and to a minor extent on the mounting conditions of the absorbing layer. For this indicator, the reduction level is of the same order as for the treatment of a rigid cavity wall. When comparing to this latter configuration, the main gain of treating the vibrating plate is seen on the first system resonance, which is substantially damped by about 30 to 40 dB in the unbonded and the bonded case, respectively.

The plate radiation efficiency is also modified to a large extent. The number of lobes is decreased and, on the average, the radiation efficiency is increased compared to the untreated case. Above the first mode of the system with bonded conditions, the radiation efficiency of the plate is larger when the plate and the porous solid phases are decoupled than when they are coupled. Beyond 300 Hz however, there are only little differences between the two configurations. This means that the noise reduction achieved, which is larger for the unbonded case above 300 Hz, is mainly due to a reduction of the vibrational velocity.

Concerning the radiated power, the two configurations examined here give reductions which are substantial on the entire frequency range. In the bonded case, the values of the radiated power level out at frequencies above 200 Hz. On the contrary, the values in the unbonded case decrease uniformly beyond this frequency. This has for consequence that, while the reduction achieved in the bonded case surpasses that achieved in the unbonded case for frequencies below 300 Hz, the reverse situation is observed at higher frequencies. However, as mentioned in Ref. 13, this results may be valid only for porous materials having a large stiffness to mass ratio. In total, the reduction level achieved goes from a few decibels around 150 Hz to almost 30 dB at 500 Hz for the unbonded situation.

It is worth mentioning that these results coincide with the expertise from aircraft industries which states that the sound absorbing treatment should have a minimum area of contact with the vibrating structure in order to be efficient. However, the results show that this statement depends on the frequency range. For instance for the situations examined here, in the frequency band below the first nonzero cavity controlled mode, the bonded configuration gives a larger noise reduction than the unbonded configuration.

B. Comparison of modeling approaches

Since it is the configuration which gives the more interesting noise reduction on the entire frequency range, it is useful to compare the different modeling approaches available, in the case where the porous layer is unbonded to the plate.

In this situation, the plate and the porous layer skeleton are decoupled and the vibrations of the solid phase of the porous layer can be neglected. Therefore, the rigid frame approaches presented in Sec. III may be used and compared to the poroelastic model, which gives the reference results. Note that, in the rigid frame models, there is no need to insert an air gap between the porous layer surface and the plate.

Figure 7 shows the power radiated by the plate in this configuration. It is observed that the poroelastic model and the limp porous model give very similar predictions. This means that the dissipation mechanisms which are dominant in this configuration are included in the simplified limp porous model, i.e., viscous and thermal effects. This confirms the previous result in Ref. 15 obtained for the transmission loss prediction of multilayer systems. Furthermore, the motionless assumption yields predicted levels which follow but underestimate the radiated power at all frequencies. This shows that even when the porous layer is not bonded to the plate, the inertial forces must be accounted for in the prediction model. This is due to the fact that the poroelastic material used here has a high mass density. It was shown in Ref.

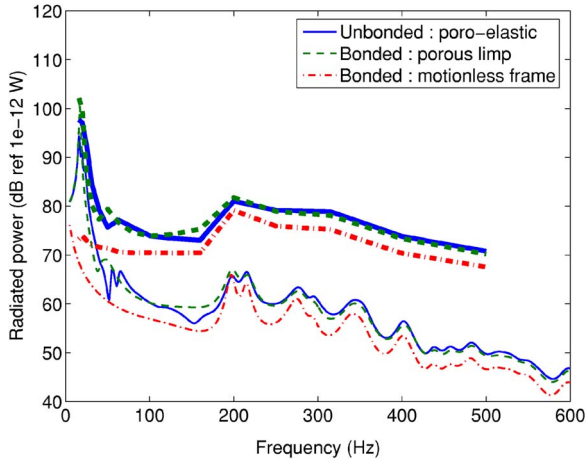


FIG. 7. (Color online) Radiated power using different modeling approaches when the porous layer is unbonded to the plate.

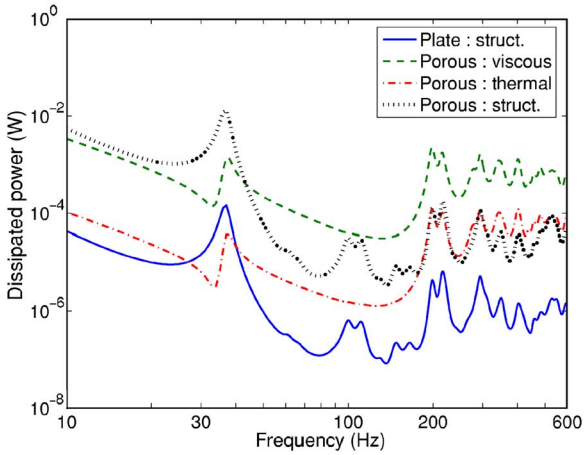


FIG. 9. (Color online) Dissipated powers in the system when the porous layer is bonded onto the plate.

33 that for materials having a lower mass density, e.g., fiberglass, the motionless frame assumption is accurate enough. Finally, this result is interesting on a numerical point of view because the limp porous model yields reduced size equation systems.

IX. DISSIPATION MECHANISMS

Now that the performances of the different treatment configurations have been compared it is interesting to try to identify the mechanisms involved in the energy dissipation. The powers dissipated by each mechanism calculated using the expressions given above are shown in Figs. 8–10. Note that for these figures a logarithmic frequency scale has been used to see better the repartition of the dissipated powers at low frequencies.

Figure 8 shows the dissipated powers inside the plate and inside the porous layer when this latter is applied on the wall facing the plate (wall $z=0$, see also Fig. 2). On most of the frequency range examined here, the viscous effects in the porous medium are dominant by at least a factor 10. Even at the first system resonance which is controlled by the first plate mode, viscous effects exceed the structural effects in the plate, as shown in the zoom-up around 30 Hz displayed inside the graph. The dissipation in the plate dominates only at very low frequencies, namely below 20 Hz and around 70 Hz and 80 Hz. These frequencies correspond to resonances of plate controlled modes (1,3) and (3,1). It should be underlined that these two modes contribute significantly to the noise radiated outside the cavity. In total for the plate/cavity/porous material system, dissipation occurs mainly inside the porous material by viscous and thermal effects.

The picture changes drastically at low frequencies when the porous layer is directly bonded onto the plate (see Fig. 9). In this configuration, the structural effects in the porous medium are dominant at low frequencies including the first system resonance. Above 40 Hz, viscous effects dominate all other dissipation mechanisms. Beyond 200 Hz, i.e., in the frequency range where cavity controlled modes are dominant, structural effects in the porous medium are of the same order of magnitude than the thermal effects, these two being 10 times larger than the dissipation due to structural effects in the plate. It is worth noting that, because the porous skeleton is directly coupled to the plate, the variation of structural effects in the plate closely follows those in the porous medium, though with a scaling factor of 100.

When the porous layer is decoupled from the plate surface (see Fig. 10), the viscous effects are dominant on the entire frequency range. However, at the first system resonance around 20 Hz, structural effects in the porous skeleton

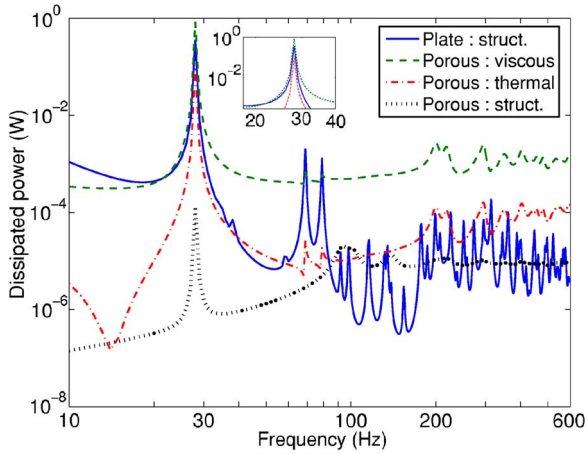


FIG. 8. (Color online) Dissipated powers in the system when the porous layer is bonded onto the cavity rigid wall $z=0$.

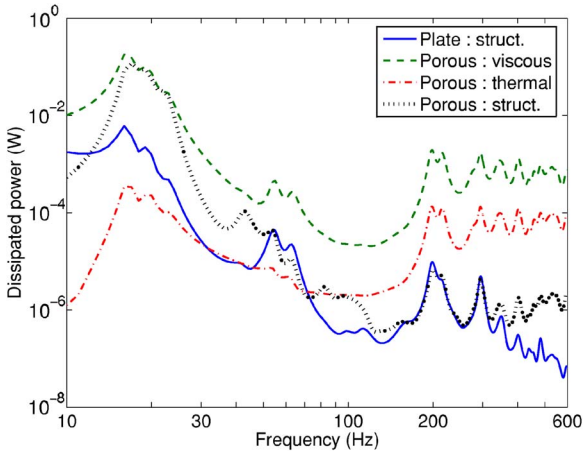


FIG. 10. (Color online) Dissipated powers in the system when the porous layer is unbonded to the plate.

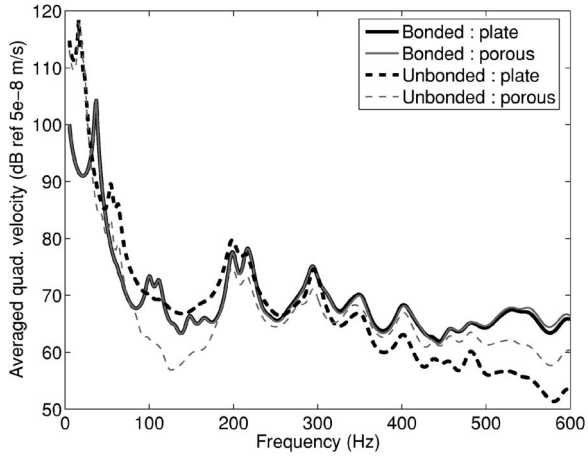


FIG. 11. (Color online) Comparison of bonded (—) and unbonded (- - -) configurations on the averaged quadratic velocities of the plate (thick lines) and of the porous material (thin lines).

are of the same order of magnitude as the viscous effects. From 300 Hz, dissipation due to structural effects in the plate drops again and becomes negligible compared to the dissipation mechanisms in the porous material. It is interesting to note that this frequency coincides with the frequency limit beyond which the unbonded configuration gives larger noise reduction than the bonded case.

In fact, results, not shown here, prove that the total dissipated powers in the system are equal for the two treatment configurations. As previously mentioned, viscous effects are responsible for the major part of dissipation in the system and are equal for the two configurations at frequencies above 200 Hz. Therefore, the sound radiation reduction is not due to an increase of dissipation in the system, in particular in the porous material.

To help interpret this change in the radiated power, the averaged quadratic velocities for the plate and for the porous material are presented in Fig. 11. It is observed that, in the case the porous material is bonded onto the plate, the velocity of the porous material and of the plate are equal in the frequency range in question here, i.e., above 300 Hz. In the unbonded case however, while being in the same order of magnitude between 200 Hz and 300 Hz, the plate velocity drops above 300 Hz and becomes negligible compared to the other vibrational levels. Moreover, in this frequency range a modification of the radiation efficiency cannot be responsible for this because they are of the same order of magnitude for the two treatment configurations. Therefore, the reduction of the sound power radiated by the plate is due to a reduction of the vibrational energy transmitted from the porous skeleton to the plate.

Note also that the velocity of the porous material is lower in the unbonded case than in the bonded case and that this difference is not followed by an increase of the dissipation due to viscous effects for the unbonded case. Moreover, at frequencies below 200 Hz, the porous material when bonded to the plate tends to damp the plate vibrations compared to the unbonded case. This means that at low frequencies, the porous material adds structural damping to the panel, leading to a more efficient noise treatment solution than the unbonded configuration.

X. CONCLUSION

In this paper, a system consisting of a baffled plate backed by a parallelepipedic rigid walled cavity has been studied. Numerous past works examined the reduction of the noise inside the cavity due to a source located outside the cavity or due to a point force acting on the plate. The main issue of this paper concerns the possibility of reducing the sound radiated outside the cavity from a source placed inside the cavity by using a porous material placed on the cavity nonvibrating walls or on the plate.

The response of the system is calculated using a finite element model for all components. The porous material is described by five different models: the complete mixed $(\underline{u},p)$ formulation of the Biot-Allard poroelasticity equations, a motionless skeleton assumption, a limp skeleton assumption, a rigid body skeleton assumption and an impedance boundary condition. All these models have been extensively used in previous works except for the rigid body model for which the authors are not aware of any previous reference.

When the porous material is applied onto rigid cavity walls, it has been shown that the material behaves mainly as if locally reacting. This result, which mainly confirms results from previous works, is important because it allows one to describe the porous material by a simple impedance boundary condition instead of using a three-dimensional model.

Given the dimensions of the system studied here, two main frequency regions are defined for the control of the noise emission. They are separated by the frequency of the first nonzero uncoupled cavity mode. Below 200 Hz, the sound radiation of the system is mainly governed by the plate controlled modes and above 200 Hz by the cavity controlled modes. This has for consequence that the treatment of the cavity walls has little effects at frequencies below 200 Hz. However, above this frequency, the noise reduction is substantial. It was also proved that the location and the surface of the sound absorbing treatment may be optimized according to, e.g., design constraints. In this respect, it was shown that a reduction of the treatment weight could be possible while keeping a treatment efficiency close to the original one.

To achieve a noise reduction on the entire frequency range, the porous material should be applied to the plate. Two mounting conditions have been examined: either the porous layer is directly bonded onto the plate or there exists an air gap between the porous slab and the plate. In the latter case, the plate and the solid phase of the porous medium are decoupled. As expected, a significant noise reduction is achieved on the entire frequency range with the treatment of the plate. In particular, numerical results show that the bonded condition leads to a larger noise reduction compared to the unbonded condition at frequencies below 300 Hz while the reverse situation is observed at higher frequencies. Moreover, when the material is not directly bonded onto the plate, it is shown that a porous limp model can be used instead to reduce calculation times of a complete Biot-Allard model, depending on the required degree of accuracy.

Finally, the assessment of the powers dissipated in the system proves that viscous effects in the porous material are

responsible for the major part of energy dissipation in the plate/cavity system with treatment. In addition, it is shown that at low frequencies, the bonded configuration provides a more efficient noise treatment because of an added structural damping to the vibrating panel. It is also proved that the sound radiation reduction observed in the unbonded case at higher frequencies is not due to an increase of dissipation inside the system, e.g., by viscous effects, but is due to a reduction of the vibrational level transmitted from the porous skeleton to the plate.

This study sheds a light on the ability of porous materials to decrease the sound transmission outside the enclosure and the practical implementations of the noise reduction treatments. In particular, modeling alternatives as well as propositions concerning the different mounting conditions, the locations and the distribution of the absorbing material are advanced. These two aspects constitute the main contribution of the present work to the current research effort for the modeling of the porous materials in complex systems.

Ongoing works involves the experimental validation of the model for the configurations examined here. In addition, further works are currently carried out to examine the potential of heterogeneous porous materials for this same purpose. For instance, the concept of double porosity has been proved to significantly increase the absorption of carefully selected porous material at low frequencies when measured in a standing wave tube. Their performance when integrated into a complex structure like a plate/cavity system and submitted to a more complex sound field still needs to be quantified. The study of their performance in transmission when coupled to a flexible structure together with the mechanisms involved in this process is also a subject for future research.

ACKNOWLEDGMENTS

This work is a part of CAHPAC research project (Capotage Acoustique Hybride Passif/Actif), which is supported by INRS (Institut National dela Recherche et de Sécurité) and CNRS (Centre National de Recherche Scientifique). The authors thank these contributors for their financial support.

[1]J. Pan and D. A. Bies, "The effect of fluid-structural coupling on sound waves in an enclosure—Theoretical part," J. Acoust. Soc. Am. **82**, 691–707 (1990).
[2]A. J. Pretlove, "Free vibrations of a rectangular panel backed by a closed rectangular cavity," J. Sound Vib. **2**, 197–209 (1965).
[3]E. H. Dowell, G. F. Gorman, and D. A. Smith, "Acoustoelasticity: General theory, acoustic natural modes and forced response to sinusoidal excitation, including comparisons with experiment," J. Sound Vib. **52**, 519–542 (1977).
[4]R. H. Lyon, "Noise reduction of rectangular enclosures with one flexible wall," J. Acoust. Soc. Am. **35**, 1791–1797 (1963).
[5]R. W. Guy, "The response of a cavity backed panel to external airborne excitation: A general analysis," J. Acoust. Soc. Am. **65**, 719–731 (1979).
[6]L. Cheng and C. Lesueur, "Influence des amortissements sur la réponse vibroacoustique: étude théorique d'une plaque excitée acoustiquement et couplée à une cavite," J. Acoust. **2**, 105–118 (1989).
[7]L. Cheng and C. Lesueur, "Influence des amortissements sur la réponse vibroacoustique. Etude théorique et expérimentale d'une plaque excitée mécaniquement et couplée à une cavité," J. Acoust. **2**, 347–355 (1989).
[8]M. Tournour and N. Atalla, "Pseudostatic corrections for the forced vibroacoustic response of a structure-cavity system," J. Acoust. Soc. Am. **107**, 2379–2386 (2000).
[9]A. F. Seybert, C. Y. R. Cheng, and T. W. Wu, "The solution of coupled interior/exterior acoustic problems using the boundary element method," J. Acoust. Soc. Am. **88**, 1612–1618 (1990).
[10]K. R. Holland and F. J. Fahy, "The radiation of sound through an aperture in a noise control enclosure via iteration around a finite-element-boundary element loop," Noise Control Eng. J. **44**, 231–234 (1996).
[11]Y.-H. Kim and S.-M. Kim, "Solution of coupled acoustic problems: a partially opened cavity coupled with a membrane and a semi-infinite exterior field," J. Sound Vib. **254**, 231–244 (2002).
[12]F. Polonio, T. Loyau, J.-M. Parot, and G. Gogu, "Acoustic radiation of an open structure: Modeling and experiments," Acta Acust. (Beijing) **90**, 496–511 (2004).
[13]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).
[14]J.-F. Allard, *Propagation of Sound in Porous Media* (Elsevier Applied Science, England, 1993), p. 300.
[15]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).
[16]P. Göransson, "A weighted residual formulation of the acoustic wave propagation through a flexible porous material and a comparison with a limp material model," J. Sound Vib. **182**, 479–494 (1995).
[17]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[18]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[19]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).
[20]N. Atalla, R. Panneton, and P. Debergue, "A mixed displacement-pressure formulation for poroelastic materials," J. Acoust. Soc. Am. **104**, 1444–1452 (1998).
[21]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).
[22]P. Debergue, R. Panneton, and N. Atalla, "Boundary conditions for the weak formulation of the mixed (u,p) poroelasticity problem," J. Acoust. Soc. Am. **106**, 2383–2390 (1999).
[23]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).
[24]O. Dazel, F. Sgard, C.-H. Lamarque, and N. Atalla, "An extension of complex modes for the resolution of finite-element poroelastic problems," J. Sound Vib. **253**, 421–445 (2002).
[25]O. Dazel, F. Sgard, and C.-H. Lamarque, "Application of generalized complex modes to the calculation of the forced response of three-dimensional poroelastic materials," J. Sound Vib. **268**, 555–580 (2003).
[26]X. Olny, J. Tran-Van, and R. Panneton, Experimental determination of the acoustical parameters of rigid and limp materials using direct measurements and analytical solutions, Forum Acusticum, Sevilla, Spain, 2002.
[27]F. C. Sgard, N. Atalla, and J. Nicolas, "A numerical model for the low frequency diffuse field sound transmission loss of double-wall sound barriers with elastic porous linings," J. Acoust. Soc. Am. **108**, 2865–2872 (2000).
[28]O. Dazel, Synthèse modale pour les matériaux poreux, Ph.D. thesis, Ecole Nationale des Travaux Publics de l'Etat, Lyon, France, 2003, p. 271.
[29]O. Dazel, F. Sgard, F.-X. Bécot, and N. Atalla, On the expressions of dissipated powers and stored energies in poroelastic media (unpublished).
[30]H. J. P. Morand and R. Ohayon, *Fluid Structure Interaction* (Wiley, New York, 1995), p. 220.
[31]N. Atalla, F. Sgard, X. Olny, and R. Panneton, "Acoustic absorption of macro-perforated porous materials," J. Sound Vib. **243**, 659–678 (2001).
[32]N. Atalla, C. K. Amédin, and F. Sgard, *Numerical and Experimental Investigation of the Vibroacoustics of a Plate Backed Cavity Coated with an Heterogeneous Porous Material* (S.A.E. World Congress, Detroit, MI, 2003).
[33]J. S. Bolton and E. R. Green, "Normal incidence sound transmission through double-panel systems lined with relatively stiff, partially reticulated polyurethane foam," Appl. Acoust. **39**, 23–51 (1993).

Spatial diversity in passive time reversal communications

H. C. Song,[a] W. S. Hodgkiss, W. A. Kuperman, W. J. Higley, K. Raghukumar, and T. Akal
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

M. Stevenson
NATO Undersea Research Centre, La Spezia, Italy

(Received 12 January 2006; revised 19 July 2006; accepted 26 July 2006)

A time reversal mirror exploits spatial diversity to achieve spatial and temporal focusing, a useful property for communications in an environment with significant multipath. Taking advantage of spatial diversity involves using a number of receivers distributed in space. This paper presents the impact of spatial diversity in passive time reversal communications between a probe source (PS) and a vertical receive array using at-sea experimental data, while the PS is either fixed or moving at about 4 knots. The performance of two different approaches is compared in terms of output signal-to-noise ratio versus the number of receiver elements: (1) time reversal alone and (2) time reversal combined with adaptive channel equalization. The time-varying channel response due to source motion requires an adaptive channel equalizer such that approach (2) outperforms approach (1) by up to 13 dB as compared to 5 dB for a fixed source case. Experimental results around 3 kHz with a 1 kHz bandwidth illustrate that as few as two or three receivers (i.e., 2 or 4 m array aperture) can provide reasonable performance at ranges of 4.2 and 10 km in 118 m deep water. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2338286]

PACS number(s): 43.60.Dh, 43.60.Gk, 43.60.Fg [DRD] Pages: 2067–2076

I. INTRODUCTION

Recently time reversal has attracted attention in underwater acoustic[1–3] and wireless channel[4–6] communications. Time reversal typically involves a source/receive array referred to as a time reversal mirror (TRM) which samples the incoming field generated by a probe source (PS). When the received signals are played back in a time-reversed fashion, they converge to the PS location without a priori knowledge of the channel.[7,8] In a multipath environment, the time reversal process also undoes the multipath and recovers the original PS signal at the focus. The spatial and temporal focusing (pulse compression) capability of time reversal immediately offers potential application to communications, especially in an environment with significant multipath. The temporal compression mitigates the inter-symbol interference (ISI) resulting from multipath propagation, while the spatial focusing achieves a high signal-to-noise ratio (SNR) at the intended receiver with a low probability of interception elsewhere. The benefit of the time reversal approach is a simple receiver structure (complexity) as opposed to the high computational complexity required for multi-channel adaptive equalizers.[9]

The preliminary system concept for active time reversal communications has been demonstrated experimentally in shallow water using a 29 element, 78 m aperture TRM, operating in the 3–4 kHz band.[1] While the temporal focusing achieved by time reversal reduces ISI significantly, there always is some residual ISI which results in a saturation of the performance at high SNR.[10] In addition, the channel varies over time in a fluctuating environment. Recently it was confirmed using at-sea experimental data[11,12] that the performance of time reversal alone can be improved significantly in conjunction with adaptive channel equalization which simultaneously removes the residual ISI and compensates for the channel variations. Indeed, Song *et al.*[12] have shown that the combination provides nearly optimal performance using the theoretical performance bounds derived in Ref. 10. Furthermore, time reversal communications have been extended to multiple-input/multiple-output multi-user communications exploiting the spatial focusing property and linearity of the system such that independent messages were sent simultaneously from a TRM (base station) to multiple receivers (users) at 8.6 km range in 105 m deep water.[13]

Passive time reversal also has been studied in the literature where the TRM needs only to receive. Reciprocity is invoked to relate passive and active time reversal and the two approaches basically are equivalent while the communication link is in the opposite direction. Indeed, the two approaches provide the same performance in theory under ideal circumstances.[10] Dowling[14] suggested the method would be useful for pulse compression and acoustic communications. Instead of rebroadcasting the wave fronts observed at a TRM, passive time reversal is implemented numerically at the receiver using the measured channel response which requires a channel probe transmission followed by the information-bearing signal[3] (see Fig. 2). Silva *et al.*[2] proposed a virtual TRM implemented electronically at the receiver array and evaluated the performance by numerical simulations. Experimental demonstration of passive time reversal was reported by Rouseff *et al.*[3] using a 14-element array, operating in the 5–20 kHz band, up to 5 km range in water between 10 and 120 m deep. The probe source was either fixed or drifting at less than 1 knot. To account for the

[a]Author to whom correspondence should be addressed; electronic mail: hcsong@mpl.ucsd.edu

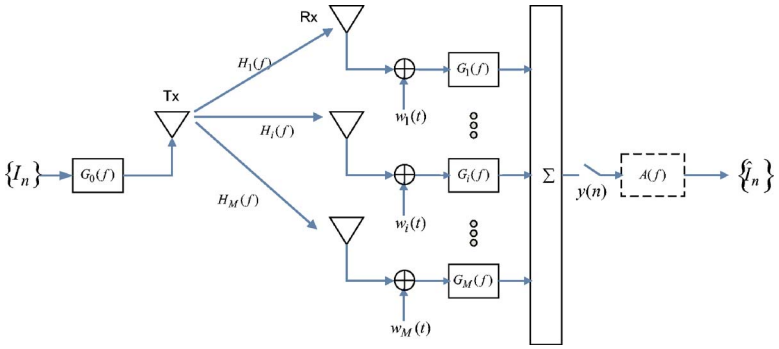


FIG. 1. (Color online) System model for passive time reversal communications followed by an equalizer (dashed box).

channel temporal variability, however, the probe signal had to be reinserted frequently to capture updated channel responses, effectively reducing the data rate by a factor of 2. To diminish the loss in data rate, a decision-directed passive phase conjugation (DDPPC) approach has been developed by the same group,[15] in which the current block of data is used to update an estimate of the channel for the next block.

As described above, both active and passive time reversal communications employ a vertical receiver array in a waveguide spanning the water column with many elements to sufficiently sample the incoming field for maximal time reversal focusing. A natural question that follows is how many elements (or how large an array aperture) are required to provide reasonable performance for practical applications of the time reversal approach. The objective of this paper is to study the impact of spatial diversity in passive time reversal communications between a single PS and a vertical receiver array, while the results are equally applicable to the active time reversal case. We take advantage of the passive time reversal approach which allows selecting a subset of array elements from a single transmission for comparison purposes. The PS is either fixed or moving.

To achieve the objective, we investigate the performance of two different approaches using at-sea experimental data in terms of output SNR as a function of the number of receiver elements: (1) time reversal alone and (2) time reversal combined with adaptive channel equalization. In the latter, channel equalization is applied to a single time series which is combined from multi-channel data using the time reversal concept (see Fig. 1), whereas multi-channel equalization typically involves (feedforward) filters applied to each channel that are updated jointly and then followed by channel combining.[9,16,17] Since time reversal is analogous to broadband matched-field processing (or generalized beam forming),[7,18,19] approach (2) can be viewed as a generalized beam former followed by channel equalization. The benefit of this approach is that the number of taps required for the post-time reversal equalizer is much smaller than the case with just an equalizer alone, thereby resulting in lower computational complexity of the equalizer.[12,20]

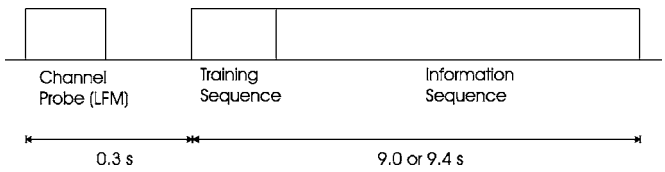


FIG. 2. Typical data format transmitted by a probe source for passive time reversal communications. The overhead resulting from a probe signal transmission and 100 training symbols is less than 5%.

This paper will present experimental results of coherent passive time reversal communications between a probe source and a 32 element vertical receiver array with 2 m spacing in 118 m deep water, operating around 3 kHz with a 1 kHz bandwidth during the focused acoustic fields 2004 (FAF-04) experiment. The theory behind passive time reversal communication is briefly reviewed in Sec. II. Section III describes the experimental setup and Sec. IV analyzes the performance for a fixed source at 10 km range. The performance analysis for a moving source at 4.2 km range is presented in Sec. V.

II. PASSIVE TIME REVERSAL: THEORY

The theory behind the use of active time reversal in the context of acoustic communications has been described in our earlier paper[1] where two-way time reversal can be seen as implementing actively a spatio-temporal matched filter of the channel response (Green's function) from a signal processing point of view. On the other hand, one-way passive time reversal requires only a receiver array and the spatio-temporal matched filter is implemented numerically at the receiver. The overall system under consideration is shown in Fig. 1 where passive time reversal is followed by an equalizer when necessary. An extensive discussion on passive time reversal communications also can be found in Ref. 3.

When a known signal $g_0(t)=s(t)$ is transmitted from a PS in a waveguide, the (noiseless) received signal on the ith element of a receiver array is $r_i(t)=s(t)*h_i(t)$ where $h_i(t)$ is the channel impulse response of the waveguide and $*$ denotes convolution. While active time reversal retransmits the time reversed version of the received signal $r_i(-t)$[12], passive time reversal applies matched filtering at each receiver element with $g_i(t)=h_i(-t)$ and combines them coherently such that

$$y(t)=\sum_{i=1}^{M} r_i(t)*g_i(t)=s(t)*\left[\sum_{i=1}^{M} h_i(t)*h_i(-t)\right]$$
$$=s(t)*q(t), \tag{1}$$

where M is the number of receiver elements and the term in the right bracket is denoted the q function representing the summation of the autocorrelation of each channel impulse response.[21] It should be mentioned that $y(t)$ is essentially

identical to the signal received at the probe source position in active time reversal $s_{ps}(-t)$ [Eq. (1) in Ref. 12]. Note that the matched filter in the frequency domain $G_i(f)=H_i^*(f)$ requires knowledge of the channel which can be obtained by a channel probe signal prior to the information-bearing signals as shown in Fig. 2. The performance of time reversal focusing depends on the complexity of the channel $h_i(t)$ (i.e., the number of multipaths), the number of array elements M, and their spatial distribution.

The application of time reversal to communications, either active or passive, relies totally on the behavior of the q function in Eq. (1). To minimize the ISI, it would be desirable to have a q function that approaches a delta function. In practice, however, there always is some residual ISI which results in a saturation of the performance.[10,22] Moreover, the channel continues to evolve over time in a dynamic ocean environment while time reversal assumes that the channel is time invariant. Thus time reversal alone may require frequent transmission of a channel probe signal at the expense of data rate to account for channel fluctuations[3] while the loss in data rate can be somewhat diminished by the DDPPC approach[15] as mentioned in Sec. I. In this paper, the passive time reversal approach will be combined with adaptive channel equalization to simultaneously eliminate the residual ISI and compensate for channel fluctuations without compromising the data rate. This is especially true for a moving source as will be shown in Sec. V.

III. EXPERIMENT

A time reversal experiment was conducted jointly with the NATO Undersea Research Center in July 2004 both north and south of Elba Island, off the west coast of Italy.[12] The passive time reversal communications experiment reported in this paper was carried out in a flat region of 118 m deep water north of Elba on July 17 (JD199). A 32-element vertical receiver array (VRA) was deployed spanning the water column from 42 to 104 m with 2 m element spacing which corresponds to about 4λ at 3 kHz. Sound speed profiles collected during the experiment are shown in Fig. 3 featuring an extended thermocline down to 70 m depth. The VRA was moored for stable operation. The probe source was either fixed at 90 m depth (+) or moving at 70 m depth (○) as marked in Fig. 3. We begin with analysis of the stationary source data in the next section.

IV. PERFORMANCE: STATIONARY SOURCE

In this section, the performance of passive time reversal communications is addressed between a stationary source and the VRA separated by 10 km as shown in Fig. 4(a). For a fixed probe source, a single element at 90 m depth was selected from a source/receiver array which we have used previously for active time reversal communications.[12] The source level was 179 dB *re* 1 μPa. The probe signal $g_0(t)$ was a 150 ms, 2.5–4.5 kHz linear frequency modulation (LFM) chirp with a Hanning window, resulting in an effective 100 ms, 3–4 kHz bandwidth chirp. Note that this probe signal is identical to the one used for active time reversal communications reported in Refs. 12 and 13 and the duration

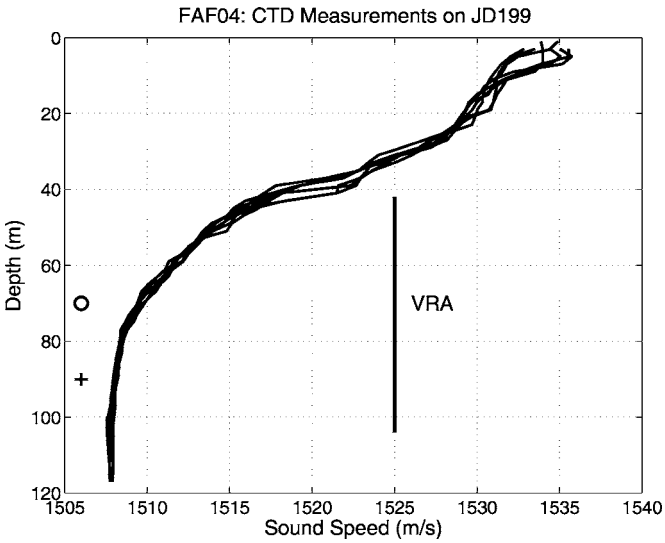


FIG. 3. The sound speed profile measured during the passive time reversal communications experiment on July 17, 2004 (JD199) along with the depth coverage of the VRA. The probe source depths are also denoted: fixed (+) and moving (○).

of the chirp after compression (matched filtering) is $T=2$ ms. Binary phase-shift keying (BPSK) was used to encode the data stream with a symbol rate of $R=1/T=500$ symbols/s such that the 150 ms chirp signal $g_0(t)$ was overlapped every 2 ms with polarity ±1 to generate the information-bearing signal $v(t)=\Sigma_n I_n g_0(t-nT)$. $\{I_n\}$ is the sequence of information symbols and $g_0(t)$ is applied as a shaping (modulation) filter (see Fig. 1). The communications sequence was 9.4 sec long with $N=4700$ symbols as shown in Fig. 2 and the sampling frequency was $f_s=12$ kHz. The resulting spectral efficiency is 0.5 bit/s/Hz for a 1 kHz bandwidth. It should be pointed out that the overhead resulting from the probe transmission and 100 training symbols reduces the data rate by less than 5%.

The channel response captured by the VRA from the PS at 10 km range is displayed in Fig. 4(b) showing a complicated multipath structure in an acoustic waveguide. The received signal is noisy even after compression of the chirp wave form with SNRs of 1.5–7.8 dB depending on the receiver depth. The delay spread is about 40 ms resulting in an ISI of 20 symbols. The measured channel response (before compression) will be used as a demodulation filter $g_i(t)$ in Fig. 1 such that $g_i(t)=g_0(-t)*h_i(-t)$ for matched filtering of both the probe signal and the channel impulse response at once. Before discussing spatial diversity with multiple receiver elements, the performance of a single receiver is investigated.

A. Single receiver: $M=1$

The performance of single element processing is illustrated in Fig. 5(a) in terms of bit error rate (BER) and output signal-to-noise ratio (SNR_o) at the receiver depth using two different approaches: (1) time reversal alone (Δ) and (2) time reversal combined with adaptive channel equalization (○). The output SNR_o (normalized by the signal energy)[23] is the reciprocal of the mean-square error (MSE) denoted by J,

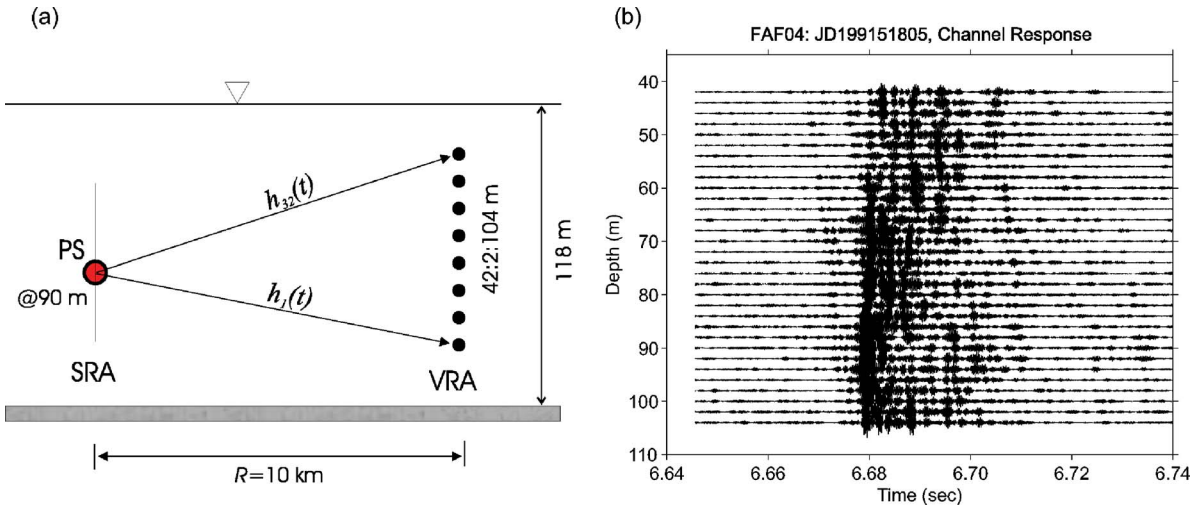


FIG. 4. (Color online) (a) Schematic of passive time reversal communication with a stationary probe source. (b) The channel responses (after compression) received by the VRA from a probe source at 90 m depth and 10 km range.

$$J = \mathrm{SNR}_o^{-1} = E|I_n - \hat{I}_n|^2, \tag{2}$$

where E denotes expectation, I_n is the information symbol (± 1) and $\hat{I}_n$ is the estimate of that symbol. Examples of scatter plots with $\{\hat{I}_n\}$ are shown in Fig. 5(b) for two different receiver depths: 46 and 78 m. For convenience, the scatter plots displayed throughout this paper are normalized with respect to the maximum value of $\{|\hat{I}_n|\}$. To estimate the input SNR ($*$), noise power is calculated outside the communication signal interval (before and/or after) while signal-plus-noise power is calculated within the communication signal interval. The difference between the two calculations is an estimate of the signal power. The input SNR ($*$) is superimposed in Fig. 5(a). The receiver for approach (1) is identical to the optimum receiver for signals corrupted by added white Gaussian noise in the absence of ISI[24] while the receiver for approach (2) is described in Ref. 12. Phase tracking was carried out by using a decision-feedback phase-locked loop (DFPLL)[24] averaged over 20 symbols.

Two observations can be made. First, the combination ($\bigcirc$) always outperforms time reversal alone (Δ) although the improvement is minimal at low input SNR ($*$). It is interesting that even a single receiver at 78 m depth provides reasonable performance (BER=0.4%, SNR_o=6.5 dB) for an input SNR of 7 dB [see Fig. 5(b)]. The BER is under 20% for all receivers where the input SNR is relatively low at levels of 1.5–7.8 dB. In general, a higher input SNR yields better performance.

An adaptive equalizer, either linear feedforward or nonlinear decision feedback (DFE) whichever yields better performance, has been applied independently to each receiver. Interestingly, it is found that when the BER of time reversal alone exceeds about 10%, a linear equalizer usually outperforms a nonlinear DFE which uses previously detected symbols to suppress the ISI in the present symbol. Since the

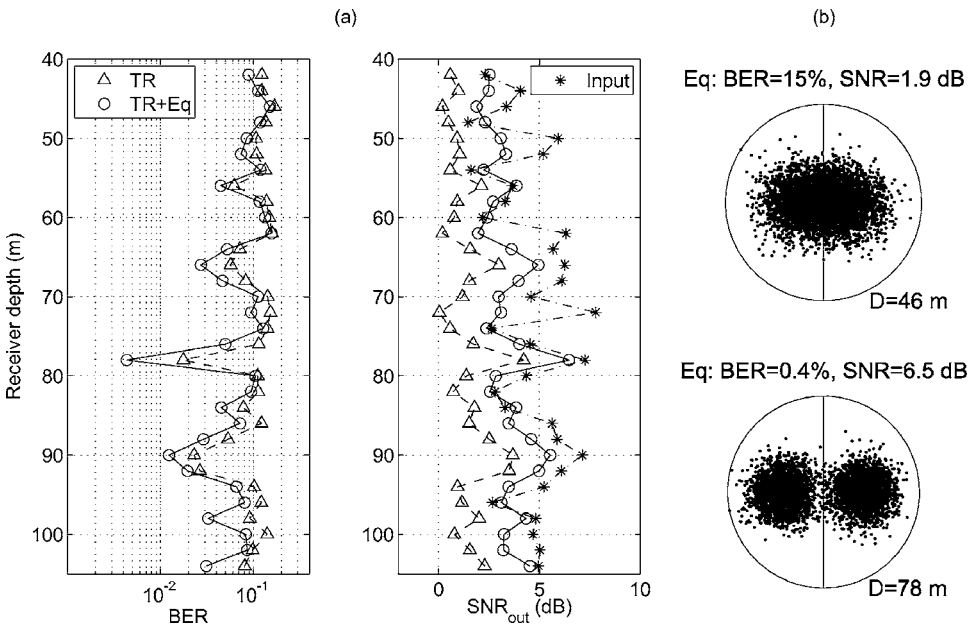


FIG. 5. Performance of single element processing: time reversal alone (Δ) and time reversal combined with an adaptive channel equalizer ($\bigcirc$). (a) BER and output SNR_o as a function of receiver depth. The input SNR ($*$) is also displayed on the right column. (b) Example of scatter plots of time reversal combined with channel equalization for two different receiver depths: D=46 m (SNR_o=1.9 dB) and D=78 m (SNR_o=6.5 dB).

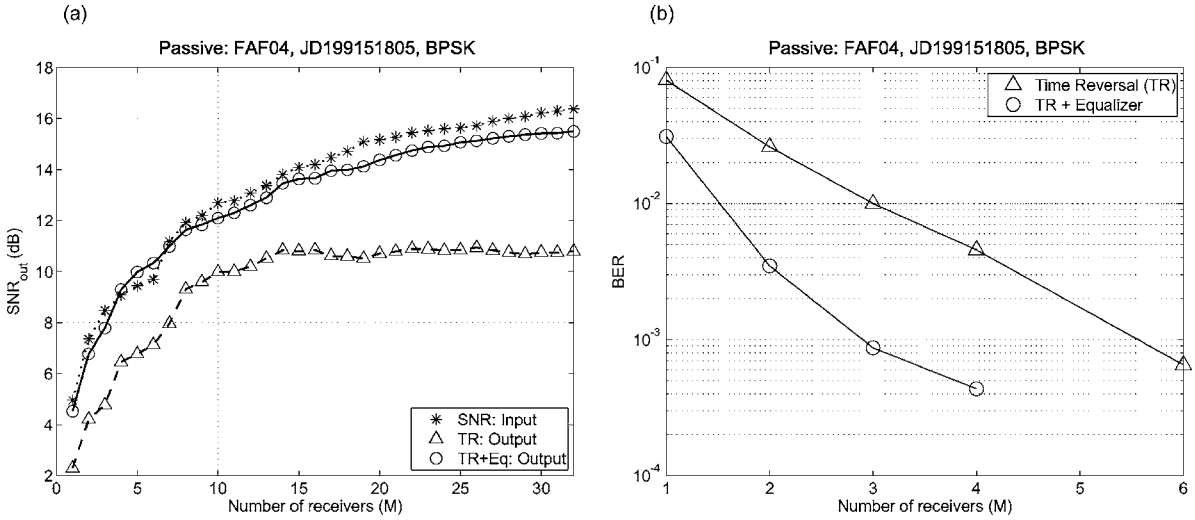


FIG. 6. Performance of multi-channel processing as a function of the number of receiver elements M: time reversal alone (Δ) and time reversal combined with an adaptive DFE equalizer ($\bigcirc$). The receiver elements are selected from the bottom (see Fig. 4). (a) Output SNR_o along with input SNR ($*$). (b)BER. There are no errors beyond the M marked.

symbol rate $R=500$ symbols/s is only half the signal bandwidth of 1 kHz, a fractionally spaced equalizer (FSE) with feedforward tap spacing of $(1/2)T$ or less should be used to avoid compensating for the aliased received signal. Here we use a tap spacing of $(1/4)T$ as in active time reversal communications resulting in the best performance.[12] The number of taps for feedforward filters n_f varies from 8 to 40. Note that even with $n_f=40$ spans only half the 20 symbols of ISI are shown in Fig. 4(b) due to the temporal compression provided by the time reversal process, which is the benefit of the combined approach ($\bigcirc$) as described in Sec. I. The recursive least squares (RLS) algorithm has been used for implementing the adaptive equalizer with forgetting factor 0.995.

It also is interesting to observe that the performance at shallower depths (around 40–60 m) appears worse in general than at deeper depths, which can be explained as follows. First, generally the energy at shallower depths is smaller than at deeper depth due to the sound speed profile shown in Fig. 3 where the energy tends to refract downward. Second, the receivers at shallower depths are located in the middle of thermocline, where they are more susceptible to environmental fluctuations.

B. Multiple receivers

The impact of spatial diversity is illustrated in Fig. 6 where the performance of time reversal communications is shown as a function of the number of receivers M in terms of (a) output SNR_o and (b) BER for two different approaches as before: time reversal alone (Δ) and time reversal combined with an adaptive DFE ($\bigcirc$). The elements are selected sequentially from the bottom such that, for example, $M=4$ includes the bottom-most four elements. Note that the horizontal axis can be replaced by aperture of the corresponding subarray ranging from 0 to 62 m. Here a nonlinear adaptive DFE has been used to generate the results of Fig. 6 which yields better performance. The number of taps used for the feedforward and feedback portions of the DFE are $n_f=20$ and $n_b=4$, respectively, and the RLS forgetting factor is 0.995. In Fig. 6(b), there are no errors beyond the M marked. Recall from Fig. 1 that the equalization is applied to a single time series which is combined from multichannel data. Phase tracking is also carried out on the single time series using a DFPLL prior to the channel equalization.

Note that the performance of time reversal (Δ) improves quite rapidly (almost linear) up to about $M=10$ (i.e., 20 m aperture), then slowly increases and eventually saturates. This happens when there is no additional gain from spatial diversity given the channel complexity and the side lobe levels of q function remain unchanged. On the other hand, the performance of time reversal combined with adaptive channel equalization ($\bigcirc$) continues to improve as the number M increases although the enhancement is gradual as compared to the significant improvement with the first several elements. This observation indicates that spatial diversity can be exploited maximally using a few elements with a small aperture given the environmental complexity. Overall, the combination outperforms time reversal alone from 2 dB up to 5 dB. In addition, the output SNR_o ($\bigcirc$) is almost comparable to the input SNR ($*$).

Taking advantage of spatial diversity requires an appropriate element spacing to ensure that the channel impulse responses are sufficiently different from one another.[22] Thus the side lobes of the autocorrelation function of each channel response interfere destructively, while the main lobes of the autocorrelation functions add up coherently in the q function. The VRA element spacing is 2 m which corresponds to $\sim 4\lambda$ at 3.5 kHz and apparently provides sufficient element spacing across the array in Fig. 4(b). One way to confirm that there is sufficient spacing between the elements is to compare the performance of different sets of receiver elements for a fixed M (e.g., $M=4$) and we did not find any noticeable differences.

A few examples of scatter plots corresponding to Fig. 6 are displayed in Fig. 7 when $M=1$, 2, 4, and 32. The top panels illustrate that the $q(t)$ function approaches delta function behavior with increasing M, resulting in improved per-

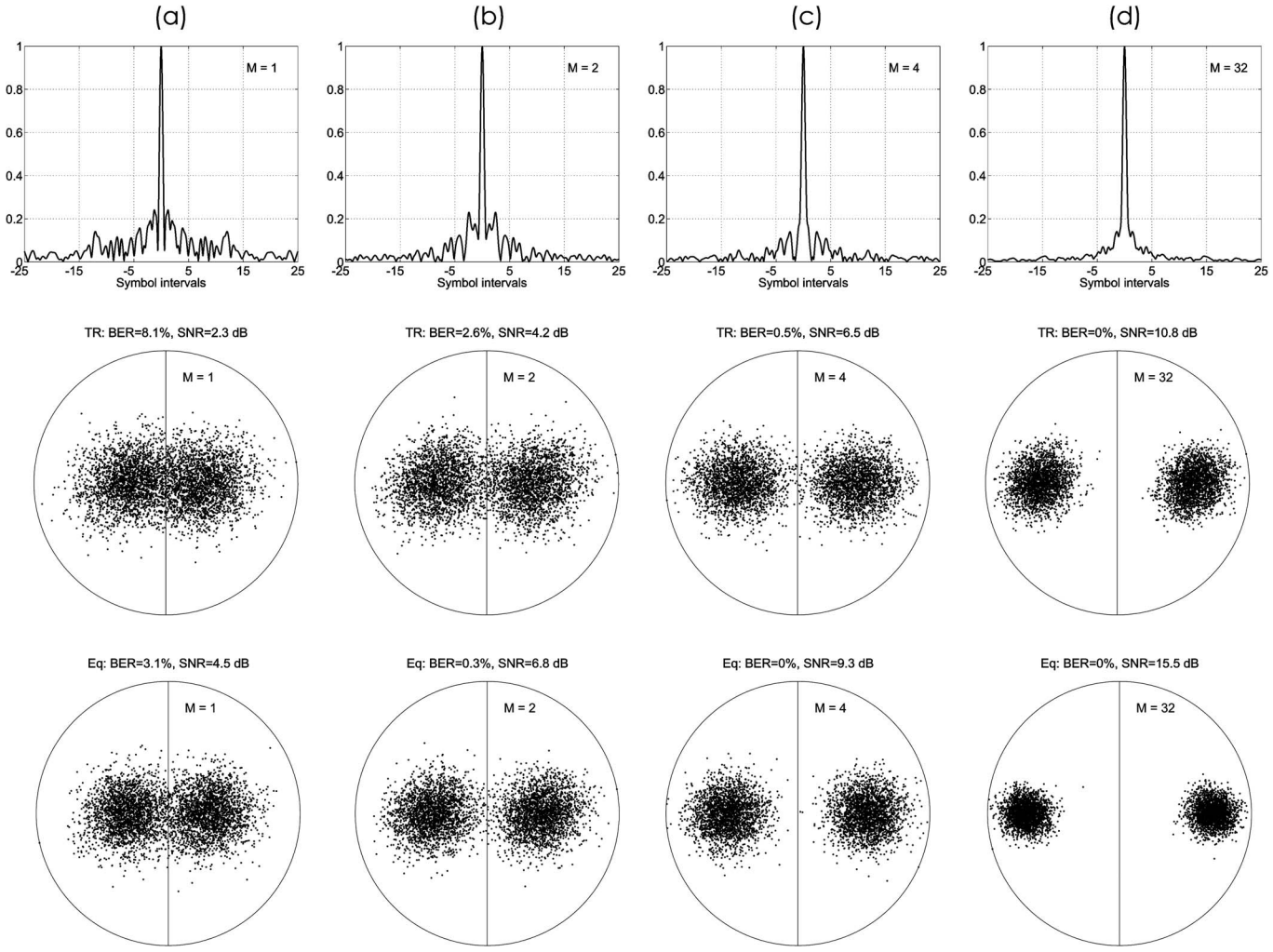


FIG. 7. Performance of multi-channel processing for various numbers of receivers M: (a) $M=1$ (single element), (b) $M=2$, (c) $M=4$, and (d) $M=32$ (entire array). The normalized $q(t)$ functions are displayed on the top row. The scatter plots are shown in the middle row for time reversal alone (Δ) and the bottom row for time reversal combined with an adaptive DFE ($\bigcirc$).

formance for time reversal communications. The middle and bottom panels are the scatter plots for time reversal alone (Δ) and time reversal combined with an adaptive DFE (o), respectively. The scatter plots suggest that as few as two receivers (or 2 m array aperture) can provide reasonable performance in this example.

V. PERFORMANCE: MOVING SOURCE

In this section, passive time reversal communications are investigated between a moving source and the VRA separated by 4.2 km as shown in Fig. 8(a). The probe source was an ITC-2007 (formerly ITC-1000) towed at about 4 knots at 70 m depth away from the VRA. The source level was 200 dB *re* 1 μPa. The channel probe signal was a 100 ms, 2–4 kHz LFM chirp and the corresponding channel impulse responses (envelope) are shown in Fig. 8(b) indicating a delay spread of about 100 ms. Each data symbol $g_0(t)$ was a 1 ms, 3 kHz continuous wave tone, as opposed to the chirp signal used for the fixed source in Sec. IV. The symbol rate was $R=1000$ symbols/s using BPSK modulation. The communication sequence was 9 s-long with $N=9002$ symbols and the spectral efficiency is 1 bit/s/Hz. As in the fixed source case, the overhead resulting from a probe transmission and 100 training symbols was less than 5%. In the presence of source motion, a rough estimate of the Doppler shift is required in order to re-sample the original data.

A. Doppler estimation and resampling

Figure 9 shows the block diagram for Doppler compensation processing. First, a coarse estimate of the Doppler shift arising from source motion is obtained from the initial phase tracking applied to the original data at the carrier frequency $f_c=3$ kHz. The result for the bottom element is displayed in Fig. 10(a) where the slope corresponds to the Doppler shift of $\hat{f}_d=-4.710$ Hz. Note that no separate signaling scheme (e.g., a pilot tone[25]) is required to estimate the Doppler shift.

The next step is to re-sample the original data using a new sampling frequency of $\tilde{f}_s=f_s+\hat{f}_d$ to remove the Doppler shift. An efficient polyphase interpolator[26] and linear interpolation[27] is used for sampling rate conversion since the Doppler shift is less than 0.2%. Phase tracking applied to resample data is shown in Fig. 10(b) indicating that there is a

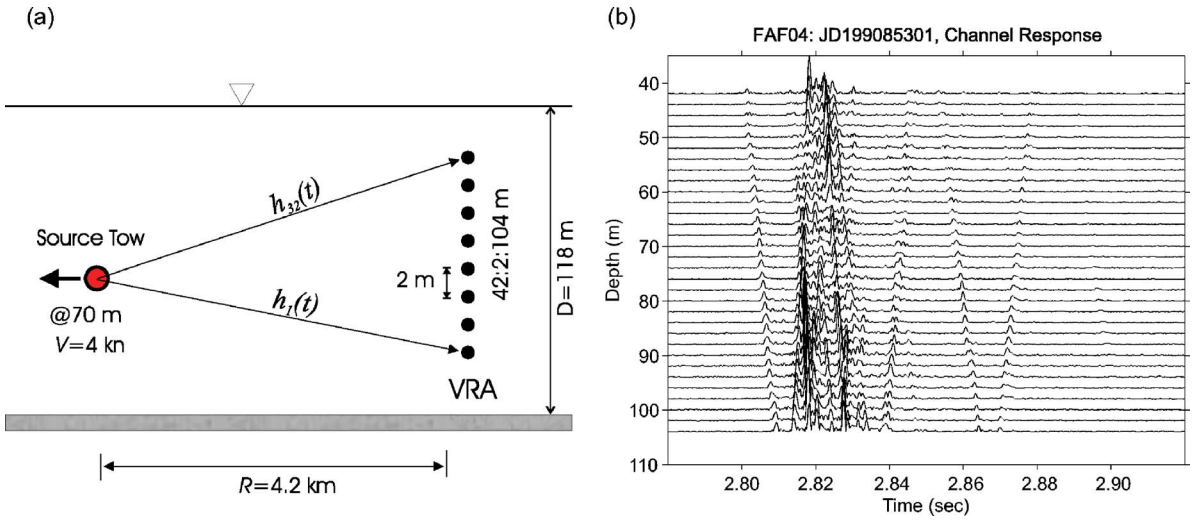


FIG. 8. (Color online) (a) Schematic of passive time reversal communications with a moving source at about 4 knots. (b) The channel responses (envelope) received by the vertical receiver array from a probe source at 70 m depth and 4.2 km range.

residual Doppler shift of $\hat{f}_d$=0.064 Hz. The overall Doppler shift is then f_d=−4.710+0.064=−4.646 Hz. There also is a small Doppler shift due to mismatch in sampling rate (f_m =−0.045 Hz) which was observed in our previous experiment.[13] The source velocity can be estimated assuming a sound speed of c=1508 m/s from Fig. 3:

$$\hat{v} = c\left[\frac{f_d - f_m}{f_c}\right] \approx -2.3 \text{ m/s}. \tag{3}$$

The ship speed during this experiment was 4.2 knots (2.1 m/s) according to the navigation data. For a single receiver case (M=1), Doppler compensation was applied independently to each element. The mean Doppler shift was −4.702 Hz with a standard deviation of 0.051 Hz. For the multi-element case ($M>1$), we applied the same Doppler shift estimated for the single bottom element to resample all receiver elements.

B. Single receiver: M=1

The performance of single element processing is illustrated in Fig. 11 in terms of BER and SNR_o using two approaches: time reversal alone (Δ) and the combination of time reversal and adaptive channel equalization (○). Note that time reversal alone (Δ) shows very poor performance such that SNR_o=−1.6–2.7 dB and BER=6%–30%, which is not surprising. Time reversal involves a correlation process (matched filtering) using the measured channel probe signal, assuming that the channel does not vary during the transmission of a data packet. However, the channel responses continue to evolve over time due to source motion even when the environmental fluctuations are minimal. Consequently, the matched filtering introduces an undesirable mismatch which gradually changes over time (9 s) even after resampling. This is the limitation of time reversal approach unless it is followed by an adaptive channel equalizer to compensate for the channel variations due to source motion. Therefore we focus mainly on the combined approach (○) in this subsection.

It is interesting that the performance in the middle of the water column (60–90 m) is worse with BER of 10%–30% while some receivers above and below the region show reasonable performance with BER less than 1%. Examples of scatter plots are displayed in Fig. 11(b) for two different depths: 50 m (best) and 84 m (worst). There are no bit errors at 50 and 94 m (no marker shown). Note also the symmetry between the BER (○) and SNR_o (○) which is the reciprocal of the MSE [See Eq. (2)]. In general, the output SNR_o (○) shows a similar trend as the input SNR (∗) which varies around 26–36 dB depending on the receiver depth while the input SNR displayed is offset by 20 dB for convenience. The large difference between the input and output SNRs is not well understood, but is due in part to the differential Doppler between different ray paths which leads to overall Doppler spread. In fact, a large discrepancy (10–15 dB) in the presence of strong multipath is also reported in Ref. 27 for a moving source.

Since the symbol rate is approximately equal to the sig-

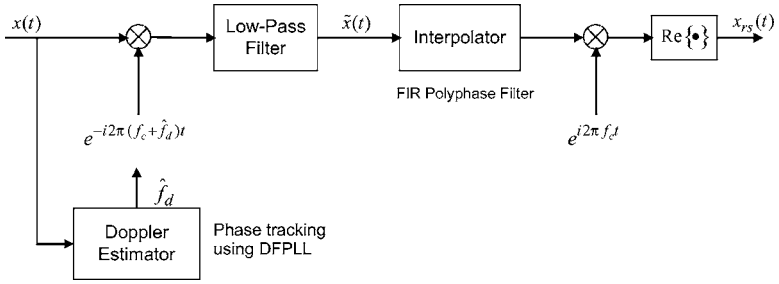


FIG. 9. Block diagram for Doppler compensation processing. A Doppler shift $\hat{f}_d$ is estimated from initial phase tracking results using a DFPLL. Linear interpolation is carried out on the complex base band signal $\tilde{x}(t)$ after carrier phase correction and $x_{\text{rs}}(t)$ is the output signal re-sampled using the new sampling frequency of $\hat{f}_s = f_s + \hat{f}_d$.

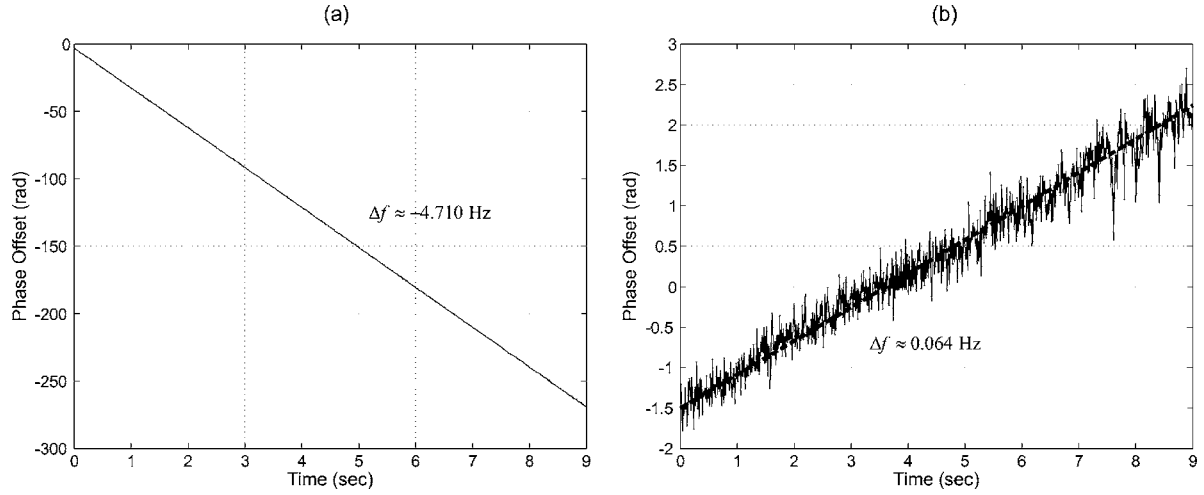


FIG. 10. Phase tracking results using a DFPLL applied to: (a) original data and (b) re-sampled data. A coarse Doppler estimate of $\hat{f}_d = -4.710$ Hz is obtained from (a). Plot (b) indicates that there is a residual Doppler shift of about $\hat{f}_d = 0.064$ Hz.

nal bandwidth[24] of 1 kHz (i.e., $W \approx 1/T$), we can use a FSE with feedforward tap spacing of $(1/2)T$ with RLS forgetting factor of 0.99. A linear equalizer has been applied to array elements in the middle of the water column and two additional depths (48 and 104 m) where the BER is greater than 10%, as mentioned in Sec. IV A, and the number of taps is $n_f = 20$. The remaining depths employ an adaptive DFE with $n_f = 20$ and $n_b = 8$.

C. Multiple receivers

The impact of spatial diversity in the presence of source motion is illustrated in Fig. 12. The performance of time reversal communications is shown in terms of (a) output SNR_o and (b) BER as a function of the number of receivers M for two different approaches: time reversal alone (Δ) and time reversal combined with an adaptive DFE (○). There are no errors beyond $M = 3$ for the combination (○) in Fig. 12(b). As before, the elements are selected sequentially from the bottom. Except when $M = 1$, where a linear equalizer with $n_f = 16$ has been employed, an adaptive DFE is used with $n_f = 20$ and $n_b = 8$ and the RLS forgetting factor is 0.99.

Three observations can be made. First, the performance characteristics are very similar to those shown earlier in Fig. 6 such that the initial significant improvement is followed by a slow increase then either a minimal gradual improvement (○) or saturation (Δ). On the other hand, the combination (○) outperforms time reversal alone (Δ) as much as 13 dB as compared to 5 dB for the stationary source case. The saturation effect is visible in both BER and SNR_o. This observation clearly indicates that the channel variation due to source motion requires channel adaptivity and the time reversal approach should be used in conjunction with an adaptive channel equalizer to compensate for the channel variations. Finally, the input SNR (∗) shows similar behavior to the output SNR (○) while there is about 25 dB discrepancy as in Fig. 11. Note that the input SNR (∗) displayed is offset by 20 dB.

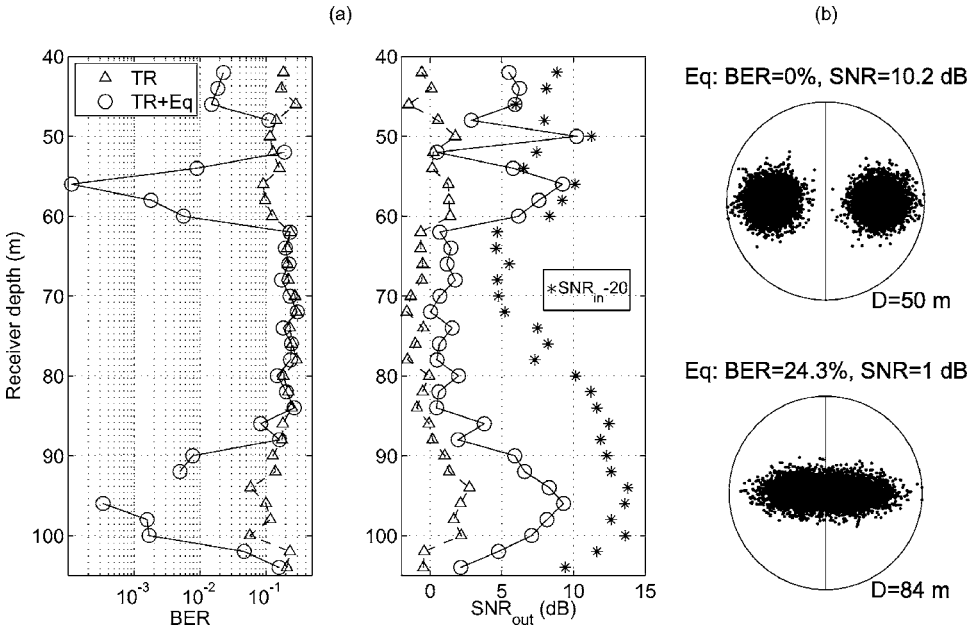


FIG. 11. Performance of single element processing: time reversal alone (Δ) and time reversal combined with an adaptive channel equalizer (○). (a) BER and SNR_o as a function of receiver depth. The input SNR (∗) is also displayed on the right column with an offset of 20 dB. (b) Example of scatter plots of time reversal combined with channel equalization for two different receiver depths: 50 and 84 m.

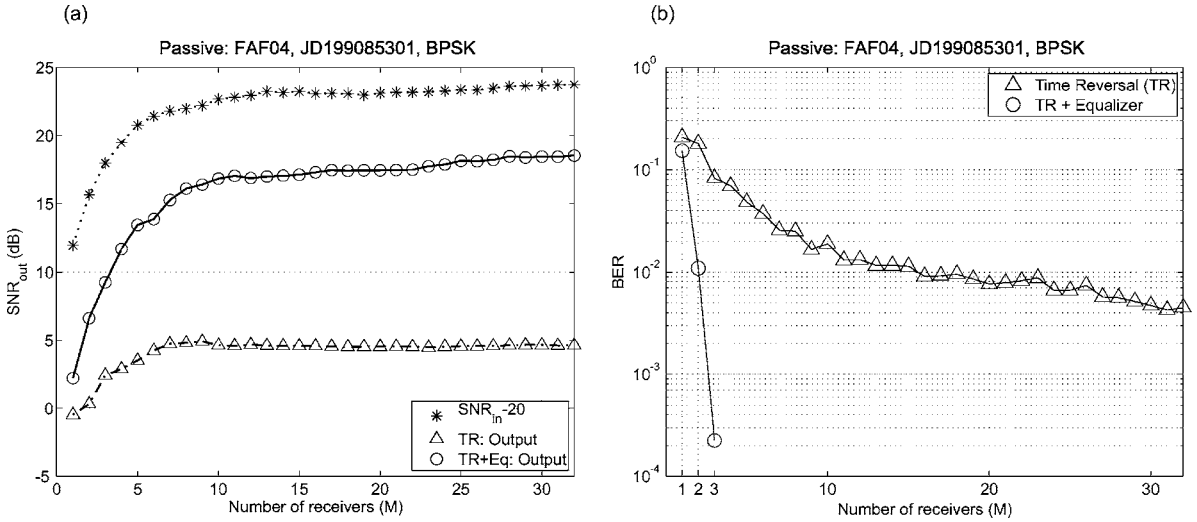


FIG. 12. Performance of multi-channel processing as a function of the number of receiver elements M: time reversal alone (Δ) and time reversal combined with an adaptive DFE (○). (a) Output SNR_o. The input SNR (*) is also displayed with an offset of 20 dB. (b) BER. The combination (○) shows no errors beyond $M=3$.

A few examples of scatter plots corresponding to Fig. 12 are displayed in Fig. 13 for $M=1$, 2, 3, and 10. The top panels show that the $q(t)$ function approaches a delta function with increasing M. The middle panels confirm again that

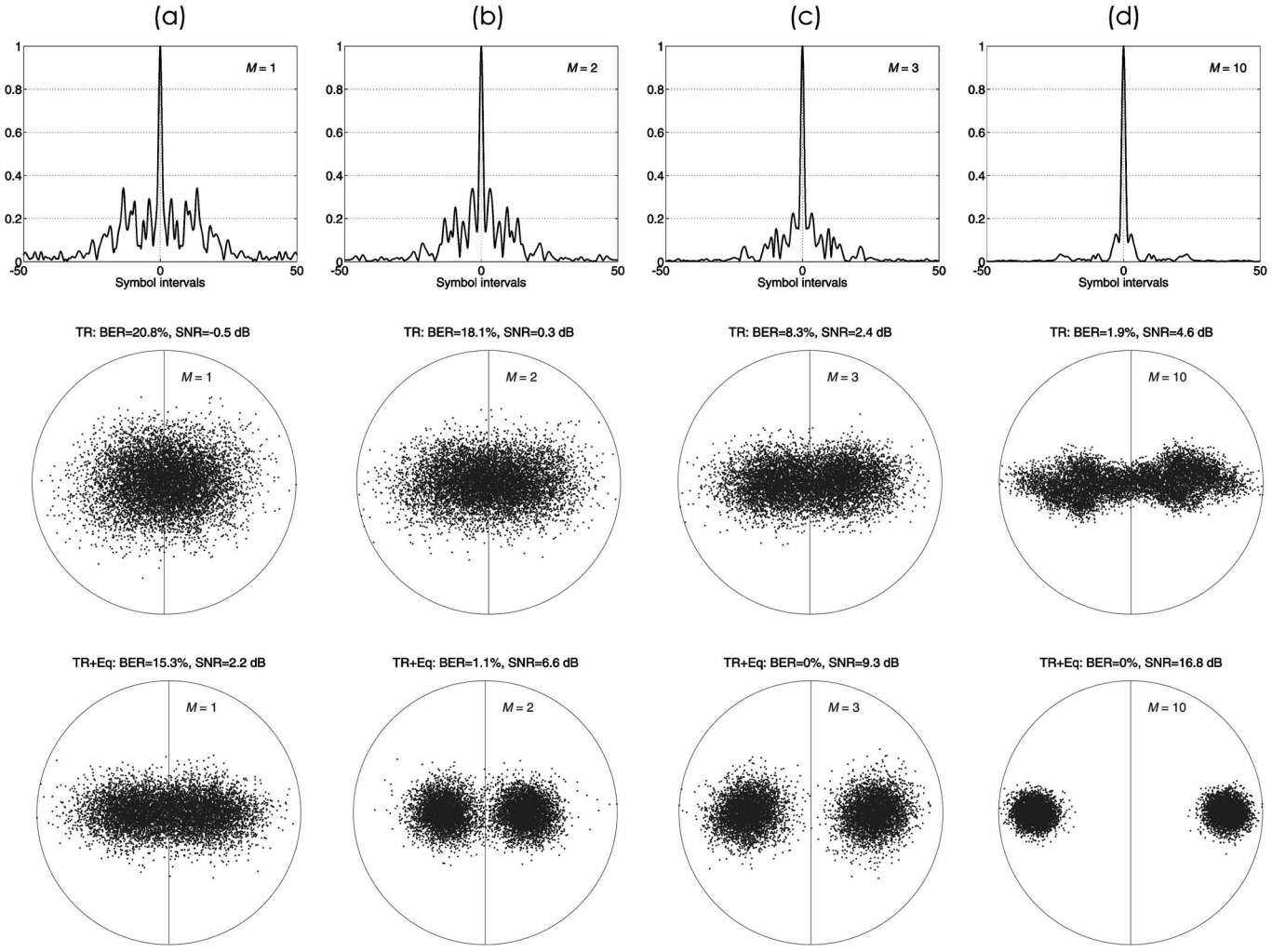


FIG. 13. Performance of multi-channel processing for various numbers of receivers M: (a) $M=1$ (single element), (b) $M=2$, (c) $M=3$, and (d) $M=10$. The normalized $q(t)$ functions are displayed on the top row. The scatter plots are shown in the middle row for time reversal alone (Δ) and the bottom row for time reversal combined with an adaptive DFE (○).

the performance of time reversal alone (Δ) cannot be improved with increasing M unless the channel variations are accounted for by an adaptive channel equalizer as shown in the bottom (o). The scatter plots suggest that as few two receiver elements (or 2 m aperture) can provide reasonable performance in this example.

VI. CONCLUSION

Time reversal mirrors, either active or passive, exploit spatial diversity to achieve spatial and temporal focusing, a useful property for communications in an environment with significant multipath. Taking advantage of spatial diversity involves using a number of receivers distributed in space. While the analysis is equally applicable to the active case, for practical purposes we investigated the impact of spatial diversity in passive time reversal communications between a single probe source and a vertical receive array using the data from our July 2004 experiment. The probe source was either fixed (90 m depth and 10 km range) or moving at about 4 knots (70 m depth and 4.2 km range) from the 32-element vertical receiving array spanning the water column from 42 to 104 m with 2 m spacing in 118 m deep water, operating in the 2–4 kHz band. The performance of two different approaches was compared in terms of output signal-to-noise ratio and bit error rate as a function of the number of receivers: (1) time reversal alone and (2) time reversal combined with adaptive channel equalization.

Two conclusions have been made. First, approach (1) saturates when there is no additional gain from spatial diversity given the channel complexity. Second, approach (2) always outperforms approach (1) because the adaptive equalizer simultaneously eliminates the residual ISI and compensates for channel fluctuations if any exist. This is especially true for a moving source which introduces time-varying channel responses such that the performance enhancement amounts up to 13 dB as compared to 5 dB for a fixed source case. Finally, experimental results around 3 kHz with a 1 kHz bandwidth illustrate that as few as two or three receivers (i.e., 2 or 4 m array aperture) can provide reasonable performance at ranges of 10 km (fixed source) and 4.2 km (moving source) in 118 m deep shallow water.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research under Grant Nos. N00014-05-1-0263 and N00014-06-1-0128.

[1]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).
[2]A. Silva, S. Jesus, J. Gomes, and V. Barroso, "Underwater acoustic communications using a virtual electronic time-reversal mirror approach," *Proc. Fifth European Conference on Underwater Acoustics*, 531–536 (2000).
[3]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).
[4]G. Lerosey, J. de Rosney, A. Tourin, A. Derode, G. Montaldo, and M. Fink, "Time reversal of electromagnetic waves," Phys. Rev. Lett. **92**, 193904 (2004).
[5]H. Nguyen, J. Andersen, and G. Pederson, "The potential use of time reversal techniques in multiple element antenna systems," IEEE Commun. Lett. **9**, 40–42 (2005).
[6]B. E. Henty and D. D. Stancil, "Multipath-enabled super-resolution for rf and microwave communication using phase-conjugate arrays," Phys. Rev. Lett. **93**, 243904 (2004).
[7]W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. Jackson, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**, 25–40 (1998).
[8]C. Feuillade and C. Clay, "Source imaging and sidelobe suppression using time-domain techniques in a shallow-water waveguide," J. Acoust. Soc. Am. **92**, 2165–2172 (1992).
[9]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).
[10]M. Stojanovic, "Retrofocusing techniques for high rate acoustic communications," J. Acoust. Soc. Am. **117**, 1173–1185 (2005).
[11]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).
[12]H. Song, W. Hodgkiss, W. Kuperman, M. Stevenson, and T. Akal, "Improvement of time reversal communications using adaptive channel equalizers," IEEE J. Ocean. Eng. [in press (2006)].
[13]H. Song, P. Roux, W. Hodgkiss, W. Kuperman, T. Akal, and M. Stevenson, "Multiple-input/multiple-output coherent time reversal communications in a shallow water acoustic channel," IEEE J. Ocean. Eng. **31**, 170–178 (2006).
[14]D. R. Dowling, "Acoustic pulse compression using passive phase-conjugate processing," J. Acoust. Soc. Am. **95**, 1450–1458 (1994).
[15]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).
[16]M. Stojanovic, J. A. Capitovic, and J. G. Proakis, "Reduced-complexity spatial and temporal processing of underwater acoustic communication signals," J. Acoust. Soc. Am. **98**, 961–972 (1995).
[17]Q. Wen and J. Ritcey, "Spatial diversity equalization applied to underwater communications," IEEE J. Ocean. Eng. **19**, 227–241 (1994).
[18]A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**, 401–424 (1993).
[19]R. K. Brienzo and W. S. Hodgkiss, "Broadband matched-field processing," J. Acoust. Soc. Am. **94**, 2821–2831 (1993).
[20]T. C. Yang, "Correlation-based decision-feedback equalizer for underwater acoustic communications," IEEE J. Ocean. Eng. **30**, 865–880 (2005).
[21]T. Yang, "Temporal resolutions of time-reversed and passive-phase conjugation for underwater acoustic communications," IEEE J. Ocean. Eng. **28**, 229–245 (2003).
[22]H. C. Song and A. Dotan, "Comments on retrofocusing techniques for high rate acoustic communications [J. Acoust. Soc. Am. **117**, 1173–1185 (2005)]," submitted to J. Acoust. Soc. Am. (2006).
[23]M. Stojanovic, J. A. Capitovic, and J. G. Proakis, "Phase-coherent digital communications for underwater acoustic channels," IEEE J. Ocean. Eng. **19**, 110–111 (1994).
[24]J. Proakis, *Digital Communications* (McGraw–Hill, New York, 2001).
[25]J. Dhanoa, R. Ormondroyd, and E. Hughes, "An improved digital communication system for doubly-spread underwater acoustic channel using evolutionary algorithms," in *Proc. Oceans 2003*, 109–114 (2003).
[26]M. Johnson, L. Freitag, and M. Stojanovic, "Improved doppler tracking and correction for underwater acoustic communications," in *Proc. ICASSP'97*, 575–578 (1997).
[27]B. Sharif, J. Neashan, O. Hinton, and A. Adams, "A computationally efficient doppler compensation system for underwater acoustic communication," IEEE J. Ocean. Eng. **25**, 52–61 (2000).

Tracking fin whale calls offshore the Galicia Margin, North East Atlantic Ocean

Oriol Gaspà Rebull,[a] Jordi Díaz Cusí, Mario Ruiz Fernández, and Josep Gallart Muset
Dep. Geofisica i Tectònica, Institut de Ciències de la Terra "Jaume Almera" (CS1C), c/ Solé i Sabaris s/n. 08028 Barcelona, Spain

(Received 22 March 2006; revised 13 July 2006; accepted 20 July 2006)

Data recorded during a temporary deployment of ocean bottom seismometers (OBSs) are used in this study to monitor the presence of fin whales around the array. In the summer of 2003, ten OBSs were placed 250 km from the NW coast of Iberia in the Galicia Margin, NE Atlantic Ocean for a period of one month. The recorded data set provided a large variety of signals, including fin whale vocalizations identified by their specific acoustic signature. The use of a dense array of seafloor receivers allowed investigation into the locations and tracks of the signal-generating whales using a seismological hypocentral location code. Individual pulses of different sequences have been chosen to study such tracks. Problems related to the correct identification of pulses, discrimination between direct and multiple arrivals, and the presence of more than one individual have been considered prior to location. Fin calls were concentrated in the last two weeks of the deployment and the locations were spread around the area covered by the array. These results illustrate that, besides its classical seismological aim, deployment of semipermanent seafloor seismic arrays can also provide valuable data for marine mammal behavior studies. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2336751]

PACS number(s): 43.60.Jn, 43.30.Sf, 43.80.Ka [WWA] Pages: 2077–2085

I. INTRODUCTION

Ocean bottom seismometers (OBSs) were primarily designed to record seismic-acoustic signals at the seafloor that are generated by man-made devices at the sea surface (active sources) with short-term exploration purposes. Typically, these OBSs included a hydrophone that records pressure fluctuations on the sea water and were equipped with a three component seismometer of frequency sensitivity ranging from 1–4 Hz to near 50 Hz. Nowadays, new instruments with enlarged bandwidth at lower frequencies are becoming operational for longer periods of time to also undertake seismological studies (passive sources). The frequency sensitivity of the instruments has allowed the recording of signals with a wide range of origins, from earthquake activity (Sato *et al.*, 2004), volcanic tremors (Talandier and Okal, 1987), T-phases (Walker and Bernard, 1993), monochromatic infrasound waves probably related to gas or fluid venting (Pontoise and Hello, 2002), seafloor hydrothermal flows (Sohn *et al.*, 1995), or different marine mammal vocalizations (McDonald *et al.*, 1995; Clark *et al.*, 2002). The availability of this wide variety of signals has increased in the last few years in parallel with technical improvements making it possible to deploy OBSs during several weeks or months continuously recording of the incoming signal.

Between 23 August and 21 September 2003, ten OBSs, provided by IRD-Géosciences Azur (France), were deployed approximately 250 km offshore the NW coast of Iberia [Fig. 1(a)] as part of a multidisciplinary study of the zone where the oil tanker "Prestige" sank on 13 November 2002. The OBSs were equipped with three-component 4 Hz geophones and in most cases, an additional hydrophone. Data were acquired continuously at a sampling rate of 100 Hz, although in the recording procedure every disk access entails some lost signal (about 27.3 s every 11 min). The area covered was roughly 50 km^2 centered around 42°N 12°W, with distances between adjacent instruments from 8 to 10 km [Fig. 1(b)]. After deployment of the OBS array, six multichannel vertical seismic profiles were acquired in the study area using the seismic technology available at the Spanish BIO-Hespérides vessel. The water-wave arrivals from those profiles recorded at the different OBSs were used to infer their accurate position on the sea bottom. In a second phase, starting on 25 August 2003, the OBSs operated as a passive seismic network, recording a wide variety of signals, including local, regional, and teleseismic earthquakes (Díaz *et al.*, 2006a), local activity in the form of harmonic tremors and short-duration events (Díaz *et al.*, 2006b), as well as another type of event identified hereafter as whale calls. The purpose of this contribution is to analyze the latter and introduce a new method to determine the position and pathway of the whales based on a seismic location algorithm, using the analogy between the acoustic waves generated by whale calls and seismic compressional waves.

II. IDENTIFICATION OF FIN WHALE CALLS

The data set obtained by the OBS network was thoroughly explored to identify any type of earthquake activity and this exploration revealed the presence of fin whale (*Balaenoptera physalus*) calls. These signals are characterized by well-known features (Watkins *et al.*, 1987): (i) short and regular pulses of about 0.6 s of duration, with a regular pulse interval of 12.6 s recorded by several instruments simulta-

[a]Electronic mail: ogaspa@ija.csic.es

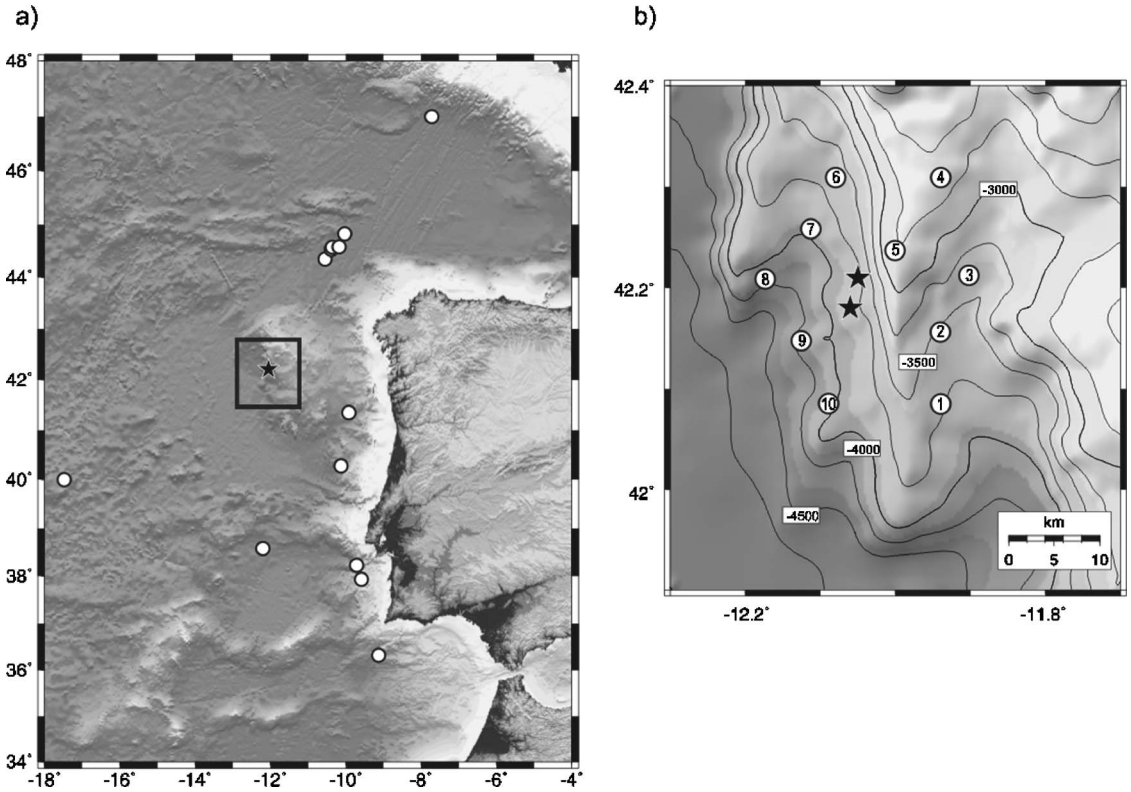


FIG. 1. (a) Location of the oil tanker Prestige wreckage (black star) (Galicia Margin, NE Atlantic), with fin whale sightings from 1969 to 2003 from OBIS-SEAMAP (white dots). (b) Location of the ten OBSs (circles), prow and stern of the Prestige (black stars).

neously; (ii) trains of pulses of variable duration but generally ranging between 3 and 12 min, separated by rest intervals, typically of 1–3 min; (iii) narrow amplitude spectra ranging from 18 to 26 Hz, with a peak close to 20 Hz. It is now well established from combined visual observations and instrumental recordings (Schevill *et al.*, 1964) that these characteristics correspond to the acoustic waves generated by the most common call type of the fin whale, often described as "doublet 20-Hz" and related to male breeding (Watkins *et al.*, 1987; Thompson *et al.*, 1992; Croll *et al.*, 2002). The call series are usually associated with the respiration cycle of whales and are usually described using the terms pulse interval, rest, and gap (Watkins *et al.*, 1987). The pulse interval is the time measured from a particular point on one pulse to the same point on succeeding pulses. Rests correspond to the time intervals between different train pulses and are associated with the breathing of the whale, while the pulse series duration is related to diving times, when the whale generates its calls. Finally, gaps are defined as long periods of time between calls [Fig. 2(a)].

Fin whales (*Balaenoptera physalus*) are the second largest animal on Earth, reaching up to 25 m and weighing up to 70 000 kg. They are found in all oceans of the world and may migrate from subtropical waters during the winter to the colder areas of the Arctic and Antarctic for feeding during the summer (Perry *et al.*, 1999). They have been observed all along the North Atlantic Ocean, from Iceland in the north to the coast of northwest Africa in the south as documented by OBIS-SEAMAP (Read *et al.*, 2003) [Fig. 1(a)].

Fin whale vocalizations recorded in our experiment were observed to concentrate in the last two weeks of recording [Fig. 2(b)]. Within this period, the presence of whales near the OBS array was observed at more than three instruments simultaneously for approximately 80 h. Rest intervals (associated with breathing) have an average length of 125 s. Most of the analyzed rest series have a fairly uniform duration around 90 s, except for Julian days 249 and 250, which show a more variable rest interval duration, reaching up to 310 s. Call series (associated with diving) have an average duration of 450 s. Calling sequences recorded during days 249, 250, and 251 have an average duration of about 350 s, but most of the bouts recorded during days 257–258 present clearly longer sequence durations close to 600 s. Some extremely short sequence durations, with less than 200 s, are also present within the whole period [Fig. 2(c)]. Although these values are similar to those reported previously by both visual and acoustic surveys (Patterson and Hamilton, 1964; Watkins *et al.*, 1987; McDonald *et al.*, 1995; Sirovic *et al.*, 2004; Nieukirk *et al.*, 2004), the wide and scattered range of call intervals observed for days 257 and 258, suggest that several individuals were possibly present within the area. This is supported by the data presented in Fig. 2(b): OBS6 and OBS10, almost 30 km apart, both show a significant increase in the daily time with whale calls recordings for days 257 and 258 (14 and 15 July).

Only call sequences with more than six simultaneous observations and a high signal-to-noise ratio were retained for their further location to avoid problems of signal identification. The frequency content of the recorded pulses has been analyzed using spectrograms. Typical frequencies lie between 18 and 26 Hz with a peak at 20 Hz (Fig. 3). Some authors (e.g., McDonald and Fox, 1999) have reported that

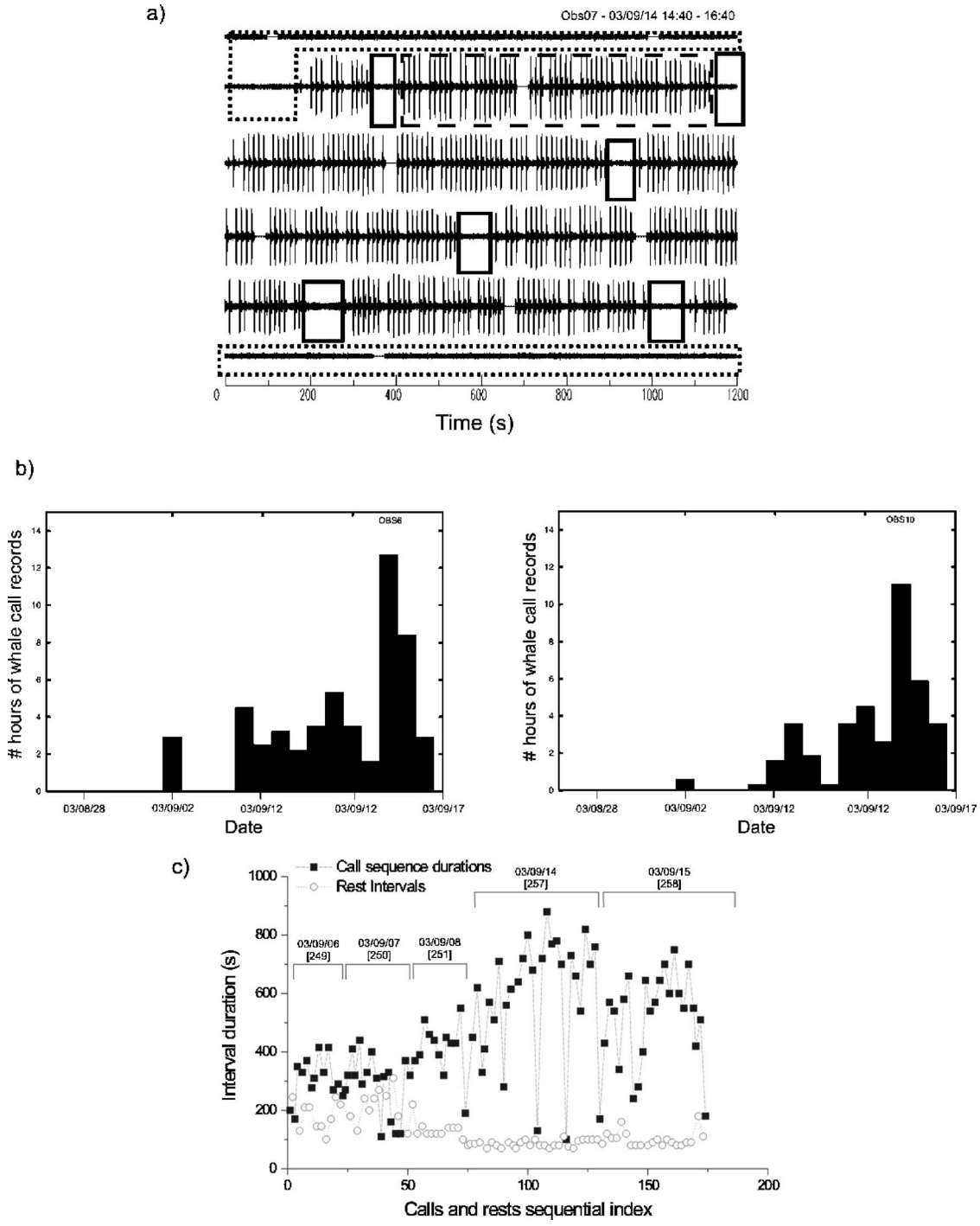


FIG. 2. (a) Two hours of recording (20 ft per row) from Obs07 vertical component, depicting trains of pulses (associated with diving times, slashed boxes), rests (associated with breathing, solid boxes), and gaps, large periods of time between calls (dotted boxes). (b) Daily number of hours of fin whale calls detected at Obs06 and Obsl0. (c) Rest and calling time interval durations detected through the registration period [brackets: Julian days].

the presence of different individuals may be inferred by a close inspection of the spectral signature. In our case, we could not observe any significant difference between pulses, even in the case where the presence of more than one animal is suggested by other features.

III. LOCATION AND TRACKING

The existence of a dense array of ocean bottom seismometers encouraged the location of the signal-generating whales and determination of their tracks by means of algorithms primarily designed for earthquake hypocentral determination, based on classic ray-path theory. We tested several codes and finally used the NonLinLoc package (Lomax *et al.*, 2000), which allows one to perform hypocentral determinations in a probabilistic, nonlinear grid-search approach. In this code, the region of interest is covered by a grid, where a two- or three-dimensional (2D or 3D) velocity model is defined which can take care of regional sound speed and bathymetric variations. Theoretical travel times through the model are calculated from each point of the grid to the receivers and

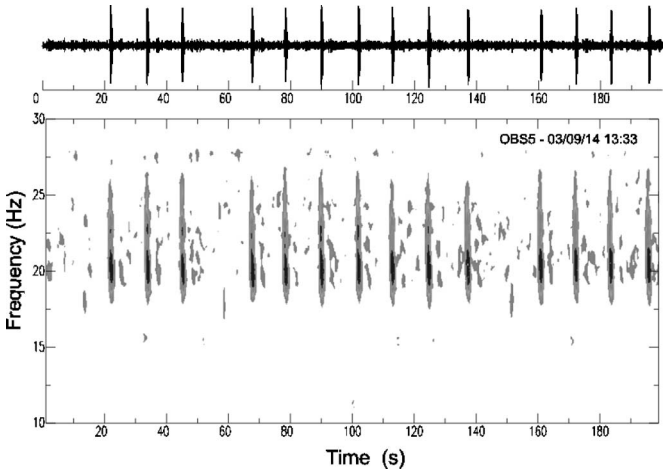


FIG. 3. Spectrogram of a calling sequence recorded at Obs05 showing the typical frequency content of the whale calls, ranging between 18 and 26 Hz with a frequency peak around 20 Hz.

compared to the observed travel times. This approach provides posterior density functions (PDFs) which assign a degree of probability to each point of the grid, allowing a reliable probabilistic evaluation of the location uncertainties. These PDFs include location uncertainties related to the relative location of the network and the source, to errors in the picking of the arrival times, and to differences between the real velocity distribution and the velocity model used for the calculation of travel times.

Classical methods for acoustic location of marine mammals fix the whale's position by comparing the measured and theoretical time differences between multiple pairs of receivers using least squares hyperbolic approach (Marchand, 1964; Clark and Ellison, 2000; Tiemann *et al.*, 2002; Janik *et al.*, 2000). Recent improvements to these methods allow the average speed of sound to vary between different receivers and the source (Spiesberger and Wahlberg, 2002) or explicitly account for ray-refraction effects arising from a depth-dependent sound speed profile (Thode, 2005). The seismological routine we considered follows a similar approach, using in this case the absolute time arrivals at each receiver and taking care of regional sound speed and bathymetric variations. There are not *a priori* restrictions to the depth and distance to the receivers of the signal generating animal.

In order to perform the location of the acoustic signal with this seismic package, some aspects ought to be taken into consideration:

(a) Pulse identification: Due to the short delay between successive pulses (~12.6 s) and the distance between the OBSs, it is not obvious to ensure that the same pulse is chosen in all the stations. A mismatch in the pulse selection may obviously lead to significant errors in the source location. To prevent those errors, the rests between trains of pulses are used as reference framework, as illustrated in Fig. 4(a).

(b) Multipath effects: Multipath arrivals are usually generated by the multiple reflections of the whale call between the sea bottom and the free surface of the ocean, although the presence of water layers with different densities can also contribute to multipathing. These secondary arrivals will be

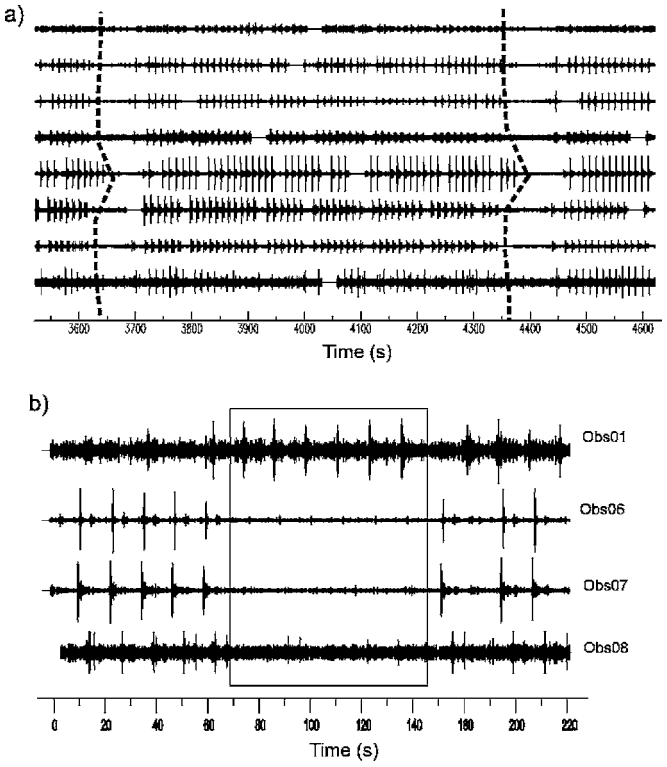


FIG. 4. (a) Pulse bout recorded at eight different OBSs used for the location procedure. Using the rest intervals, the same pulse can be clearly identified. (b) Example of a sequence where signals from at least two different individuals are observed. The pulses recorded at Obs01 are coincident with the rest interval at Obs07 and Obs08 (box). Obs06 records calls from both individuals; the most energetic arrivals show the same pattern as in Obs07 and 08, while smaller amplitude pulses are consistent with those recorded at Obs01.

detected in a distance range that depends on the bathymetry and can be significantly more energetic than the direct waves (McDonald *et al.*, 1995). As the location program only deals with direct arrivals, it must be verified that the picked times do not include secondary arrivals.

(c) Recognizing individuals: All pulses used to determine the location must be generated by the same individual. Fin whales are found most often alone, but groups of three to seven individuals are also common (Perry *et al.*, 1999). As already discussed, it has not been possible to discern between individuals through differences in the spectral signature. To prevent this source of errors, we have visually checked the consistency of the call and rest intervals between the whole OBSs array. In the cases where the rest intervals are not coincident in time between the different receivers [Fig. 4(b)] we assumed that more than one individual were present and discarded that sequence for the location procedure to avoid mislocations.

(d) The regular use of NonLinLoc includes readings of shear wave arrivals, which do not exist in this case. The strong seismic wave dependence between depth determination and shear wave arrivals makes it necessary to increase the minimum number of pickings for each call to maintain the location accuracy. We have mainly worked with cases when no less than five good quality picks of the same pulse were available.

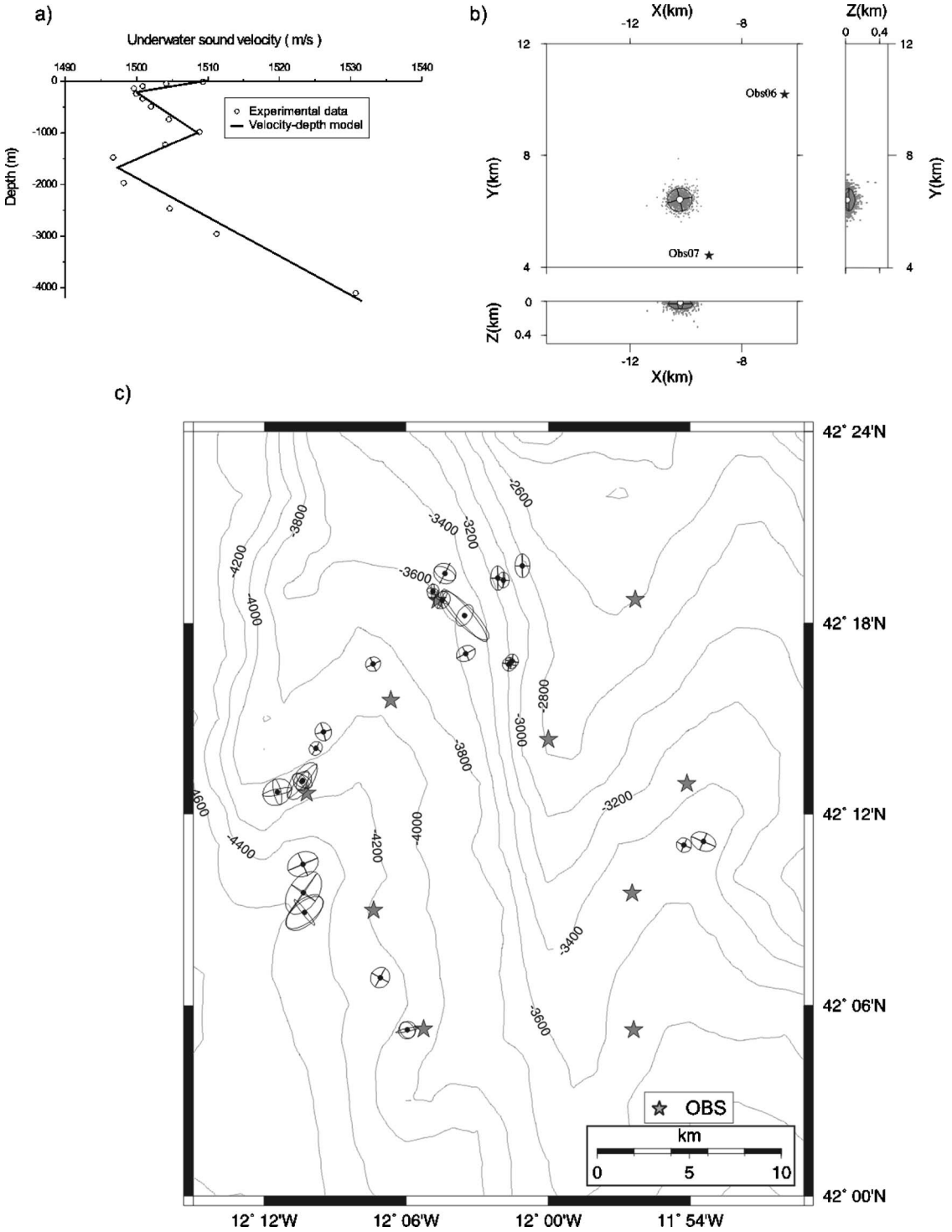


FIG. 5. (a) Velocity-depth model used for the location determination (solid line) based on experimental data (dots). (b) Example of call location. The white point denotes the maximum likelihood hypocentral solution. Dots represent the 1000 points of the grid with higher probability density for this location, while ellipses represent the projection of the 68% confidence ellipsoid on three different projections (XY, XZ, YZ). The area covered by this ellipses, illustrates the location accuracy. (c) Locations of all the 23 resolved vocalizations during the recording period. For each position, a black dot denotes the optimal epicentral solution and the ellipses represent the map projection of the axis of the 68% confidence ellipsoid.

Underwater sound velocity has temperature, salinity and pressure dependence. Using Mackenzie's nine-term equation (MacKenzie, 1981) and constant temperature and salinity, underwater sound velocity variation between 0 and 4500 m can reach 60 m/s. Velocity depth profiles close to the study zone have been calculated from vertical data sections acquired by the "BORD-EST" campaign from IFREMER, between May and June of 1988 (http://sam.ucsd.edu/talley_url/atlases/atlases.html), showing that in this zone the velocity values range between 1.495 km/s at 1500 m and 1.535 km/s at 4500 m.

To perform the location we used a velocity model [Fig. 5(a)] reproducing this velocity profile in the water layer over a layer of unconsolidated sediments. The differences in

TABLE I. Whale location results. Columns labeled "Az" and "Dip" refer to the azimuth and dip defining the position of the first axis of the ellipsoid. Len1, Len2, and Len3 refer to the length of the three orthogonal semiaxes of this ellipsoid.

Date	Origin time	Latitude (°)	Longitude (°)	Depth (m)	Az (°)	Dip (°)	Len1 (m)	Len2 (m)	Len3 (m)
2003-09-02	05:43:02.02	42.3311	−12.0183	3	343	89	65	440	692
2003-09-02	06:13:27.86	42.3234	−12.0358	3	317	89	69	418	709
2003-09-02	06:27:37.84	42.2996	−12.0532	3	63	24	619	846	2310
2003-09-02	08:10:34.42	42.1165	−12.1188	3	129	−88	165	545	618
2003-09-07	15:00:16.17	42.1860	−11.8885	5	152	87	203	524	754
2003-09-07	15:52:29.32	42.1844	−11.9047	3	133	89	70	416	443
2003-09-07	23:14:02.54	42.1619	−12.1704	5	174	−85	501	858	1400
2003-09-08	00:12:18.38	42.1747	−12.1724	5	331	−87	342	702	928
2003-09-08	00:47:56.19	42.1493	−12.1728	10	312	44	865	1010	1330
2003-09-08	03:05:19.72	42.0886	−12.1006	5	345	6	498	575	963
2003-09-14	06:42:40.49	42.2116	−12.1934	5	292	64	635	772	969
2003-09-14	09:36:20.04	42.2339	−12.1643	3	219	−89	54	363	416
2003-09-14	10:33:58.79	42.2777	−12.1255	3	94	−89	78	419	447
2003-09-14	11:39:04.27	42.2838	−12.0577	3	51	−89	73	438	562
2003-09-14	13:34:28.48	42.2804	−12.0253	3	109	−89	41	376	424
2003-09-14	14:08:19.61	42.2777	−12.0278	3	66	−89	44	390	400
2003-09-14	15:17:11.36	42.3121	−12.0752	3	337	−68	371	458	550
2003-09-14	18:15:23.09	42.3217	−12.0323	3	291	−89	27	351	452
2003-09-15	10:00:49.69	42.3278	−12.0730	5	206	−49	505	670	760
2003-09-15	17:07:14.08	42.2273	−12.1606	5	314	14	605	898	1370
2003-09-15	18:48:49.47	42.2430	−12.1599	3	303	89	97	480	529
2003-09-15	21:36:16.28	42.2232	−12.1710	10	307	12	478	582	1330
2003-09-16	01:36:55.46	42.3184	−12.0829	5	93	−5	353	428	1210

bathymetry between OBSs sometimes reached 1500 m and had to be corrected prior to the location procedure. All readings were obtained by manually picking the onset time of each pulse. To avoid errors due to the reading method, four consecutive calls were processed in each case and the location was retained only if the results were coherent. This approach led to reducing the large initial number of detected calls to the 23 locations finally retained (Table I). An example of location is presented in Fig. 5(b). The optimal hypocenter is depicted by the white circle. Dots represent the 1000 points of the grid with higher probability density, while the ellipsoid encloses all points within the 1σ (68%) confidence level. The locations obtained in this study are presented in Fig. 5(c). Solutions are well constrained in an approximate area of about 0.5 km^2. In most cases, the location procedure fixes a source origin close to the surface, consistent with the usually accepted values of 50 m for fin whales maximum calling depth (Watkins *et al.*, 1987).

After location, traces were displayed in a distance versus time plot to verify the consistency of the solution [Fig. 6(a)]. A fair agreement is observed between the first arrivals and the hodochrone for a 1.5 km/s water velocity. Other arrivals corresponding to multiple reflections are visible with significant amplitude at distances greater than 8 km from the source. Arrival times of direct and multiple phases lack parallelism in some cases due to the bathymetric variations along the profile.

Simple synthetic analyses using a classical 2D ray-tracing program (Zelt and Smith, 1992) have been performed to assess the consistency of the locations. The model accounts for the 2D velocity-depth profile used for location and for the bathymetric variations along the network [Fig. 6(a)]. Synthetic seismograms including direct and multiple waves can be calculated for the different OBSs [Fig. 6(b)], leading finally to a synthetic distance versus time plot [Fig. 6(c)] that can be compared to real data. The fit between the synthetic and real data illustrates the quality of the whale location determination. Possible confusions between direct and multipath arrivals can be checked with this synthetic data.

As a further test, we have compared our results with those provided by the methods classically used in whale localization (Clark *et al.*, 1996). For this, we have calculated the arrival-time differences between multiple pairs of receivers using classical cross-correlation methods that provide the bearing lines for each pair of receivers. The intersection of those bearing lines defines the location of the vocalizing whale. The use of OBS located at the sea-bottom in a zone with significant bathymetric variations makes it necessary to adapt the method that, in its usual form assumes a 2D model, to a 3D configuration. The offset between the locations obtained through this procedure and those from the seismological routine we used is not larger than 1–2 km, falling within our location accuracy.

Once the location of the different calls was established, we investigated whether they could be attributed to the same individual and being part of a single trajectory. We considered only those periods without gaps greater than 1 h and verified that the signature of the recorded signals varied smoothly through the time intervals between locations, without evidence of anomalous features such as indications for

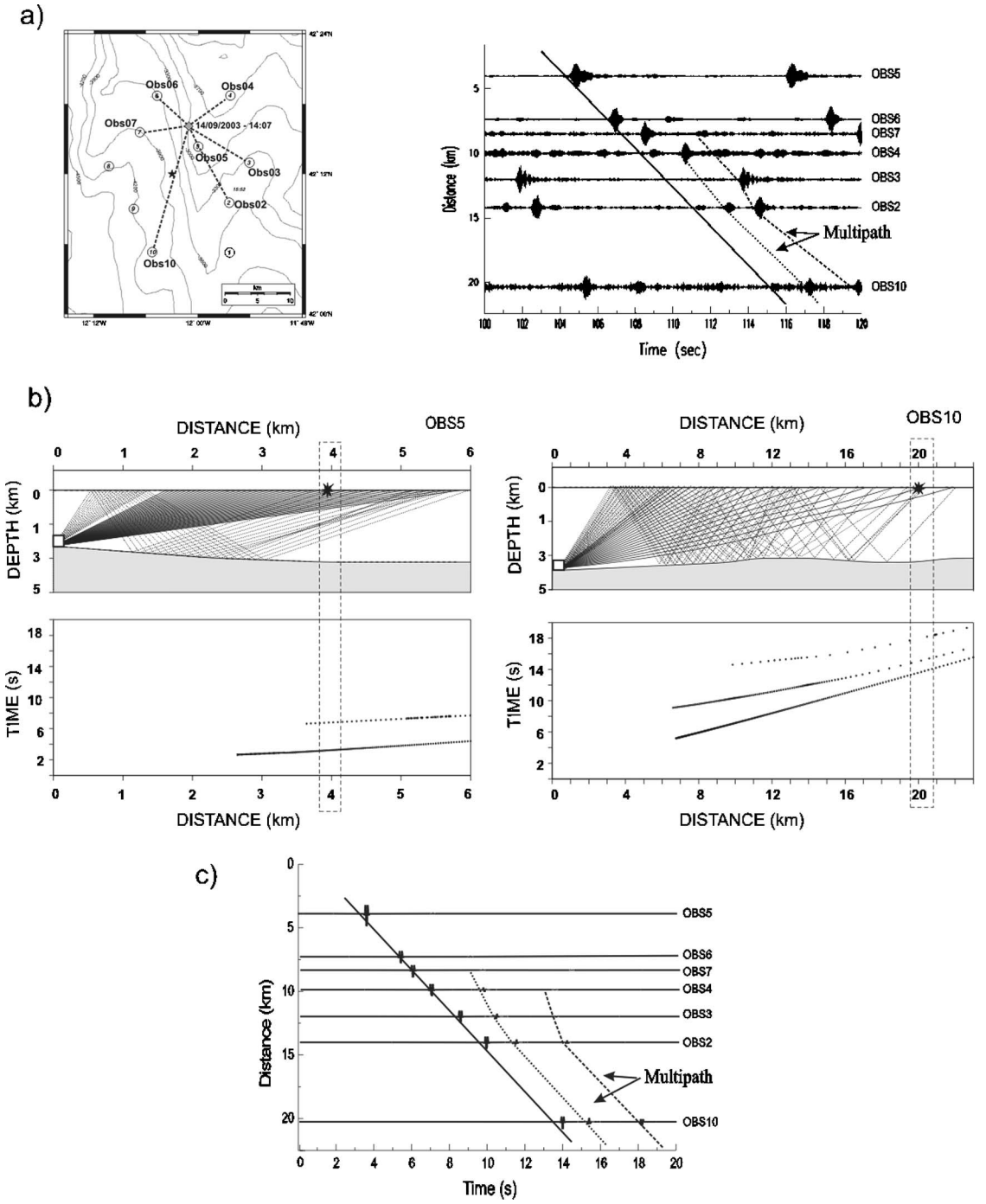


FIG. 6. (a) Location obtained for the first pulse of the sequence starting the day 257 (14/09/2003) at 14:08 and distance vs time plot for the real data analyzed. (b) Examples of raypaths for Obs05 and Obs10, including direct and multiple waves calculated using the velocity model used to locate the whale calls. (c) Distance vs time plot for the synthetics signals calculated using the ray-tracing modeling based on the pulse location. The solid lines show the direct arrival and the dashed lines its multiples. Those features are comparable with the real data graphic shown above.

the presence of different individuals. The use of the correlograms between different pairs of receivers (Fig. 7) has been useful to assess the continuity of the calls series. Following those criteria, up to four tracks have been established (Fig. 8), with durations ranging from 1 to 8 h. It can be noted that the apparent speed, as deduced from the time difference between consecutive locations, is far from regular. For some periods, the apparent speed is in the range 4–7 km/h, consistent with the range reported from visual inspections (Perry *et al.*, 1999). However there are some periods (day 250 15–16 h; day 257 between 11:40 and 14:08) in which the whale seems to remain nearly static while its vocalizations remain regular.

IV. CONCLUSIONS

The presence of fin whales in the Galicia Margin, NE Atlantic has been assessed through the sounds of their calls recorded in a temporary array of ten autonomous ocean bottom seismometers.

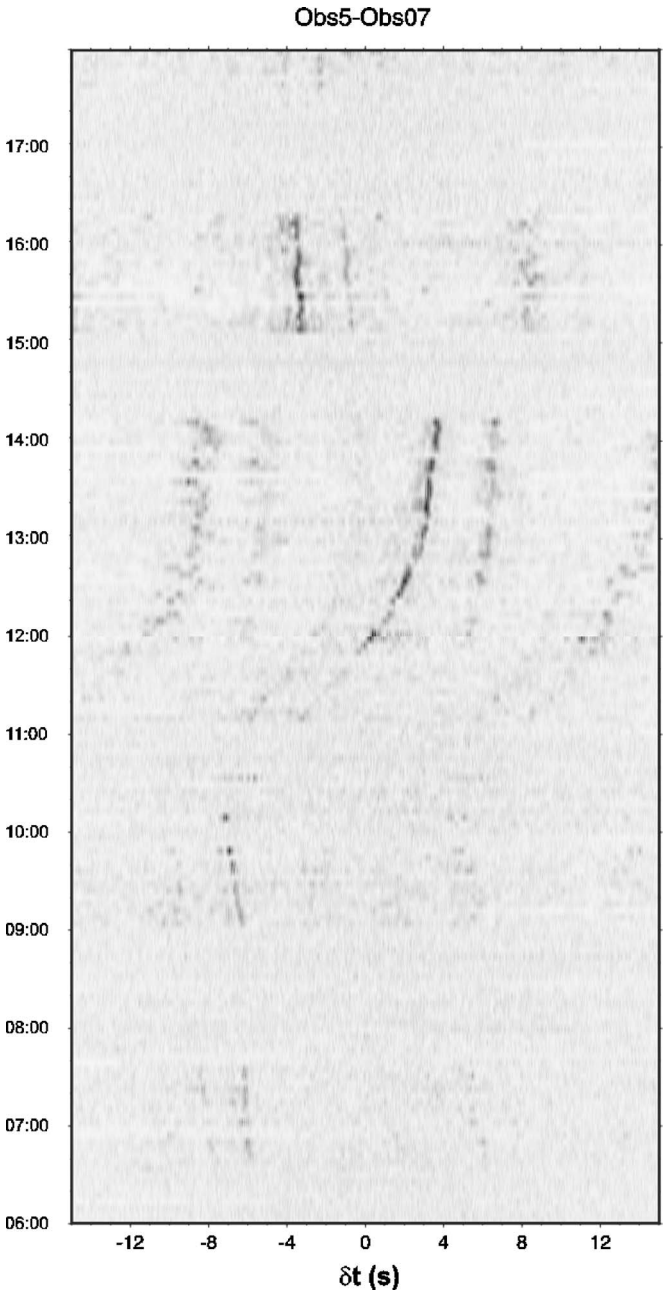


FIG. 7. Correlogram between the signals of Obs05 and Obs07 (7 km apart) recorded between 06:00 and 18:00 of day 257 (14/09/2003). Each row corresponds to the correlation of a 30 s window. Data are bandpass filtered prior to correlation. *Y* axis (labeled in hours) denotes the time of each correlation, while the *X* axis denotes the time delay in seconds between both signals. The darkest shadow represents the highest correlation coefficient and the best trajectory to use, while lighter shadows are fakes due to cross correlations between different multipath arrivals and correlations with precedent and next pulses (separated 12.6 s). The relative displacement of the whale can be established in the intervals 06:40/07:40, 09:00/10:00, 11:20/14:20. The change observed between 14:20 and 15:00 can be interpreted as the end of a single track and the beginning of another or the appearance of another call-generating whale.

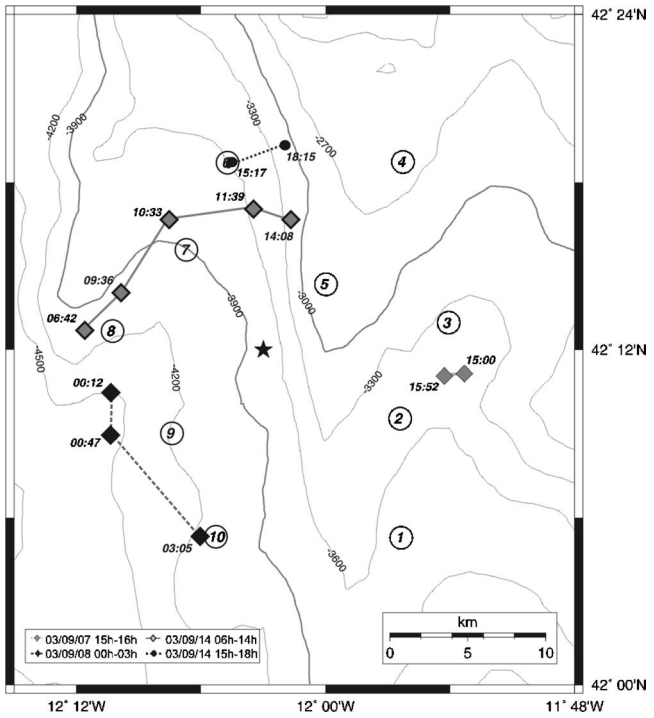


FIG. 8. Tracks of the four identified sequences.

We have used a hypocentral location routine, usual in seismological research, to fix the positions of the sound generating whales. The algorithm considered does not involve any restriction in the velocity depth distribution or in the source depth and can easily account for bathymetric variations. The probability density functions provided by the algorithm are not based on any *a priori* assumptions on the velocity model and geometrical features. Therefore the results obtained, based on the grid search routine, allow reliable probabilistic evaluation of the location uncertainties.

The location of successive calls along the same series defined whale tracks around the array area. The possible sources of error associated with the presence of more than one individual and to the misidentification between direct waves and multiple reflections have been analyzed. The quality of the locations has been tested from techniques commonly used in seismological practice, as the distance/time diagrams or synthetic models. The presence of more than one individual during some periods can be established from the detailed inspection of the recorded data and it is probably coincident with an increase of the call interval duration.

These results illustrate that deployments of semipermanent seafloor seismic arrays can be a valuable instrument not only for seismological purposes, but also to track and monitor the movements of fin whales and other marine mammals without human disturbance, allowing in this way the study of their natural behavior. Therefore, future geophysical projects that include long-term deployment of OBSs could be considered as challenges for more integrated bio-geosciences research.

ACKNOWLEDGMENTS

This research has been funded by the Spanish Special Action "Identificación de riesgos geoambientales potenciales y su valoración en la zona de hundimiento del buque Prestige." Additional support from the REN2001-1734-C03-01 program. OBSs have been provided by IRD-Géosciences Azur, Villefranche-sur-Mer, France, under a specific contract. We acknowledge support from IRD engineering staff and

participation of Yann Hello, Valenti Sallarés, and Audrey Gailler, in the marine cruises. Additional support is provided by the Dept. of Universities, Research and Society of the Generalitat de Catalunya government. Two anonymous reviewers provided valuable suggestions to improve the manuscript.

Clark, C. W., Borsani, F., and Notarbartolo-di-Sciaar, G. (**2002**). "Vocal activity of fin whales, *Balaenoptera physalus*, in the Ligurian Sea," Marine Mammal Sci. **18**, 286–295.

Clark, C. W., Charif, R. A., Mitchell, S. G., and Colby, J. (**1996**). "Distribution and behavior of the bowhead whale, *Balaena mysticetus*, based on analysis of acoustic data collected during the 1993 spring migration off Point Barrow, Alaska," Scientific Report, International Whaling Commission Vol. **46**, pp. 541–552.

Clark, C. W., and Ellison, W. T. (**2000**). "Calibration and comparison of the acoustic location method used during the spring migration of the bowhead whale (*Balaena mysticetus*), off Pt. Barrow, Alaska, 1984-1993," J. Acoust. Soc. Am. **107**(1), 3509–3517.

Croll, D. A., Clark, C. W., Acevedo, A., Tershy, B., Flores, S., Gedamke, J., and Urban, J. (**2002**). "Only male fin whales sing loud songs," Nature (London) **417**, 809.

Díaz, J., Gallart, J., and Gaspà, O. (**2006b**). "Subsurface seismic activity at the Galicia Margin, North Atlantic Ocean, evidenced from harmonic tremors and short-duration seismic events," Tectonophysics (submitted).

Díaz, J., Gallart, J., Gaspà, O., Ruiz, M., and Córdoba, D. (**2006a**). "Seismicity in the Galicia Margin, NW Iberia, in relation with the sinking of the 'Prestige' oil-tanker," Mar. Geol. (to be published).

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.

Lomax, A., Virieux, J., Volant, P., and Berge, C. (**2000**). "Probabilistic earthquake location in 3D and layered models: Introduction of a Metropolis-Gibbs method and comparison with linear locations," in *Advances in Seismic Event Location*, edited by C. H. Thurber and N. Rabinowitz (Kluwer, Amsterdam), pp. 101–134.

MacKenzie, K. V. (**1981**). "Nine-term equation for the sound speed in the oceans," J. Acoust. Soc. Am. **70**(3), 807–812.

Marchand, N. (**1964**). "Error distribution of best estimate position from multiple time difference hyperbolic networks," Aerosp. Navigational Electron. **11**, 96–100.

McDonald, M. A., and Fox, C. G. (**1999**). "Passive acoustic methods applied to fin whale population density estimation," J. Acoust. Soc. Am. **105**(5), 2643–2651.

McDonald, M. A., Hildebrand, J. A., and Webb, S. C. (**1995**). "Blue and fin whales observed on a seafloor array in the Northeast Pacific," J. Acoust. Soc. Am. **98**(2), 712–721.

Nieukirk, S. L., Stafford, K. M., Mellinger, D. K., Dziak, R. P., and Fox, C. G. (**2004**). "Low-frequency whale and seismic arigun sounds recorded in the mid-Atlantic Ocean," J. Acoust. Soc. Am. **115**(4), 1832–1843.

Patterson, B., and Hamilton, G. R. (**1964**). "Repetitive 20 cycle per second biological hydroacoustic signals at Bermuda," in *Marine Bio-Acoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 125–144.

Perry, S. L., DeMaster, D. P., and Silber, G. K. (**1999**). "The fin whale," Marine Fisheries Review **61**(1), pp. 44–51.

Pontoise, B., and Hello, Y. (**2002**). "Monochromatic infra-sound waves recorded offshore Ecuador: Possible evidence of methane release," Terra Nova **14**(6), 425–435.

Read, A. J., Halpin, P. N., Crowder, L. B., Hyrenbach, K. D., Best, B. D., and Freeman, S. A. (editors). (**2003**). "OBIS-SEAMAP: mapping marine mammals, birds and turtles," World Wide Web electronic publication. http://seamap.env.duke.edu. Accessed on August 8, p. 105.

Sato, T., Kasahara, J., Taymaz, T., Ito, M., Kamimura, A., Hayakawa, T., and Tan, O. (**2004**). "A study of microearthquake seismicity and focal mechanisms within the Sea of Marmara (NW Turkey) using ocean bottom seismometers (OBSs)," Tectonophysics **391**, 303–314.

Schevill, W. E., Watkins, W. A., and Backus, R. H. (**1964**). "The 20-cycle signals and Balaenoptera (fin whales)," in *Marine Bio-acoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 147–152.

Širović, A., Hildebrand, J. A., Wiggins, S. M., McDonald, M. A., Moore, S. E., and Thiele, D. (**2004**). "Seasonality of blue and fin whale calls and the influence of sea ice in the Western Antarctic Peninsula," Deep-Sea Res., Part II **51**, 2327–2344.

Sohn, R. A., Hildebrand, J. A., Webb, S. C., and Fox, Ch. G. (**1995**). "Hydrothermal microseismicity at the megaplume site on the southern Juan de Fuca Ridge," Bull. Seismol. Soc. Am. **85**(3), 775–786.

Spiesberger, J. L., and Wahlberg, M. (**2002**). "Probability density functions for hyperbolic and isodiachronic locations," J. Acoust. Soc. Am. **112**(6), 3046–3052.

Talandier, J., and Okal, E. A. (**1987**). "Seismic detection of underwater volcanism: The example of French Polynesia," Pure Appl. Geophys. **125**, 919–950.

Thode, A. (**2005**). "Three-dimensional passive acoustic tracking of sperm whales (Physeter macrocephalus) in ray-refracting environments," J. Acoust. Soc. Am. **118**, 3575–3584.

Thompson, P. O., Findley, L. T., and Vidal, O. (**1992**). "20-Hz pulses and other vocalizations of fin whales, *Balaenoptera physalus*, in the Gulf of California, Mexico," J. Acoust. Soc. Am. **92**(6), 3051–3057.

Tiemann, C. O., Porter, M. B., and Hildebrand, J. A. (**2002**). "Automated model-based localization of marine mammals near California," Oceans 2002 MTS/IEEE, pp. 1360–1364.

Walker, D. A., and Bernard, E. N. (**1993**). "Comparison of T-phase spectra and tsunami amplitudes for tsunamigenic and other earthquakes," J. Geophys. Res. **98**, 12557–12565.

Watkins, W. A., Tyak, P., Moore, K. E., and Bird, J. (**1987**). "The 20-Hz signals of fin whales (*Balaenoptera physalus*)," J. Acoust. Soc. Am. **82**(6), 1901–1912.

Zelt, C. A., and Smith, R. B. (**1992**). "Seismic traveltime inversion for 2-d crustal velocity structure," Geophys. J. Int. **108**, 16–34.

A forward model and conjugate gradient inversion technique for low-frequency ultrasonic imaging

Koen W. A. van Dongen[a] and William M. D. Wright[b]
Ultrasonics Research Group, Department of Electrical and Electronic Engineering, University College Cork, College Road, Cork, Ireland

(Received 31 August 2005; revised 19 July 2006; accepted 20 July 2006)

Emerging methods of hyperthermia cancer treatment require noninvasive temperature monitoring, and ultrasonic techniques show promise in this regard. Various tomographic algorithms are available that reconstruct sound speed or contrast profiles, which can be related to temperature distribution. The requirement of a high enough frequency for adequate spatial resolution and a low enough frequency for adequate tissue penetration is a difficult compromise. In this study, the feasibility of using low frequency ultrasound for imaging and temperature monitoring was investigated. The transient probing wave field had a bandwidth spanning the frequency range 2.5–320.5 kHz. The results from a forward model which computed the propagation and scattering of low-frequency acoustic pressure and velocity wave fields were used to compare three imaging methods formulated within the Born approximation, representing two main types of reconstruction. The first uses Fourier techniques to reconstruct sound-speed profiles from projection or Radon data based on optical ray theory, seen as an asymptotical limit for comparison. The second uses backpropagation and conjugate gradient inversion methods based on acoustical wave theory. The results show that the accuracy in localization was 2.5 mm or better when using low frequencies and the conjugate gradient inversion scheme, which could be used for temperature monitoring. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2336752]

PACS number(s): 43.60.Pt, 43.20.Dk, 43.20.El, 43.80.Qf, 43.60.Uv [TDM] Pages: 2086–2095

I. INTRODUCTION

Ultrasonic tomographic imaging[1–3] is a well-established technique for medical diagnosis, with the majority of ultrasonic imaging systems relying on backscattered ultrasonic energy to produce an image. There has long been an interest in using ultrasound for the imaging of temperature distributions in the human body (e.g., Johnson *et al.*[4]), particularly during high temperature hyperthermia cancer treatments such as high intensity focused ultrasound (HIFU).[5,6] During this treatment, necrosis of cancer cells is obtained by increasing the tumor temperature to 50–55 °C for a duration of 1 or 2 min or over a shorter period of time for temperatures over 60 °C.[7,8] For successful treatment, it is important that the tumor temperature is sufficiently high to induce necrosis while the normal tissue surrounding the tumor remains at or near normal body temperature to prevent excessive damage. The noninvasive nature of this method results in a lack of direct visual control and therefore requires a monitoring system. Ultrasound as a guidance method has the advantages of being relatively cheap and compatible with HIFU apparatus.

Several methods of ultrasonic estimation of temperature using backscattered ultrasound have been devised, with the majority of techniques based on the observation that biological tissues can be described by semiregular lattices of which the scattering properties change as a function of temperature[9–13] and others based on thermally induced strain.[14–16] As the acoustic contrast between many different soft tissues (or between regions of the same tissue at different temperatures) is relatively small, using backscattered ultrasound to provide the necessary quantitative information for clinical diagnosis or temperature estimation is often difficult. Under these circumstances, other imaging methods utilizing through-transmission may be more appropriate,[17] which have also been used for temperature estimation.[18,19]

Such imaging techniques typically require a compromise between using a high enough frequency for adequate spatial resolution, and a low enough frequency for adequate tissue penetration. It would seem, therefore, that certain benefits could be obtained by imaging at low frequencies, and this has been used with some success for strongly reflective or scattering media such as bone,[20,21] and other applications.[22] The focus of the current study is to use simulation to assess the feasibility of using low frequency (2.5–320 kHz) and therefore highly penetrating ultrasound to image the weakly scattering contrasts expected during HIFU treatment, where a change in temperature from 37 to 50 °C typically produces a contrast in sound speed of only 6 m/s.[23,14,15]

The various imaging methods using diagnostic ultrasound that have been investigated previously are based on different approximations and are consequently not usually compared directly with each other. Common assumptions are that the Born approximation holds, and that only backscattered ultrasound is considered, as this greatly simplifies the mathematics of the inverse scattering problem,[1] a topic still under much scrutiny.[24] In this study, both the transmitted and the backscattered ultrasound are considered together in a full solution. Two types of methods that could be used for measuring the temperature distributions or contrast functions in

[a]Electronic mail: koen@rennes.ucc.ie
[b]Electronic mail: bill.wright@ucc.ie

and around the tumor with low-frequency ultrasound are compared using simulation. The first is computed tomography[25] and is based on optical ray theory. Here, filtered backprojection, Fourier techniques, or algebraic reconstruction techniques are used to relate projection data, i.e., time delays obtained via tomographic measurements, to temperature distributions.[18,26] Good comparisons between the methods within this group have been published by Nawata[18] and Kak *et al.*[25] Despite the obvious disadvantages of using straight ray or simple diffraction assumptions, the major advantage of these techniques is that they require little processing and therefore are fast and easy to implement. As such, there are still many practical applications and recently a system was proposed where these methods will be used for early breast cancer diagnosis.[27] However, most of these techniques require large data sets and tend to neglect the combined effects of diffraction, attenuation, and multiple scattering. As a representative method from this group of techniques, the Fourier slice technique[25,28] has been selected for comparison in this study.

The second group of imaging methods uses acoustical wave theory as the starting point for which various inversion methods have been developed.[29] Here, backpropagation and a conjugate gradient iterative inversion scheme[30,31] have been chosen as representative examples.

Testing of the imaging algorithms is done using synthetic data. These data are obtained by solving the forward (scattering) problem for known contrast functions representing regions positioned in a homogeneous background medium. In some applications, the approach for solving the forward problem is to neglect effects like diffraction and to use far field or Born approximations as done, e.g., in Refs. 32 and 36. However, in the past few years, various methods have been developed to compute synthetic data which take these effects into account, e.g., by using finite element techniques,[33] finite difference time domain techniques,[34] or by solving the representative integral equations via conjugate gradient (CG) inversion schemes.[35–37] In this study, the conjugate gradient technique is used to solve the forward problem in the temporal Laplace domain. In this way, the same computational scheme is used for both the forward problem and the imaging. Fourier transformations are used to obtain synthetic time domain data.

II. FORWARD PROBLEM

The forward problem is solved in the temporal Laplace domain with Laplace parameter $\hat{s}$. Hence, frequency domain results are obtained by taking the limit $\hat{s} \rightarrow -i\omega$, with $i^2 = -1$ and ω the temporal angular frequency. The caret symbol is used to show the frequency dependence of a parameter. In addition, a position in the spatial domain $\mathbb{R}^3$ is notated by the vector $\mathbf{x}$.

By applying reciprocity[38] on the acoustic wave field equations it can be shown that the total pressure wave field $\hat{p}^{\text{tot}}(\mathbf{x})$ and the total velocity wave fields $\hat{v}_k^{\text{tot}}(\mathbf{x})$ and $\hat{v}_j^{\text{tot}}(\mathbf{x})$ for $\{k,j\} = 1$, 2, or 3 equals

$$\hat{p}^{\text{tot}}(\mathbf{x}) = \hat{p}^{\text{inc}}(\mathbf{x}) + \hat{p}^{\text{sct}}(\mathbf{x}), \tag{1}$$

$$\hat{v}_k^{\text{tot}}(\mathbf{x}) = \hat{v}_k^{\text{inc}}(\mathbf{x}) + \hat{v}_k^{\text{sct}}(\mathbf{x}), \tag{2}$$

where $\hat{p}^{\text{inc}}(\mathbf{x})$ and $\hat{v}_k^{\text{inc}}(\mathbf{x})$ are the incident pressure and velocity fields and where $\hat{p}^{\text{sct}}(\mathbf{x})$ and $\hat{v}_k^{\text{sct}}(\mathbf{x})$ refer to the scattered pressure and velocity fields. In the presence of objects represented by contrasts in the acoustic medium parameters compressibility κ and volume density of mass ρ, these scattered wave fields are equal to

$$\hat{p}^{\text{sct}}(\mathbf{x}) = \hat{G}^{pq}(\mathbf{x},\mathbf{x}')[\Delta\hat{\eta}(\mathbf{x}')\hat{p}^{\text{tot}}(\mathbf{x}')] + \sum_{j=1}^{3} \hat{G}_j^{pf}(\mathbf{x},\mathbf{x}') \times [\Delta\hat{\zeta}(\mathbf{x}')\hat{v}_j^{\text{tot}}(\mathbf{x}')], \tag{3}$$

$$\hat{v}_k^{\text{sct}}(\mathbf{x}) = \hat{G}_k^{vq}(\mathbf{x},\mathbf{x}')[\Delta\hat{\eta}(\mathbf{x}')\hat{p}^{\text{tot}}(\mathbf{x}')] + \sum_{j=1}^{3} \hat{G}_{k,j}^{vf}(\mathbf{x},\mathbf{x}') \times [\Delta\hat{\zeta}(\mathbf{x}')\hat{v}_j^{\text{tot}}(\mathbf{x}')], \tag{4}$$

where the contrasts $\Delta\hat{\eta}(\mathbf{x})$ and $\Delta\hat{\zeta}(\mathbf{x})$ are defined by the acoustic medium parameters of the background medium (bg) and the object medium (obj) via

$$\Delta\hat{\eta}(\mathbf{x}') = \hat{s}(\kappa^{\text{bg}} - \kappa^{\text{obj}}(\mathbf{x}')), \tag{5}$$

$$\Delta\hat{\zeta}(\mathbf{x}') = \hat{s}(\rho^{\text{bg}} - \rho^{\text{obj}}(\mathbf{x}')), \tag{6}$$

while the Green's tensor operators applied to a volume density of injection rate source $\hat{q}(\mathbf{x})$ or a volume density of force $\hat{f}_j(\mathbf{x})$ read as follows:

$$\hat{G}^{pq}(\mathbf{x},\mathbf{x}')[q(\mathbf{x}')] = \hat{s}\rho^{\text{bg}}\hat{G}(\mathbf{x},\mathbf{x}') * q(\mathbf{x}'), \tag{7}$$

$$\hat{G}_j^{pf}(\mathbf{x},\mathbf{x}')[f_j(\mathbf{x}')] = -\partial_j\hat{G}(\mathbf{x},\mathbf{x}') * f_j(\mathbf{x}'), \tag{8}$$

$$\hat{G}_k^{vq}(\mathbf{x},\mathbf{x}')[q(\mathbf{x}')] = -\partial_k\hat{G}(\mathbf{x},\mathbf{x}') * q(\mathbf{x}'), \tag{9}$$

$$\hat{G}_{k,j}^{vf}(\mathbf{x},\mathbf{x}')[f_j(\mathbf{x}')] = \frac{1}{\hat{s}\rho^{\text{bg}}}[\partial_k\partial_j\hat{G}(\mathbf{x},\mathbf{x}') * f_j(\mathbf{x}') + \delta_{k,j}\delta(\mathbf{x}-\mathbf{x}') * f_j(\mathbf{x}')], \tag{10}$$

with ∂_k the partial derivative in the x_k direction, $\delta_{k,j}$ Kronecker's delta function, $\delta(\mathbf{x}-\mathbf{x}')$ the three-dimensional (3D) Dirac delta function and with $\hat{G}(\mathbf{x},\mathbf{x}')$ the scalar form of Green's function. Note that, $g(\mathbf{x},\mathbf{x}') * h(\mathbf{x}')$ defines a convolution over the spatial domain $\mathbb{D}$ containing the contrasts and the transducers. Hence, in the absence of any contrast, the acoustic wave field will travel with the speed of sound c,

$$c = \frac{1}{(\kappa^{\text{bg}}\rho^{\text{bg}})^{1/2}}, \tag{11}$$

through the medium.

This integral formulation is applied to the situation shown in Fig. 1. Here, an object is positioned in a homogeneous background medium, while around the object at equiangular positions β_m a set of transducers is positioned. In the mono-static situation, only one transducer is used which acts as both a transmitter and a receiver, hence $\mathbf{x}^{\text{src}} = \mathbf{x}^{\text{rec}}$. Multiple measurements are made by rotating the transducer around

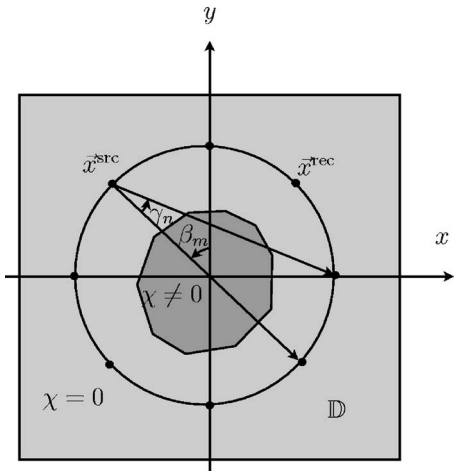


FIG. 1. The setup used for fan-beam measurements. A transmitter positioned at $\mathbf{x}^{\mathrm{src}}$ will create an acoustic wave field which is observed by the receivers located at $\mathbf{x}^{\mathrm{rec}}$.

the object, or vice-versa. For tomographic fan-beam measurements, a multi-static setup is used. Therefore, one transducer will act as a transmitter while the remainder act as receivers. Again, a complete set of measurements is obtained by rotating the transmitter and receivers around the object.

Equations (3) and (4) form an integral equation of the second type. After discretization, these equations are solved by using a conjugate gradient inversion scheme, as described by Refs. 36, 37, 39, and 40. The method includes a recursive procedure to compute the unknowns $\hat{p}^{\mathrm{tot}}(\mathbf{x})$ and $\hat{v}_k^{\mathrm{tot}}(\mathbf{x})$ with Polak-Ribière conjugate gradient update directions. The scheme is based on the minimization of the L_2-norm of an error functional. This error functional is defined as the difference between the known incident fields $\hat{p}^{\mathrm{inc}}(\mathbf{x})$ and $\hat{v}_k^{\mathrm{inc}}(\mathbf{x})$ and the approximated incident fields based on the computed scattered and total fields at the nth iteration step, $\hat{p}^{\mathrm{tot}}(\mathbf{x}) - \hat{p}^{\mathrm{sct}}(\mathbf{x})$ and $\hat{v}_k^{\mathrm{tot}}(\mathbf{x}) - \hat{v}_k^{\mathrm{sct}}(\mathbf{x})$. This method is usually referred to as the "conjugate gradient method applied to the normal equations."

III. INVERSE PROBLEM

The aim of imaging is to localize contrasts and to compute speed of sound profiles from which temperature distributions in the region of interest can be obtained. In this paper, three imaging methods are compared. The first imaging method is described well by Kak and Slaney[25] and is based on the Fourier slice theorem and uses changes in travel times as observed in tomographic fan-beam projection data. The second and third methods are based on the acoustic wave equations. In these cases, the Born approximation is used to find an approximate solution of an integral equation similar to the one used for the forward modeling. Both a single step and an iterative scheme are considered.

A. Fourier slice method

The fan-beam slice theorem[25] is based on optical ray theory, where only changes in time delays are taken into account. These time delays are related to changes in the medium parameters density and compressibility, while effects like attenuation and diffraction are neglected. Within these approximations, the direct problem is solved by computing the two-dimensional (2D) Radon transform,

$$P_\beta(\gamma_n) = \int_{(x,y)=-\infty}^{\infty} c^{-1}(x,y)\,\delta\left(x\cos\left(\frac{1}{2}\pi-\beta-\gamma\right) + y\sin\left(\frac{1}{2}\pi-\beta-\gamma\right) - R\sin(\gamma)\right)dxdy, \tag{12}$$

with R the radius of the circle defining the transducer positions, β the angle between the y axis and the transmitter, and γ_n the position of a ray in a fan-beam as shown in Fig. 1.

Using the formulation for the forward problem as given in Eq. (12), images are obtained in three steps. First the projection (Radon) data are multiplied by $R\cos(\gamma_n)$, hence

$$P'_{\beta_i}(\gamma_n) = P_\beta(\gamma_n) R\cos(\gamma_n). \tag{13}$$

Next, the modified projection data are filtered by convolving $P'_{\beta_i}(\gamma_n)$ with $g(\gamma_n)$,

$$Q_{\beta_i}(\gamma_n) = P'_\beta(\gamma_n) * g(\gamma_n), \tag{14}$$

where the filter function reads

$$g(\gamma_n) = \begin{cases} \dfrac{1}{8(\Delta\gamma)^2} & \text{when } n=0 \\ 0 & \text{when } n \text{ is even} \\ -\dfrac{1}{2}\left(\dfrac{1}{\pi\sin(\gamma_n)}\right)^2 & \text{when } n \text{ is odd.} \end{cases} \tag{15}$$

To improve the image quality, further filtering is often applied,

$$Q_{\beta_i}(\gamma_n) = P'_\beta(\gamma_n) * g(\gamma_n) * k(\gamma_n), \tag{16}$$

where $k(\gamma_n)$ is typically a cosine, Hamming, or Shepp-Logan filter. Finally, the amplitude of the reciprocal of the sound-speed profile is obtained by summing all the normalized filtered projections,

$$c^{-1}(x,y) = \Delta\beta\sum_{i=1}^{M} \frac{1}{|\mathbf{x}^{\mathrm{src}}(\beta_i)-(x,y)|^2} Q_{\beta_i}(\gamma'_n) \times \delta\left(x\cos\left(\frac{1}{2}\pi-\beta-\gamma\right) + y\sin\left(\frac{1}{2}\pi-\beta-\gamma\right) - R\sin(\gamma)\right), \tag{17}$$

where $|(x,y)|^2$ equals the L_2-norm of a vector. Note that the filter function $g(\gamma_n)$ shown in Eq. (15) is slightly different from that shown previously by Kak and Slaney[25] where $g(\gamma_n)$ reads

$$g(\gamma_n) = \left(\frac{1}{\pi\sin(\gamma_n)}\right)^2, \tag{18}$$

when n is odd, and was obtained after a thorough analysis of the original work presented in Ref. 25.

B. Backpropagation and conjugate gradient technique

The second group of reconstruction methods is based on the acoustic wave equation formulation used for the forward problem. Starting with Eq. (3) and assuming that there is only contrast in compressibility ($\Delta\hat{\zeta}(\mathbf{x})=0$) and that the problem is defined within the Born approximation, results in an approximation of the scattered pressure field which reads

$$\hat{p}^{\mathrm{sct}}(\mathbf{x}) = \hat{G}^{pq}(\mathbf{x},\mathbf{x}')\hat{s}\chi(\mathbf{x}')\hat{p}^{\mathrm{inc}}(\mathbf{x}'). \tag{19}$$

An estimate for $\chi(\mathbf{x}')$ is retrieved from a minimum norm solution of the error functional Err, where Err reads

$$\mathrm{Err} = \sum_{s,r,\omega} |\hat{p}^{\mathrm{sct}}(\mathbf{x}) - \hat{G}^{pq}(\mathbf{x},\mathbf{x}')\hat{s}\chi(\mathbf{x}')\hat{p}^{\mathrm{inc}}(\mathbf{x}')|^2, \tag{20}$$

with the subscripts s and r referring to a summation over all possible source and receiver positions. Minimization of the error functional Err in a single step is in literature referred to as backpropagation. With backpropagation, we approximate $\chi(\mathbf{x}')$ by

$$\chi(\mathbf{x}') = \alpha\Delta\chi(\mathbf{x}'), \tag{21}$$

where α is a real constant. The error functional Err in Eq. (20) tends toward a minimum when α reads

$$\alpha = \frac{\Re\left(\sum_{s,r,\omega} \hat{G}^{pq}(\mathbf{x},\mathbf{x}')\hat{s}\Delta\chi(\mathbf{x}')\hat{p}^{\mathrm{inc}}(\mathbf{x}')[\hat{p}^{\mathrm{sct}}(\mathbf{x}')]^*\right)}{\sum_{s,r,\omega} |\hat{G}^{pq}(\mathbf{x},\mathbf{x}')\hat{s}\Delta\chi(\mathbf{x}')\hat{p}^{\mathrm{inc}}(\mathbf{x}')|^2}, \tag{22}$$

where $\Re(\ldots)$ is used to denote that the real part is taken. We observe that, apart from a constant, the numerator is maximized by taking as update direction

$$\Delta\chi(\mathbf{x}') = \sum_{s,r,\omega} (\hat{p}^{\mathrm{inc}}(\mathbf{x}')\hat{s}^2\rho\hat{G}(\mathbf{x},\mathbf{x}'))^*\hat{p}^{\mathrm{sct}}(\mathbf{x})\Delta V, \tag{23}$$

with ΔV the size of a volume element and where $(\ldots)^*$ is used to denote that the complex conjugate of the operator is taken. Substituting this in Eq. (22) we obtain for α,

$$\alpha = \frac{\sum_{s,r,\omega} \hat{p}^{\mathrm{sct}}(\mathbf{x}')[\hat{p}^{\mathrm{sct}}(\mathbf{x}')]^*}{\sum_{s,r,\omega} |\hat{p}^{\mathrm{sct}}(\mathbf{x}')|^2}. \tag{24}$$

Backpropagation is similar to the first step in a CG inversion scheme. With the CG method, the set of equations present in Eq. (19) is solved iteratively by minimizing the error functional Err in Eq. (20).[36,37]

Since the Fourier method is a 2D reconstruction method, the computation of $\Delta\chi(\mathbf{x}')$ is restricted to the plane containing the transducers. Finally, a speed of sound profile is obtained from the computed contrast profile via

$$c(\mathbf{x}) = \left(\frac{1}{|\kappa^{\mathrm{bg}} - \Re(\chi)|\rho^{\mathrm{bg}}}\right)^{1/2}, \tag{25}$$

where it is assumed that changes in the density are compensated for by changes in the compressibility. This is based on the observation that the scattered pressure field is partly caused by the velocity field via changes in density, see the second term on the right-hand side of Eq. (3). However, this term is neglected in Eq. (19) and will be compensated via additional changes in the compressibility.

TABLE I. The medium parameters compressibility, κ, volume density of mass, ρ, and speed of sound, c, for various tissues.

	Compressibility $k[10^{-9}(\mathrm{Pa})^{-1}]$	Density $\rho[\mathrm{kg/m^3}]$	Speed of sound $c[\mathrm{m/s}]$
Liver at 37 °C	0.366 48	1056.6	1607
Liver at 45 °C	0.105 33	1053.3	1612
Liver at 50 °C	0.365 71	1051.0	1613
Fat	0.481 9	950.0	1478

IV. RESULTS AND DISCUSSION

Synthetic data for three different configurations was computed by solving Eqs. (3) and (4) in the frequency domain and transforming the obtained results into time domain using fast Fourier transforms (FFTs). In all cases, a homogeneous background medium was used with acoustic medium parameters similar to liver, in order that a good impression of the accuracy of the methods employed could be obtained. In practice, the background medium is unlikely to be homogeneous, but the techniques employed may also use an inhomogeneous background medium provided that its profile is known in advance of any heating, i.e., the changes in the starting profile are related to the changes in temperature. Various contrasts were present in the homogeneous background medium used in this simulation, representing regions of either fat or heated liver tissue. The corresponding medium parameters are shown in Table I and were obtained by combining the results presented by Refs. 14, 15, and 23. Note that the changes in speed of sound due to heating are relatively small, viz. 5 m/s between 37 and 45 °C and 1 m/s between 45 and 50 °C. The three-dimensional (3D) volume under investigation contained $64\times64\times8$ elements of size $2.5\times2.5\times2.5$ mm^3. Each real valued temporal signal contained 256 points, with 1.56 μs temporal step size. All acoustic sources were defined via a real valued volume density of injection rate with amplitude equal to one in the temporal Laplace domain.

Typical results for the incident, scattered, and total pressure wave fields for a configuration containing two heated regions are shown in Fig. 2. The heated regions form a known semi-infinite 3D contrast profile of which a cross section is shown in Fig. 2(a). The results show that the amplitude of the scattered wave field is one order of magnitude lower than the magnitude of the incident wave field. Consequently, the presence of the objects is hardly visible in the total wave field, despite the fact that all spatial dimensions of all the objects are bigger than the smallest wavelength present in the probing signal.

Projection data were obtained by computing the total wave fields for 36 transmitter positions marked by an open circle in Fig. 3(a). Next, standard cross correlation was used to compute the time delays between the incident and the total pressure wave fields at the remaining receiver positions. To

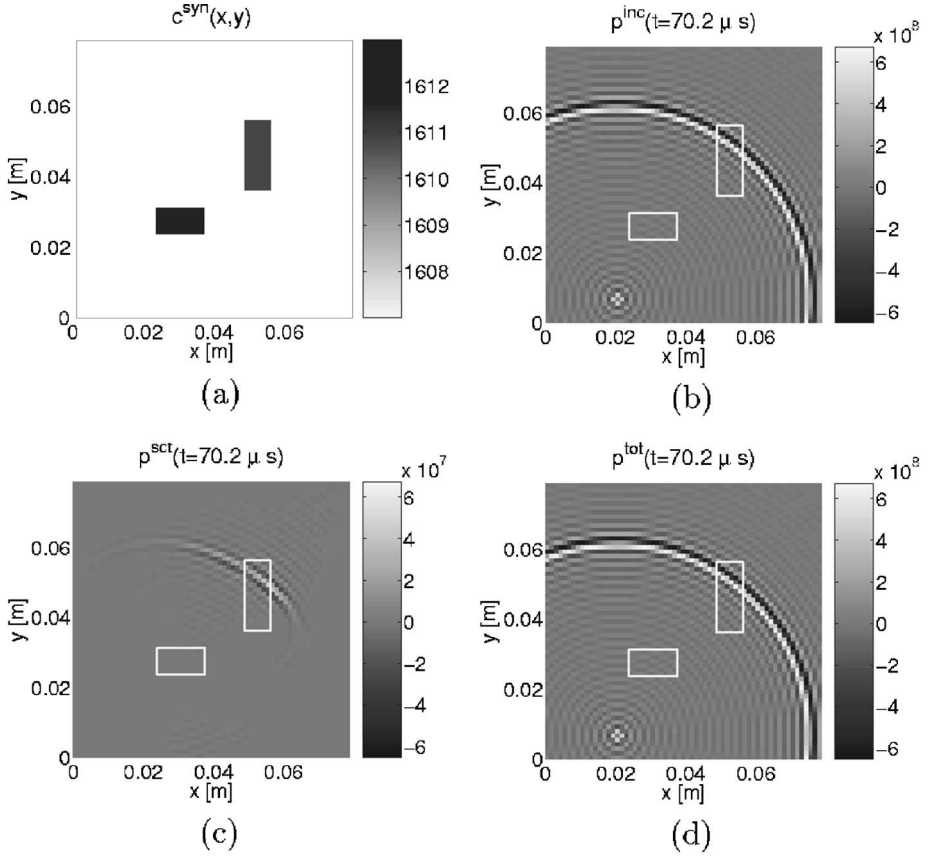


FIG. 2. Due to the presence of various contrasts (a), the incident wave field (b)generates a scattered field (c), which combined results in the total wave field (d). The boxes represent the location of the objects.

increase the temporal resolution, cubic spline interpolation was applied to both time domain signals to increase the number of samples by a factor of 1000. An example for the result obtained with cross correlation applied to the signals shown in Fig. 3(b) is given in Fig. 3(c), showing a time delay of 11 ns. The signals were obtained from the transmitter-receiver combination indicated by the line in Fig. 3(a). Repeating this procedure for all transmitter-receiver combinations resulted in the Radon transform shown in Fig. 4(a). For the same configuration the (ideal) Radon transform was also computed within the optical ray approximation using Eq. (12). The obtained results are shown in Fig. 4(b). Comparing the results, it is observed that the images show similar patterns, but with different amplitudes. Hence, time delays are observed for the same transmitter-receiver combinations, but with smaller time delays for the full solution. This is mainly due to diffraction which is not taken into account in the optical ray approximation.

Projection data as shown in Fig. 4 were used as input data for the Fourier slice algorithm. It is found that the image quality improves by correcting the projection data before processing for the background medium, i.e., when changes in a known starting background profile are being imaged. Hence, the projection $P_\beta(\gamma_n)$ used for the imaging reads

$$P_\beta(\gamma_n) = P_\beta^{\text{meas}}(\gamma_n) - P_\beta^{\text{bg}}(\gamma_n), \tag{26}$$

with $P_\beta^{\text{meas}}(\gamma_n)$ the measured projection and $P_\beta^{\text{bg}}(\gamma_n)$ the projection computed in the absence of any objects based on the known background medium parameters.

In total, three different configurations were used for testing the imaging algorithms. In the first configuration, two regions are present which are different in size and have different sound-speeds. A cross section of the sound-speed profile is shown in Fig. 5(a) while the corresponding Radon transforms are given in Fig. 4. Results with the Fourier method based on data within the optical ray approximation are shown in Fig. 5(b) and show that the regions are clearly present in the reconstructed sound-speed profiles with correct magnitudes. It is found that by increasing the number of measurements and decreasing the grid size the high amplitude noise at the edge of the image near the transducers will decrease and that sharper edges defining the regions will be obtained. However, the optical ray approximation effectively ignores diffraction. Next, the same Fourier method was used on Radon data based on the full solution of the forward problem. The reconstructed speed of sound profile from this data is shown in Fig. 5(c) and shows that both regions are localized. However, comparing these results with the reconstructions from data obtained via the line integrals shows that the region boundaries have become more blurred and that the reconstructed sound-speeds have decreased. This is in accordance with the results shown in Fig. 4, where the amplitude of the Radon data based on a full solution of the forward problem is smaller than the amplitude of the Radon data based on the line integral. Note that the region with the smallest spatial dimension but with the highest speed of

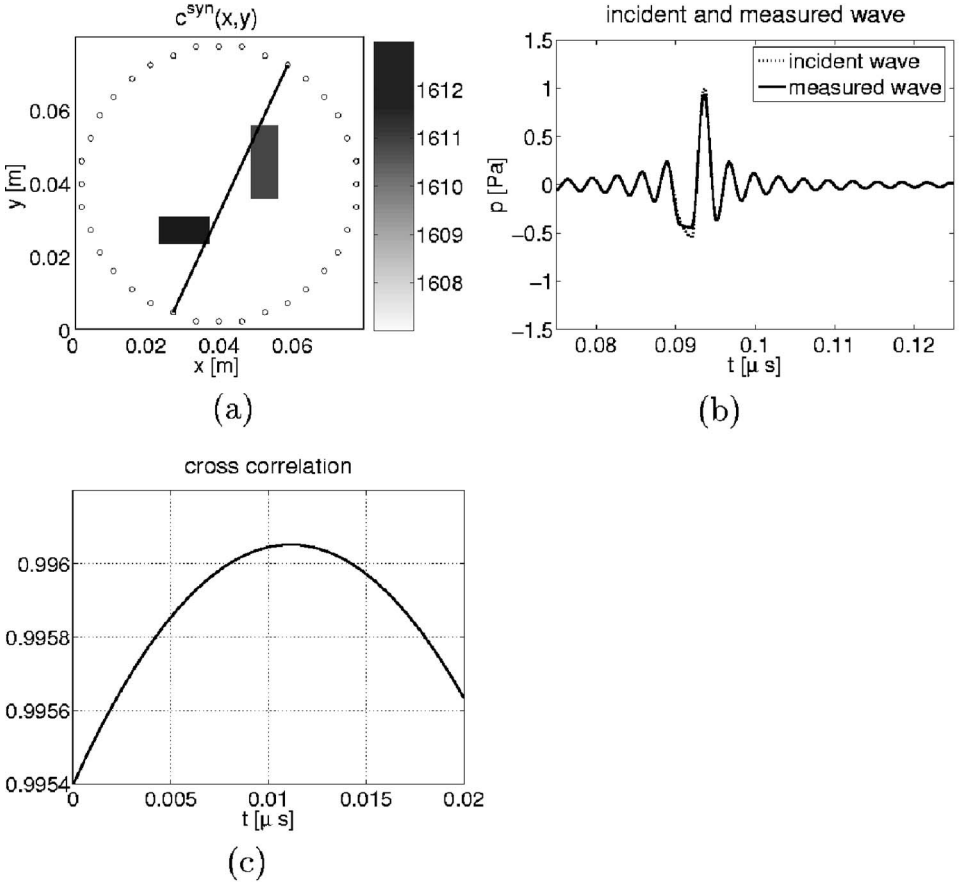


FIG. 3. For one transmitter-receiver combination, indicated by the line in (a), the incident and measured waves are obtained (b). Standard cross correlation is used to compute the time delay for this transducer combination (c).

sound has the lowest sound-speed in the reconstructed image. Next, we transformed the time domain data based on the full solution of the forward problem back to the frequency domain. Hence, each real valued temporal signal resulted in 127 complex valued data points in the frequency domain where we omitted the zero and highest frequency component of the signal. This corresponds to acoustic frequencies from 2.5 up to 320.5 kHz. On this frequency domain data set we first applied backpropagation where we used all frequency components. The resulting image is shown in Fig. 5(d). With this method the regions are localized and less blurred when compared to the results obtained with the Fourier method. However, the amplitudes of the reconstructed sound-speed profiles are lower than in the original profiles. The results improve by using the conjugate gradient method on the same complex valued frequency domain data. After only three iterations, the edges have become sharper and the magnitudes of the reconstructed velocities are higher as can be observed from Fig. 5(e). Due to the roughness in amplitudes of the reconstructed heated regions the error in some individual elements can be as high 2 m/s, therefore, the amplitude is clipped at some locations in the image. More iterations do not improve the image.

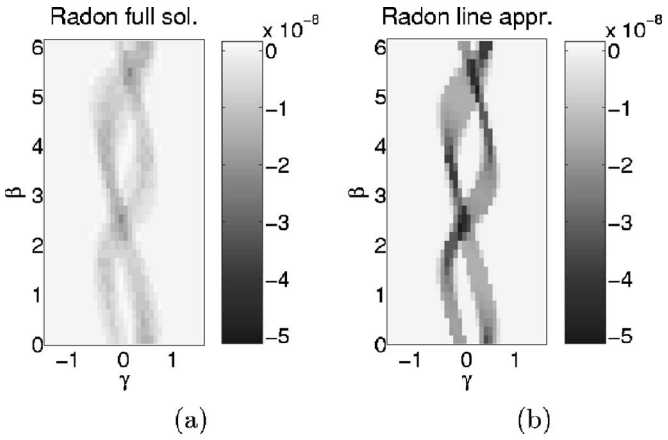


FIG. 4. The Radon transformation based on (a) the full solution of the forward problem and on (b) optical ray theory.

From the above-presented results, it is observed that the size of a region influences the amplitude of the reconstructed image in the case where data based on the full solution of the forward problem are used. In order to verify this observation, a configuration was taken where there are three regions with different spatial dimensions but identical sound-speeds (all regions represent tissue heated to 50 °C). Applying the same computational schemes resulted in the contrast profiles shown in Fig. 6. The results show that with the Fourier method and the full solution data, the smallest region is not localized, while with the backpropagation and conjugate gradient method it is localized successfully. It also shows that the amplitude of the reconstructed sound-speed profiles indeed depend on the size of the region. These results suggest that, with the CG method, smaller regions can be detected than with the Fourier method. Consequently, a third configuration has been considered where there is a small region of the same dimension with high acoustic contrast, i.e., a small region of fat. The results for this configuration are shown in Fig. 7. In this case, only the Fourier method is capable of

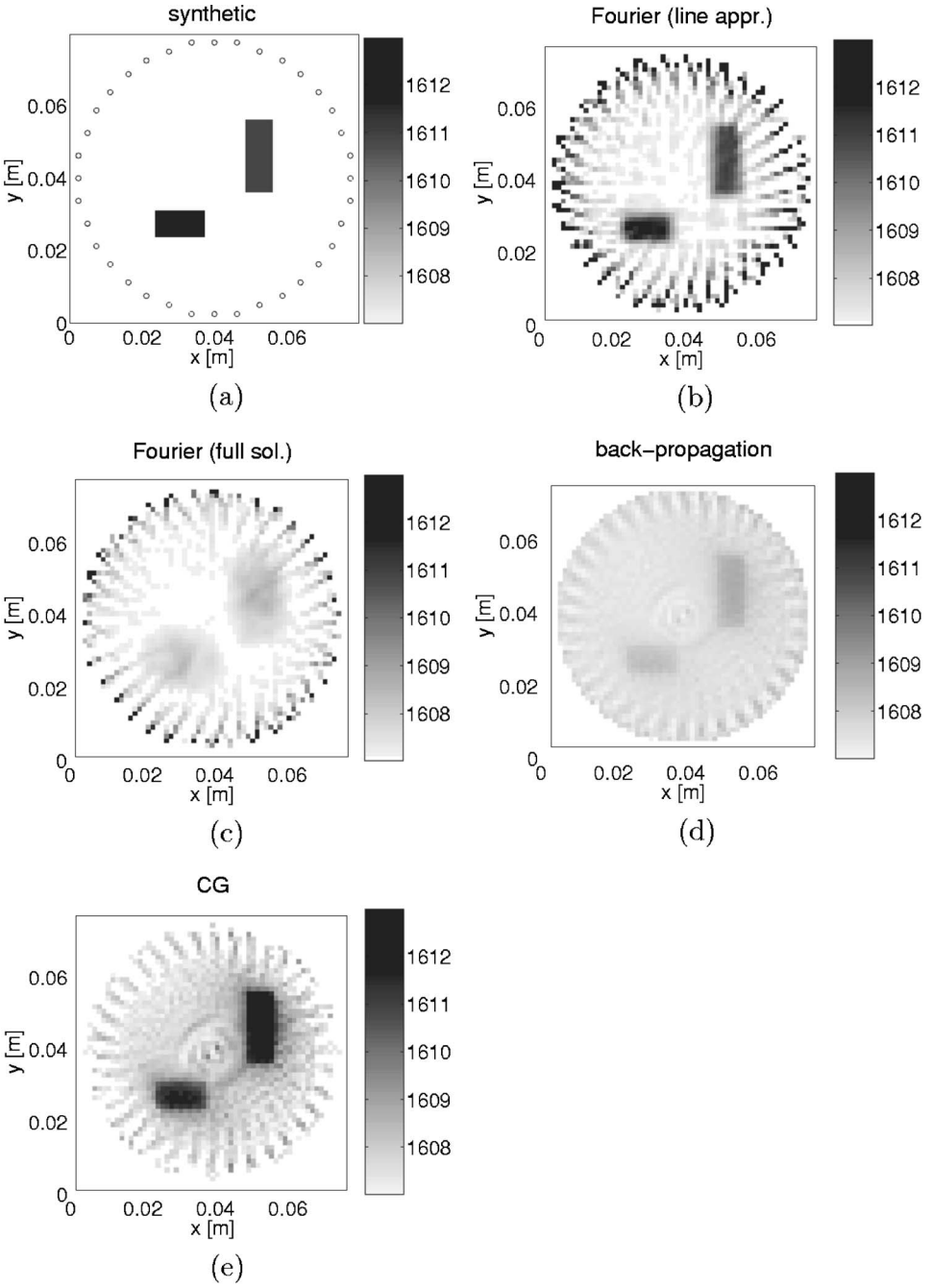


FIG. 5. The cross section (a) of a semi-infinite 3D known contrast function based on two different heated regions. Next, two images are shown with results obtained by applying the Fourier technique on synthetic data (b) computed via line integrals and (c) computed by solving the full vectorial forward problem. The results improve when (d) backpropagation or an (e) conjugate gradient scheme is applied.

localizing the region of fat, while the remaining two heated regions disappear in the background noise. With the backpropagation and conjugate gradient method, none of the regions are localized due to the noise present in the image. This noise is caused by the combination of the high frequency content of the signal and the high contrast of the fat. This problem is solved by using the first 32 frequencies (i.e., up to 20 kHz) of the signal at the expense of the obtained spatial resolution. Hence, the location of the fat region becomes clear in the backpropagation image, while the reconstruction obtained with the CG method also reveals the location of the heated regions after applying 32 iterations. The reconstructed speeds of sound in the regions are higher than the expected values due to the presence of the high contrast of the fat region. The images are expected to improve if this contrast is taken into account in the initial background profile, as such a high contrast would already be observed prior to heating.

In order to demonstrate the capabilities of the imaging methods we have chosen contrasts with sharp edges. In reality a heated region would have smooth edges and a Gaussian temperature distribution. If this is taken into account during reconstruction it would result in a decrease of the noise in the distribution and therefore a further increase of the accuracy of the reconstruction.

Finally, it should be noted that the Fourier method has the advantage over the backpropagation or CG method of being relatively easy to implement; as the algorithm is less complex and the required Radon data are easier to obtain by applying simple cross-correlation techniques on measured

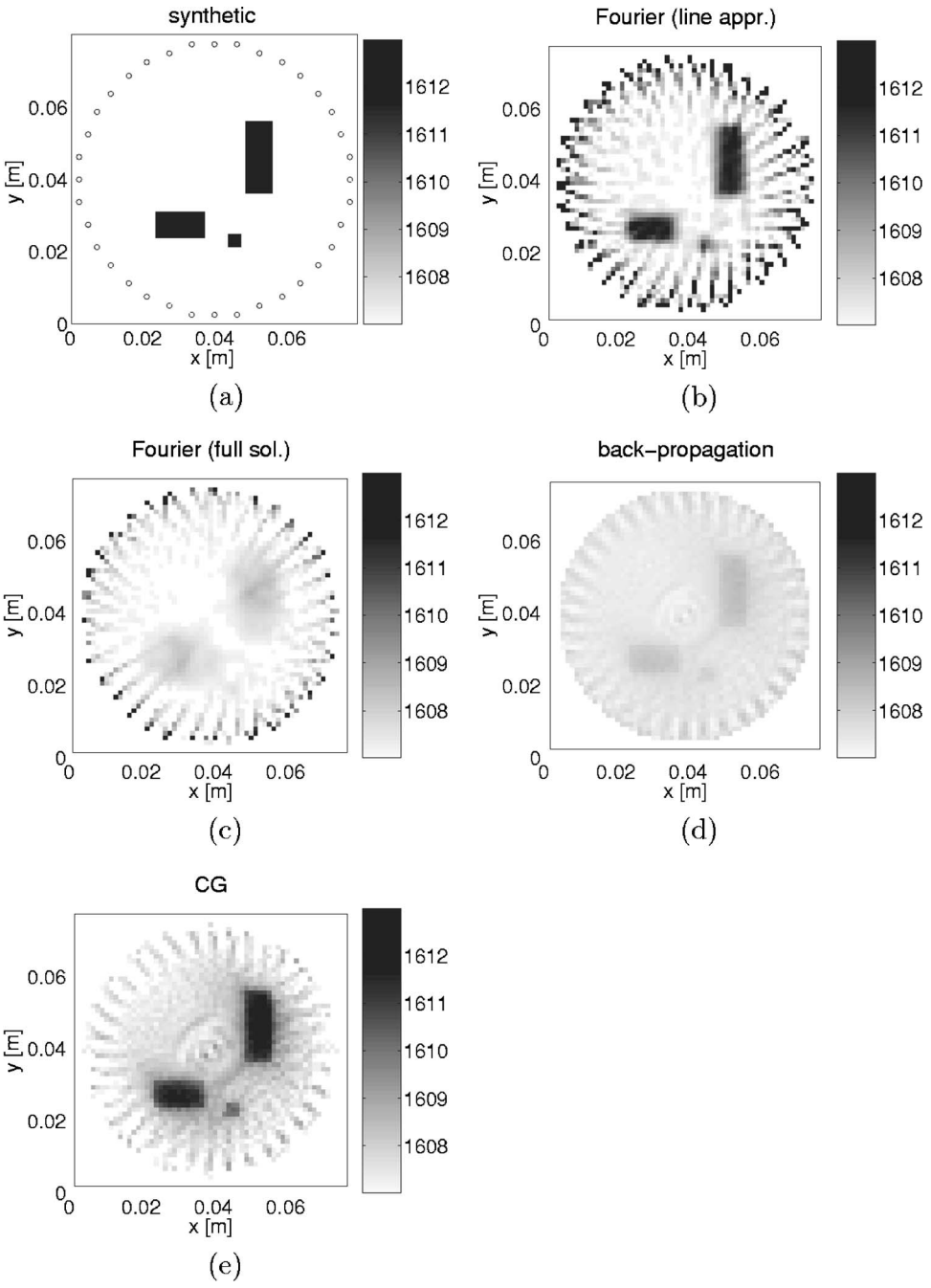


FIG. 6. The cross section (a) of a semi-infinite 3D known contrast function based on three heated regions. Next, two images are shown with results obtained by applying the Fourier technique on synthetic data (b) computed via line integrals and (c) computed by solving the full vectorial forward problem. The results improve when (d) backpropagation or an (e) conjugate gradient scheme is applied.

tomographic data. In addition, the Fourier method has the advantage of being computationally less expensive than the backpropagation or the CG method.

V. CONCLUSION

The feasibility of using low frequency (2.5–320 kHz) ultrasound to image the weakly scattering contrasts expected during HIFU treatment has been investigated. Representative imaging algorithms from two groups of imaging methods have been compared and tested on the same synthetic data set. Both groups are usually used independently and have not been compared directly. The first group uses Radon or projection data to compute 2D sound-speed profiles of the volume of interest. The reconstruction methods rely on optical ray theory and the Fourier slice technique was taken as a representative example from this group of imaging methods. The second group uses the acoustic wave equations as starting point. A backpropagation and a conjugate gradient inversion scheme were considered, both defined within the Born approximation. All schemes were restricted to the 2D situation.

Three configurations were used to test the imaging methods on (weakly) scattering contrasts which typically occur during hyperthermia cancer treatment. In all cases, the contrasts represented regions of fat or heated tissue which are embedded in a homogeneous background medium with acoustic medium properties equivalent to that of human liver at 37 °C. Synthetic data were computed by solving the complete vectorial forward problem for both the pressure and velocity wave fields, taking changes in compressibility and volume density of mass into account. Only the pressure wave fields were used for imaging.

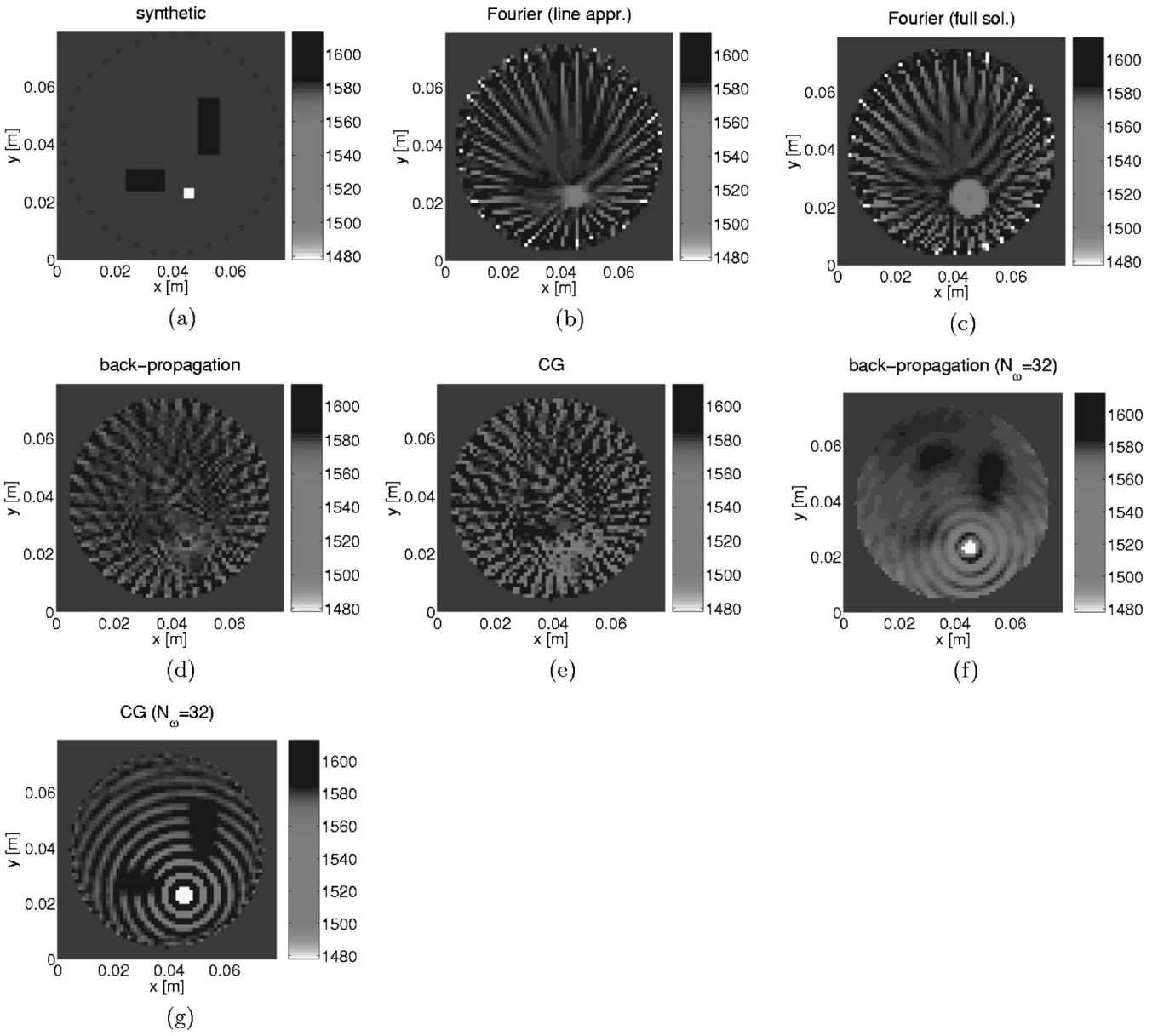


FIG. 7. The cross section (a) of a semi-infinite 3D known contrast function based on two different heated regions and one fat region. Next, two images are shown with results obtained by applying the Fourier technique on synthetic data (b) computed via line integrals and (c) computed by solving the full vectorial forward problem. The results obtained with backpropagation and a conjugate gradient scheme are shown in (d)–(g).

The results obtained show that all methods could detect changes in speed of sound due to an increase in temperature. Poor spatial resolution of the object was obtained with the Fourier method. In addition, the amplitudes of the reconstructed speed of sound profiles were lower than their synthetic values. To compare, in the situation where there was only one type of contrast, with an increase in speed of sound at the object location from 1607 to 1613 m/s the reconstructed profile showed only an increase up to 1609 m/s. With the backpropagation method comparable results were obtained for the amplitudes of the reconstructed sound-speed profiles. However, the results showed an increase in spatial resolution. With the conjugate gradient method the spatial resolution improved even further, while the reconstructed sound-speed values became too high; for the same configuration the increase was up to 1614 m/s.

Increases in temperature for regions as small as the shortest wavelength present in the probing signal were only detected with the backpropagation and conjugate gradient methods. In the situation where the same small object was a small region of fat with high acoustic contrast, the Fourier method was capable of localizing the fat region but not the heated regions. Both the backpropagation and the CG method could reveal the location of the fat region after removing the highest frequencies present in the measured signal, while of all methods only the CG method could reveal the locations of the heated regions.

The work has demonstrated that it is feasible to use low frequency ultrasound to image the small weakly scattering acoustic contrasts that would be expected to occur during hyperthermia cancer treatment, and that a good accuracy is obtained when using the CG reconstruction method.

ACKNOWLEDGMENTS

This work was supported by Marie Curie Intra-European Fellowship No. MEIF-CT-2003-501333. The authors also gratefully acknowledge all the fruitful discussions they had with K. M. Bograchev. Finally, they would like to thank the Boole Centre for Research in Informatics, University College Cork, Ireland, for the usage of their computer facility.

[1]S. J. Norton and M. Linzer, "Ultrasonic reflectivity imaging in three dimensions: Exact inverse scattering solutions for plane, cylindrical and spherical apertures," IEEE Trans. Biomed. Eng. **28**(2), 202–220 (1981).
[2]J. F. Greenleaf and R. C. Bahn, "Clinical imaging with transmissive ultrasonic computerized-tomography," IEEE Trans. Biomed. Eng. **28**(2), 177–185 (1981).
[3]A. J. Devaney, "A filtered back-propagation algorithm for diffraction tomography," Ultrason. Imaging **4**(4), 336–350 (1982).
[4]S. A. Johnson, D. A. Christensen, C. C. Johnson, J. F. Greenleaf, and B. Rajagopalan, "Nonintrusive measurement of microwave or ultrasound induced hyperthermia by acoustic temperature tomography," Proc.-IEEE Ultrason. Symp. 977–982 (1977).
[5]N. T. Sanghvi, F. J. Fry, R. Bihrle, R. S. Foster, M. H. Phillips, J. Syrus, A. V. Zaitsev, and C. W. Hennige, "Noninvasive surgery of prostate tissue by high-intensity focused ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 1099–1110 (1996).
[6]J. E. Kennedy, G. R. ter Haar, and D. Cranston, "High intensity focused ultrasound: Surgery of the future," Br. J. Radiol. **76**, 590–599 (2003).
[7]G. Ter Haar, "Ultrasound focal beam surgery," Ultrasound Med. Biol. **21**, 1089–1100 (1995).
[8]C. J. Diederich and K. Hynynen, "Ultrasound technology for hyperthermia," Ultrasound Med. Biol. **25**, 871–887 (1999).
[9]K. A. Wear, R. F. Wagner, M. F. Insana, and T. J. Hall, "Application of autoregressive spectral analysis to cepstral estimation of mean scatterer spacing," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 50–58 (1993).
[10]N. T. Sanghvi, R. S. Foster, F. J. Fry, R. Bihrle, C. Hennige, and L. V. Hennige, "Ultrasound intracavity system for imaging, therapy planning and treatment of focal disease," Proc.-IEEE Ultrason. Symp. **2**, 1249–1253 (1992).
[11]R. Seip, P. VanBaren, and E. S. Ebbini, "Dynamic focusing in ultrasound hyperthermia treatments using implantable hydrophone arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 706–713 (1994).
[12]R. Seip and E. S. Ebbini, "Noninvasive estimation of tissue temperature response to heating fields using diagnostic ultrasound," IEEE Trans. Biomed. Eng. **42**, 828–839 (1995).
[13]A. N. Amini, E. S. Ebbini, and T. T. Georgiou, "Noninvasive estimation of tissue temperature via high-resolution spectral analysis techniques," IEEE Trans. Biomed. Eng. **52**(2), 221–228 (2005).
[14]N. R. Miller, J. C. Bamber, and G. R. ter Haar, "Imaging of temperature-induced echo strain: Preliminary in vitro study to assess feasibility for guiding focused ultrasound surgery," Ultrasound Med. Biol. **30**(3), 345–356 (2004).
[15]R. Souchon, G. Bouchoux, E. Maciejko, C. Lafon, D. Cathignol, M. Bertrand, and J. Y. Chapelon, "Monitoring the formation of thermal lesions with heat-induced echo-strain imaging: A feasibility study," Ultrasound Med. Biol. **31**(2), 251–259 (2005).
[16]N. R. Miller, J. C. Bamber, and P. M. Meaney, "Fundamental limitations of noninvasive temperature imaging by means of ultrasound echo strain estimation," Ultrasound Med. Biol. **28**(10), 1319–1333 (2002).
[17]J.-W. Jeong, T.-S. Kim, D. C. Shin, S. Do, M. Singh, and V. Z. Marmarelis, "Soft tissue differentiation using multiband signatures of high resolution ultrasonic transmission tomography," IEEE Trans. Med. Imaging **24**(3), 399–408 (2005).
[18]Y. Nawata and K. Kaneko, "Measurement of temperature distribution in phantom body by an ultrasonic CT method," Proceedings of the ASME/JSME Joint Thermal Engineering Conference 1999, Vol. **3**, pp. 469–474.
[19]D. Kourtiche, M. Nadim, G. Kontaxakis, C. Marchal, and G. Prieur, "Temperature measurements using US tomography: Theoretical aspects," Proc. IEEE Eng. Med. Biol. Soc. **13**(1), 325–326 (1991).
[20]P. Lasaygues and J. P. Lefebvre, "Cancellous and cortical bone imaging by reflected tomography," Ultrason. Imaging **23**(1), 55–70 (2001).
[21]M. Muller, P. Moilanen, E. Bossy, P. Nicholson, V. Kilappa, T. Timonen, M. Talmant, S. Cheng, and P. Laugier, "Comparison of three ultrasonic axial transmission methods for bone assessment," Ultrasound Med. Biol. **31**(5), 633–642 (2005).
[22]T. Miyashita, "Low-frequency scattering component for quantitative circular-scanning ultrasonic diffraction tomography," Jpn. J. Appl. Phys., Part 1 **40**(5B), 3926–3930 (2001).
[23]T. D. Mast, "Empirical relationships between acoustic parameters in human soft tissues," ARLO **1**, 37–42 (2000), and references herein.
[24]A. J. Devaney and M. Dennison, "Inverse scattering in inhomogeneous background media," Inverse Probl. Eng. **19**(4), 855–870 (2003).
[25]A. T. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging* (IEEE, New York, 1988).
[26]K. M. Bograchev and W. M. D. Wright, "Computer modelling of iterative technique application for tissue thermal imaging," Proc.-IEEE Ultrason. Symp. **21**, 2038–2041 (2005).
[27]N. V. Ruiter, M. Zapf, R. Stotzka, T. O. Muller, K. SchloteHolubek, G. Göbel, and H. Gemmeke, "First images with a 3d prototype for ultrasound computer tomography," Proc.-IEEE Ultrason. Symp. **21**, 2042–2045 (2005).
[28]R. A. Crowther, D. J. Derosier, and A. Klug, "The reconstruction of a three-dimensional structure from projections and its application to electron microscopy," Proc. R. Soc. London, Ser. A **317**, 319–340 (1970).
[29]A. Abubakar, T. M. Habashy, P. M. van den Berg, and D. Gisolf, "The diagonalized contrast source approach: An inversion method beyond the Born approximation," Inverse Probl. Eng. **21**, 685–702 (2005).
[30]M. Dobroka, L. Dresen, C. Gelbke, and H. Ruter, "Tomographic inversion of normalized data-double-trace tomography algorithms," Geophys. Prospect. **40**(1), 1–14 (1992).
[31]J. A. Scales, "Tomographic inversion via the conjugate gradient method," Geophysics **52**(2), 179–185 (1987).
[32]M. M. Bronstein, A. M. Bronstein, M. Zibulevsky, and H. Azhari, "Reconstruction in diffraction ultrasound tomography using nonuniform FFT," IEEE Trans. Med. Imaging **21**, 1395–1401 (2002).
[33]A. J. Davies, *The Finite Element Method- A First Approach* (Clarendon, Oxford, 1980).
[34]A. Taflove, *Computational Electrodynamics: The Finite-Difference Time-Domain Method* (Artech House, Norwood, MA, 1995).
[35]J. W. Daniel, "The conjugate gradient method for linear and nonlinear operator equations," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **4**, 10–26 (1967).
[36]R. E. Kleinman and P. M. van den Berg, "Iterative methods for solving integral equations," *PIERS 5, Application of Conjugate Gradient Method to Electromagnetics and Signal Analysis*, edited by T. K. Sarkar (Elsevier, New York, 1991), pp. 67–102.
[37]P. M. van den Berg, "Iterative schemes based on minimization of a uniform error criterion," in Ref. 36, pp. 27–66.
[38]A. T. de Hoop, *Handbook of Radiation and Scattering of Waves: Acoustic Waves in Fluids, Elastic Waves in Solids, Electromagnetic Waves* (Academic, London, 1995).
[39]K. W. A. van Dongen, C. Brennan, and W. M. D. Wright, "A reduced forward operator for acoustic scattering problems," Proceedings of the IEE Irish Signals and Systems Conference, Dublin, Ireland, 1–2 September 2005, pp. 294–299.
[40]A. F. Peterson, S. L. Ray, and R. Mittra, *Computational Methods for Electromagnetics* (Wiley-IEEE, New York, 1998).

Coupling of earphones to human ears and to standard coupler

Dejan G. Ćirić[a] and Dorte Hammershøi
Department of Acoustics, Aalborg University, Fredrik Bajers Vej 7 B5, DK-9220 Aalborg Ø, Denmark

(Received 5 October 2005; revised 14 June 2006; accepted 6 July 2006)

A standardized acoustical coupler should enable the calibration of audiometric earphones which ensures that the hearing thresholds determined in the audiometric measurement are independent of the earphone type. This requires that the coupler approximates the average human ear closely. Nevertheless, the differences among earphones as well as between human ears and the coupler affect the results of the audiometric measurements inducing uncertainty. As the mentioned differences are related to the coupling of different earphones to human ears and to a standardized coupler, the effects of this coupling are investigated by measuring the transfer functions from input voltage of the earphone terminals to the pressure at the ear canal entrance in two situations: open and blocked canals. Since the "ear canal entrance" is not well-defined for the coupler, transfer function measurements in the coupler were carried out in a similar way but at different depths. In order to describe and compare the earphone couplings, the pressure divisions at the entrance of the ear canal are calculated from the measured transfer functions. The results indicate that significant difference appears among sound pressures generated in different individuals' ears. Also, the earphone couplings to human ears and to the coupler differ considerably. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2258929]

PACS number(s): 43.64.Ha, 43.58.Vb, 43.66.Yw [BLM] Pages: 2096–2107

I. INTRODUCTION

A. Background and previous investigations

In clinical settings, determination of hearing thresholds is commonly carried out by pure-tone audiometry using an earphone for the sound field generation. It is assumed that the earphone produces a certain sound pressure level in the ear when a specific voltage is fed to the input terminals independently or nearly independently of individual characteristics of the ear. However, the response of the earphone is affected by the acoustic load applied, that is, by the acoustic properties of the ear. Thus, the acoustic loading of the pinna, ear canal, eardrum, and ossicular chain are potential sources that influence the acoustic signal received by the eardrum.[1]

One of the important causes of unreliability of earphones in audiometry emerges from variability of the acoustic coupling between the earphone as a source and the ear as a termination.[2] Two distinct factors are related to the variability; both are frequency dependent. A factor causing the variability at low frequencies, and resulting in uncontrolled variations of the pressure, is the leakage. This leakage is different for different types of earphones. For supra-aural earphones (earphones which rest along the edge of the pinna), the coupling between the cushion and the pinna is usually not sealed effectively. Also, the coupling and thus the amount of leakage depends on the placement of the earphone. This unstable coupling leads to variable amounts of sound pressure loss usually below 500 Hz accompanied by small and variable amounts of sound pressure increase at somewhat higher frequencies.[3] On the other hand, at higher frequencies (above approximately 2000 Hz), the sound pressure depends on the wave properties of the earphone and the external ear.[2] In this frequency region, the size and shape of the cavity enclosed by the earphone, factors that are dependent on the earphone and its positioning, the geometry of the pinna, and the ear canal, become very important.

The variability of acoustical coupling between ear and earphone due to individual differences, characteristics of the earphone, or positioning of the earphone has been investigated in several studies.[4–8] For that purpose, the sound pressure level has been measured at different points in the ear, and the hearing thresholds have been determined. Extreme differences of the ear canal sound pressures produced in different ears as much as 35 dB have been reported for subjects with middle-ear pathologies present.[3,9] Large uncontrolled variations in the ear canal sound pressure lead to errors of the same magnitude in audiometric test. These errors could cause an incorrect interpretation of observed effects on hearing leading to inappropriate use of hearing aids or unnecessary surgery.

In the standard audiological testing procedures, even though the effective sound pressure level at the eardrum depends on individual subject characteristics and on the earphone itself, the stimulus level in the ear is not measured. The level of the sound stimulus is determined through a calibration procedure, where the earphone is loaded acoustically with a coupler. The input voltage to the earphone is adjusted in such a way that a given sound pressure level is obtained in the coupler.[10,11] The procedure for determining the acoustical output of earphones based on usage of either rigid couplers or so-called artificial ears has passed through different stages of development and standardization. Unfortunately, the results from previous investigations[5,6,12–16] have indicated that none of the couplers or artificial ears give completely satis-

[a]Electronic mail: dc@acoustics.aau.dk; Currently at Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia and Montenegro, dciric@elfak.ni.ac.yu

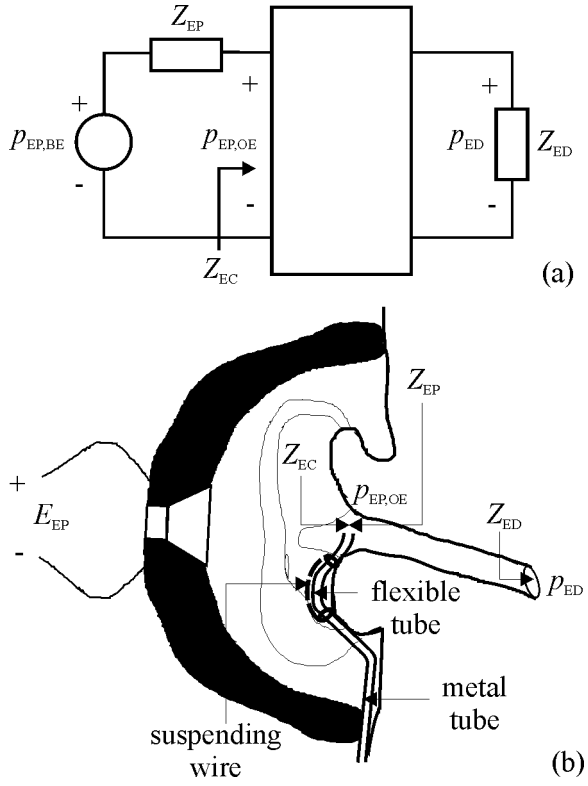


FIG. 1. Analog model for sound transmission in an ear for earphone exposure (a), and anatomic illustration together with the probe microphone tubes placed for measurements (b).

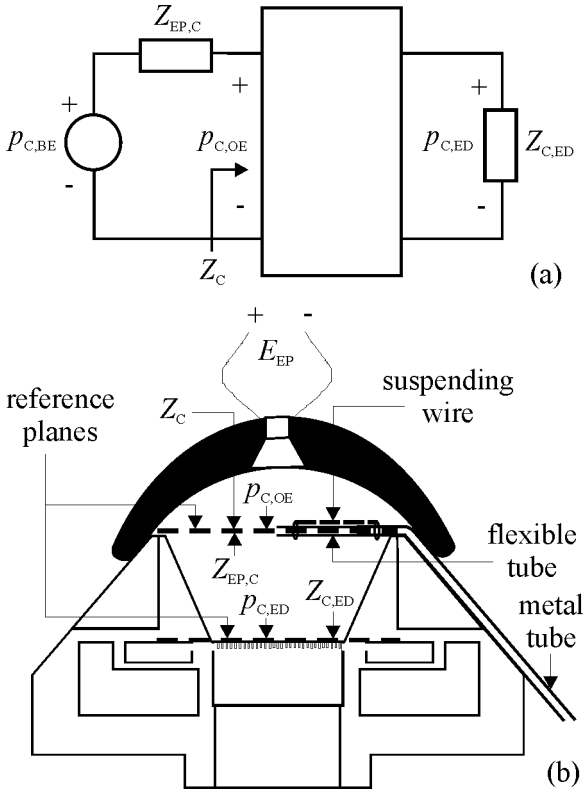


FIG. 2. Analog model for sound transmission in the coupler for earphone exposure (a), and illustration together with the probe microphone tubes placed for measurements (b).

factory results for accurate calibration of audiometric earphones. Important differences between the pressures generated in the subject ear canal and the coupler reported for different combinations of the earphones and the couplers[5,6,14,17–19] have shown that calibrating couplers have not duplicated the acoustical load of a human ear as precisely as desired for audiometric tests.

B. Models of sound transmission

The sound transmission in an ear for earphone exposure may be represented by the analog model given by Møller,[20] see Fig. 1. The sound pressure at the entrance of the ear canal is denoted $p_{EP,OE}$ (where EP is used for earphone and OE for open entrance), while the sound pressure at the eardrum is denoted $p_{EP,ED}$ (ED—eardrum). The ear canal (EC) is represented by an acoustical two-port terminated by the impedance of the eardrum Z_{ED}. The excitation part comprising everything outside the ear canal is modeled by a Thévenin equivalent with generator sound pressure $p_{EP,BE}$ (BE—blocked entrance) and its impedance Z_{EP}.

A similar model can be used for the free-field situation where sound is produced not by an earphone but by a source placed in the free field. The Thévenin equivalent for this exposure situation is given by the generator sound pressure $p_{FF,BE}$ (FF—free field) and its impedance Z_R (R—radiation), which represents the free air radiation impedance seen from the ear canal. The sound pressure at the entrance of the ear canal is $p_{FF,OE}$, and the pressure at the eardrum is $p_{FF,ED}$.

The sound transmission in the coupler for earphone exposure can be described by a model similar to the one for the sound transmission in the ear, see Fig. 2. The indices for the pressures and impedances contain the same abbreviations as in the previous model except for the character C, which is used to represent the coupler. Since the reference plane in the coupler that corresponds to the entrance of the ear canal is not well-defined, transfer functions are measured at different planes, i.e., different depths in the coupler. The reference planes on the top of the coupler's microphone and just above the orifice of the coupler are shown in Fig. 2(b). The termination impedance of the coupler including the impedance of the coupler's microphone is denoted $Z_{C,ED}$, where the term ED (eardrum) in the coupler relates to the plane of the coupler's microphone membrane.

The Thévenin pressures ($p_{EP,BE}$, $p_{FF,BE}$, and $p_{C,BE}$) do not exist physically during listening or calibration, but can be measured at the entrance of the blocked ear canal (reference plane in the coupler), e.g., blocked by the earplug (blocking tip in the coupler), since the acoustical two port is considered as an open circuit in such a condition.

For all three exposure situations, in order to yield the sound pressure at the entrance of the ear canal (reference plane), the Thévenin pressure is divided between the generator impedance Z_{EP} ($Z_{EP,C}$) and the ear canal impedance Z_{EC} (coupler impedance Z_C). Thus, the pressure division (PD) for earphone exposure in the ear is given as

$$\frac{p_{EP,OE}}{p_{EP,BE}} = \frac{Z_{EC}}{Z_{EC} + Z_{EP}}, \tag{1}$$

while the PD for free-field exposure is given as

$$\frac{p_{\mathrm{FF,OE}}}{p_{\mathrm{FF,BE}}} = \frac{Z_{\mathrm{EC}}}{Z_{\mathrm{EC}} + Z_R}. \quad (2)$$

Similar PD can be given for earphone exposure in the coupler. The ratio between two PDs for the sound transmission in the ear represents the pressure division ratio (PDR) given as

$$\frac{p_{\mathrm{FF,OE}}/p_{\mathrm{FF,BE}}}{p_{\mathrm{EP,OE}}/p_{\mathrm{EP,BE}}} = \frac{Z_{\mathrm{EC}} + Z_{\mathrm{EP}}}{Z_{\mathrm{EC}} + Z_R}. \quad (3)$$

By analyzing the PDs, valuable information about the characteristics of the impedances involved and their ratios can be obtained. In addition, the importance of all three impedances, the impedance of the ear canal, earphone, and radiation impedance, can be assessed using the PDR. In the case that the earphone impedance is close to the radiation impedance, which would probably be the lowest impedance coupled to the ear, the PDR becomes equal to unity. This is however also the case for some frequencies, even when earphone and radiation impedance differ, if the ear canal impedance is much larger than both. Earphones with unity PDR have been defined as earphones with free air equivalent coupling (FEC),[21] or "open" (in earlier Refs. 19, 20, and 22).

C. Purpose of investigation

In spite of the known variability in the acoustical coupling of earphones to the ear, the influence of the earphone fit to the ear is still present in the test of hearing sensitivity. Besides, the coupler calibration is indirect, in that the sound pressure measured does not represent the actual ear exposure during audiometry. This is why the influences of the differences between earphones as well as between human ears and the coupler are examined here by investigating the coupling, i.e., pressure divisions and pressure division ratios, of different earphones to human ears and the coupling of different earphones to the standardized coupler. The present study is preliminarily reported previously.[23,24]

The transfer functions from input voltage of the earphone terminals to the pressure at the entrance of the ear canal were measured in two situations: (1) open (earphone transfer function with open entrance $\mathrm{PTF_{EP,OE}}$) and (2) blocked (earphone transfer function with blocked entrance $\mathrm{PTF_{EP,BE}}$). In order to compare the results for human ears with the results for the coupler, similar measurements were done in the coupler. However, since it is not well-defined where the "ear canal entrance" is located in the coupler, the transfer functions ($\mathrm{PTF_{C,OE}}$ and $\mathrm{PTF_{C,BE}}$) were measured at different depths in the coupler. Based on the transfer functions, the PDs are determined in the way that will be explained later and used to describe the coupling.

In addition, the coupling of audiometric earphones to human ears is compared with the corresponding coupling of the ears to free air. The external ear represents an important part of an earphone-to-ear system, and acoustical measurements on that part alone can give some valuable information of the system as a whole. Thus, the transfer functions to open and blocked ear canal ($\mathrm{PTF_{FF,OE}}$ and $\mathrm{PTF_{FF,BE}}$) were also measured for the exposure to a free field. These functions are used for determination of the PD for free-field exposure, which, together with the PD for earphone exposure, gives the PDR.

II. MEASUREMENT METHOD

A. Method of measurements in human ears and coupler

1. Measurements in human ears

Transfer functions from voltage at the earphone terminals to the sound pressure at the entrance of the ear canal were measured; specifically, impulse responses were measured for the transmission from voltage at the input of the power amplifier to the output of the measuring microphone. At frequencies where the wavelength is comparable to or smaller than the dimensions of the ear, the sound pressure depends on the measurement position in the ear, on the individual characteristics of the ear, and in the case of earphone exposure, on the characteristics of the earphone itself. The center of the ear canal entrance is chosen as a suitable measurement point. It enables measurements at two conditions, with the ear canal blocked (by an Aero Ear Classic earplug) and with the ear canal open, without considerable discomfort for the subjects.

2. Measurements in the coupler

The ear simulator specified in IEC 60318-1 (Ref. 25) (Brüel & Kjær type 4153, here designated coupler) was used for all measurements carried out in the coupler. For the supra-aural earphones, the coupler was in its original form, while for the circum-aural earphone it was supplied with a flat plate adaptor (type 1) (Ref. 26) used as a rest for the earphone.

The transfer functions from input voltage of the earphone terminals to the sound pressure at the reference plane/depth in the coupler were measured similar to the measurements in human ears. In order to find the reference plane in the coupler that could best compare to the entrance of the ear canal, nine planes were defined for measurements beginning from the plane almost on the top of the coupler's microphone (Brüel & Kjær type 4134) shown in Fig. 2(b), designated plane 1. Next eight planes separated approximately one millimeter from each other were defined above this plane 1. The last one, plane 9, was just above the plane of the coupler's orifice, also shown in Fig. 2(b). For each of the reference planes, the central point of the circular opening (in the plane) of the coupler was defined as the measurement point, i.e., all measurements were along the axis of the circular opening.

B. Earphones

Three exposure situations were used during the measurements, earphone and free-field exposure in the ear, and earphone exposure in the coupler. For the earphone exposures, the sound was reproduced by each of five audiometric earphones: four supra-aural earphones (Telephonics TDH39, Telephonics TDH39 with noise capsules and cushions type ME70 (referred to as TDH39C), Beyerdynamics DT48, and Holmberg 95-01) and one circum-aural (Sennheiser

HDA200). For the free-field exposure, three 7 cm midrange units (Vifa M10MD-39) mounted in 15.5 cm diameter hard plastic balls[21] were placed in an anechoic chamber at the distance of 2 m from the subject, one in front of, one aside of, and one above the subject.

C. Subjects

All 34 subjects participating in the measurements had normal hearing that was tested by standard audiometry. There were 19 males and 15 females. The age ranged from 22 to 32 years (the mean age was 24.7). Their ears were not clinically controlled, but absence of physical irregularities such as perforated eardrum or presence of wax in the ear canal was examined by an otoscope. Also, none of the subjects had reported ear abnormalities that might affect the middle ear function. The subjects were not allowed to wear spectacles and earrings during the measurements, and their ears were not covered by hair.

D. Measurement system

The general purpose Maximum Length Sequence (MLS) measuring system (MLSSA – DRA Laboratories) was used. The autorange function in the MLSSA system was enabled to allow the best utilization of the dynamic range. The MLS signal of 12th order with 16 averages was chosen enabling sufficiently long excitation and still relatively short measurement (1.4 s). The sampling frequency was 48 kHz (provided by an external clock).

The stimulus amplitudes for the earphones were set in such a way to give a sound level of about 85 dB(A) at the microphone position in a head and torso simulator (Brüel & Kjær type 4128). This corresponds to a free-field sound pressure level of approximately 72 dB(A), which was the level of excitation for the free-field exposure. The signal from the MLSSA system was fed to the earphones or to the loudspeakers through the power amplifier Pioneer A-616 with a calibrated gain of 0 dB. In addition, the signal from the power amplifier was attenuated approximately 25 dB by a custom made attenuator, but only for the earphone exposure.

The choice of an appropriate microphone for the measurements in both conditions represented the compromise between a satisfying sensitivity and dimensions such that the sound field at the entrance of the ear canal was not disturbed. Thus, the sound pressure was recorded by a probe microphone Brüel & Kjær type 4182 with a 45 mm metal tube, which was bent at the end, and extended by a flexible tube. The length of the flexible tube varied from 9 to 15 mm depending on the shape of the pinna of a particular subject. The signal from the probe microphone was sent to the measuring amplifier Brüel & Kjær type 2607 and then to the MLSSA system.

The transfer functions in the coupler for all supra-aural earphones were measured with the same 45 mm metal tube bent at the end. The flexible tube attached to this metal tube was of a variable length from 5 to 10.5 mm depending on the earphone and the reference plane for measurement. For the circum-aural earphone, since the flat plate adaptor was placed on the coupler, it was impossible to use the same metal tube of the microphone, but another metal tube of the length of 100 mm bent in a specific way to follow the plate adaptor and coupler opening was attached to the probe microphone. The flexible tube of variable length, but similar to the above-mentioned lengths was then attached to this bent metal tube. The probe microphone was calibrated separately for the different metal tubes (and attached flexible tubes) for the supra-aural and the circum-aural earphones. Also, the transfer functions from input voltage of the earphone terminals to the pressure at the membrane of the coupler's microphone were measured by this microphone.

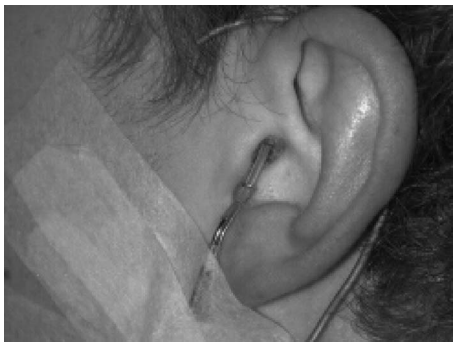


FIG. 3. Placement of probe microphone on human subject.

E. Measurement procedure

1. Procedure for measurements in ears

The subjects were seated on an elevated seat with a headrest placed approximately in the center of the anechoic chamber, so that their ears were 1.8 m above the wire mesh floor. The headrest was used for support to enable subjects to comfortably keep their heads fixed during the measurements. In order to minimize the risk of displacement of the probe tube tip, special care was taken to place and fix the microphone and tube tip securely, similar to Ref. 21. The bent part of the metal tube was positioned in the notch between the tragus and antitragus. The tip of the flexible tube was centered at the entrance to the ear canal, and its position was assured with a special suspending wire, see Figs. 1(b) and 3. A flexible metal strap, which was individually adjusted to fit the shape of the ear, was used for attaching the probe microphone to the subject's ear. The microphone arrangement was fixed by surgical tape, and additionally by a bandage and gauze with some subjects. The position of the probe tip was checked before and after each measurement.

The transfer function $PTF_{EP,BE}$ was measured first. For each earphone, the measurement was repeated five times for each of the two capsules (left and right) placing both of them on a subject's left ear only, removing and repositioning the earphone after each measurement. Care was taken to position the earphone in a way similar to its position during audiometric tests. After measurements with all five earphones, the $PTF_{FF,BE}$ was measured for sound coming from three directions. The measurement was repeated three times for each direction. Then, the earplug was carefully removed trying not to disturb the position of the probe tip. The next step was to repeat all the measurements, but measuring at the open ear canal entrance, $PTF_{FF,OE}$ first and then $PTF_{EP,OE}$. Some ad-

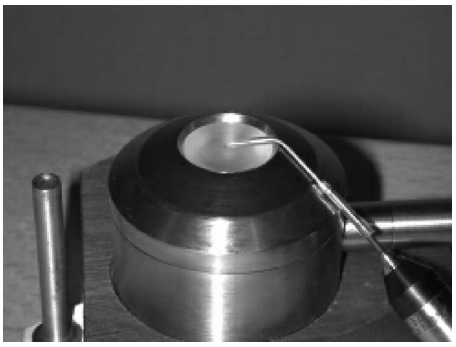


FIG. 4. Positioning of the tip of probe microphone tube for supra-aural earphones for measurements in blocked coupler.

ditional control measurements were included to check the repeatability of the measurement procedure and also to measure the signal-to-noise ratio.

Some difficulties were experienced due to disturbing effects of the physiological noise caused by pulse, and also by breathing, etc. The influence of this disturbance was especially important at low frequencies, where the signal-to-noise ratio was decreased by 10 dB or even more in some cases. Because of that, these effects were monitored during the measurements. In the case of significant disturbance, the setup with the probe microphone was readjusted taking special care to reduce the influence of the physiological noise by isolating the points of possible interference.

2. Procedure for measurements in the coupler

Measurements in the coupler (performed in an audiometry room) were also carried out in two conditions, open and blocked coupler. Thus, the $PTF_{C,BE}$ was measured first. For that purpose, since it was impossible to block the coupler in such a precise way with the earplug, the coupler was blocked with a blocking tip of conical shape fitting the shape of the coupler's cavity and made of plastic. Preliminary measurements showed that the earplug and the blocking tip had similar blocking properties. Different blocking tips were used for different measurement planes. As shown in Fig. 4, the tip of the probe microphone tube was positioned at the center point on the top surface of the blocking tip, where this surface represents the particular measurement plane.

The body of the probe microphone was firmly fixed by a vice. For measurements with the supra-aural earphones, the probe microphone was positioned in such a way that the metal tube sat on the outer surface of the coupler's ring, Fig. 4. For measurements with the circum-aural earphone, the metal tube sat on the flat plate adaptor and the upper part of the coupler's ring. For the blocked condition, the transfer function measurement was repeated three times for each of the capsules, repositioning the earphone on the coupler each time. The force that should be applied on an earphone during calibration[10] was applied during coupler measurements for both conditions.

After the measurements were completed with the blocked coupler, the blocking tip was removed keeping the measurement setup in the same position in reference to the coupler. Again, the transfer function was measured three times for each of the capsules repositioning the earphone between measurements. The whole measurement procedure for the blocked and open coupler was repeated for each of the measurement planes.

For one measurement plane, 12 transfer functions were measured for each of the earphones (six for each of the conditions—open and blocked coupler) giving 60 transfer functions in total for each measurement plane. Additional measurements were carried out in order to check the repeatability and signal-to-noise ratio.

F. Data processing

In order to determine the transfer functions, PDs and PDRs [Eqs. (1) to (3)], the measured impulse responses must be post processed, which was done in MATLAB. The impulse responses measured in human ears and in the coupler contain the desired information, i.e., the transfer function from input voltage of the earphone terminals to the sound pressure at the measurement point (H_{TF}). The impulse responses also contain the response of the equipment (the electrical part of the measurement system, the measurement microphone, and the loudspeakers for free-field exposure).

For the determination of PDs, the contribution of the measurement system cancels out, since the same system was used for the measurements in both conditions, open and blocked. Thus, the PDs were simply determined by Fourier transformation of the measured impulse responses for open and blocked conditions, and complex division in frequency domain. Since all responses were relatively short (a few milliseconds), only the first 256 samples were used for the post processing.

The determination of transfer functions requires a calibration between voltage and sound pressure, which takes the contribution of the measurement system into account. The transfer function of the electrical part of the measurement system (H_{EL}) was obtained by measurement of the electrical part of the system while its output was short circuited. The transfer function of the probe microphone (H_{MIC}) was found with reference to a reference microphone (Brüel & Kjær type 4136) in a dedicated coupler as described in the manual for the probe microphone. So, the transfer function for both conditions could be determined as

$$|H_{TF}| = \frac{|H_{IR}|}{|H_{EL}||H_{MIC}|}\frac{k}{\rho_{MIC}}, \tag{4}$$

where H_{IR} represents the frequency response obtained by Fourier transformation of the measured impulse response, ρ_{MIC} the microphone sensitivity, and k the correction factor (equal to the ratio of the output voltage of the microphone and the input voltage to the earphone) used to derive the correct absolute physical values.

III. RESULTS

A. Signal-to-noise ratio

The signal-to-noise ratio was not the same in all the measurements, since it depends on the physiological noise of a given subject, and on earphone type. Typically, the signal-

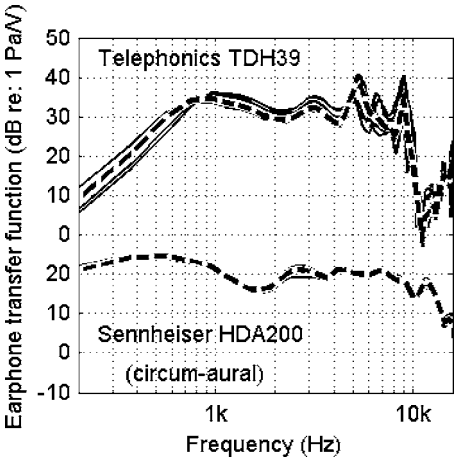


FIG. 5. Transfer functions (PTF_{BE}s) for one subject (JG) and five placements of the supra-aural (Telephonics TDH39) and circum-aural (Sennheiser HDA200) earphone (—) for the left capsule together with the average curve (- - -).

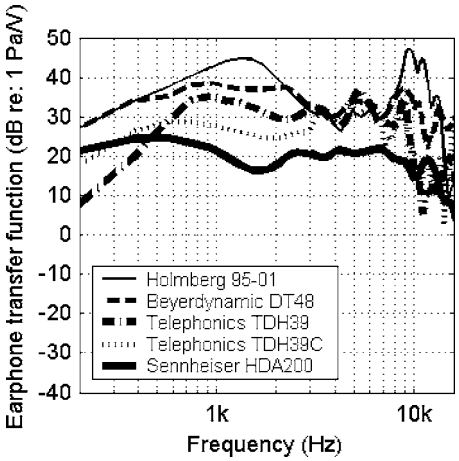


FIG. 6. Mean transfer functions (PTF_{BE}s) obtained by averaging transfer functions for left and right capsules of each earphone measured on one subject (JG).

to-noise ratio for measurements in human ears was about 40 or 45 dB, but in the worst case, in the frequency range where the earphone transfer function had a low amplitude and/or where the background noise level was high, the ratio was about 20 dB. For measurements in the coupler, the ratio was higher, since there was no physiological noise.

B. Earphone transfer functions in human ears

The PTF_{BE}s of one of the supra-aural earphones (Telephonics TDH39) together with those of the circum-aural earphone (Sennheiser HDA200) for one randomly chosen subject and one capsule obtained in five repeated measurements with the earphone repositioning are presented in Fig. 5. The variation in the earphone's response caused by its positioning on an ear depends on subject, earphone, and frequency. Thus, greater differences among the transfer functions exist at low and high frequencies, while the responses are closer to each other at midfrequencies. Besides, the differences are greater for the supra-aural earphones (they are in the order of approximately 5 to 8 dB) than for the circum-aural earphone (which are in the order of approximately 2 to 3 dB).

The earphone transfer functions from five repeated measurements are averaged. The difference between the average transfer functions for the left and right capsules of a particular earphone is relatively small in comparison to other differences in the results. The average transfer functions for both capsules are further averaged to give the mean transfer function for that measurement condition and earphone. The mean PTF_{BE}s of all earphones measured on one subject are presented in Fig. 6. These are rather different, although the reduction of PTF_{BE}s at low frequencies is common for supra-aural earphones.

Individual transfer functions, computed as the average of the five repeated measurements for each earphone (capsule) (PTF_{BE}s), are shown in Fig. 7 only for left capsules. These are considerably different, but there is a common, characteristic pattern for each earphone. The differences among individual PTF_{BE}s depend on frequency range and earphone. The largest variation caused by individual (sub-

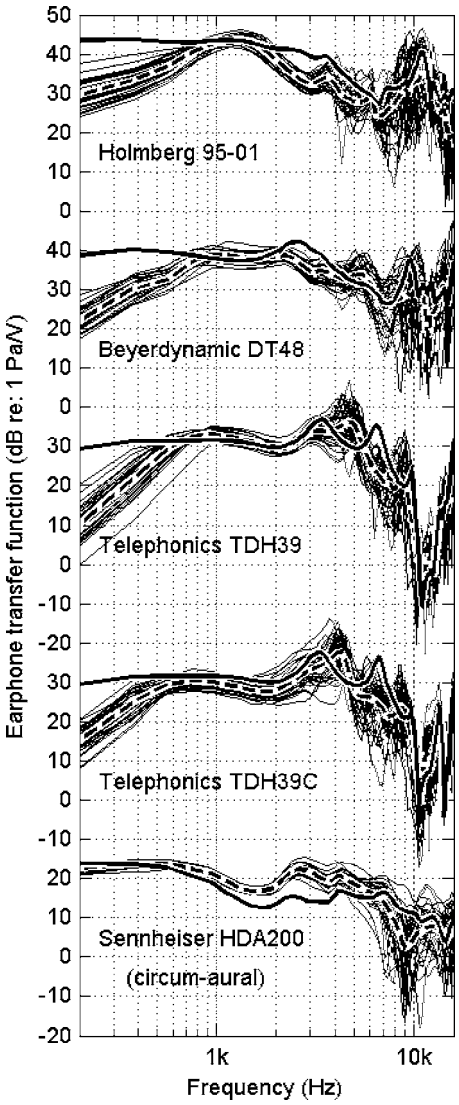


FIG. 7. Individual transfer functions (PTF_{BE}s) for all subjects and left capsule of each earphone (thin solid lines) and the average of these individual functions for each earphone (dashed lines) together with the earphone transfer function measured in the coupler by coupler's microphone (thick solid line) for each earphone.

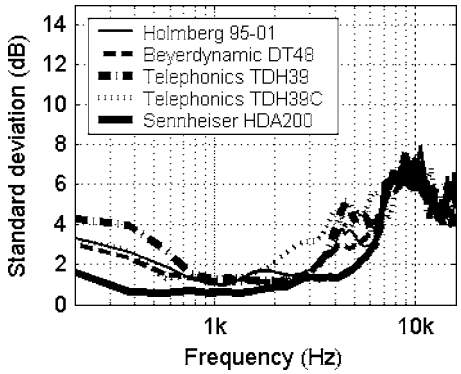


FIG. 8. Standard deviations of individual PTF_{BE}s calculated across all subjects and both capsules of each of the earphones.

ject's ear) characteristics for a particular supra-aural earphone exists at low and high frequencies, while it is somewhat smaller at midfrequencies. This is not completely the case for circum-aural earphone (Sennheiser HDA200), where the larger variation is found only at high frequencies. Besides, the variation is generally smaller for the circum-aural earphone than for the supra-aural earphones.

This can also be seen from the standard deviation calculated across all individual PTF_{BE}s for both capsules and for each earphone, presented in Fig. 8. The smallest standard deviations exist at midfrequencies between 1 and 3 kHz, while the greatest values of deviation are obtained at frequencies above 7 kHz. Besides, at low and midfrequencies, a smaller standard deviation is obtained for the circum-aural earphone than for the supra-aural earphones.

The transfer function for each of the earphones measured in the coupler by its microphone is also shown in Fig. 7. There is a significant difference between individual transfer functions measured in the ears and that one measured in the coupler, but also between the average human PTF_{BE}s and the function in the coupler. The difference between the average human PTF_{OE} and the coupler transfer function, presented in Fig. 9, is even greater for all earphones.

The intersubject variation in the earphone transfer functions is larger for the open ear canal entrance (in PTF_{OE}s).

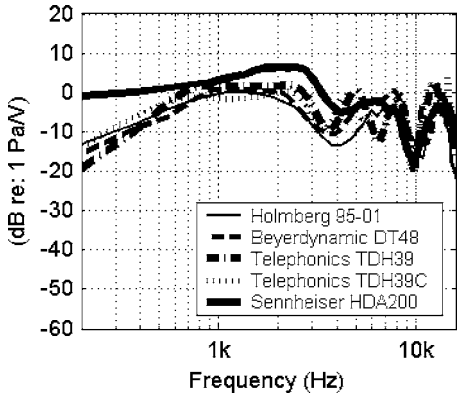


FIG. 9. Differences between average PTF_{OE}s measured in human ears for open ear canal entrance and transfer functions measured in the coupler by coupler's microphone.

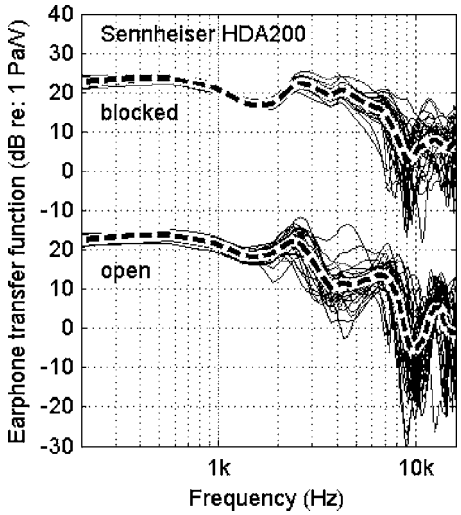


FIG. 10. Individual transfer functions (measured at both blocked and open ear canal entrances) for all subjects and left capsule of circum-aural earphone (thin solid lines) and the average of these individual functions (dashed lines).

For illustration, the individual transfer functions measured at the blocked and open ear canal entrance for only one earphone are shown in Fig. 10.

C. Earphone transfer functions in the coupler

The influence of earphone repositioning is not equally important for measurements in the coupler as for measurements in human ears. However, for the supra-aural earphones, the repositioning can result in differences in the transfer functions especially at low (below 800 Hz) and at high frequencies (above 8 kHz) in some cases. Repositioning of the circum-aural earphone causes very little difference in the transfer functions measured in the coupler.

The measurement position (depth) in the coupler affects the earphone transfer functions measured in both blocked and open coupler. The transfer functions measured by the probe microphone at eight depths (from the measurement plane 1—coupler's microphone surface, to the plane 8—just below the coupler's orifice) in the blocked coupler are presented in Fig. 11. For the supra-aural earphones, the shortest blocking tip (the measurement plane 1) gives the lowest amplitude of the transfer function at frequencies up to 5 or even 8 kHz depending on earphone. The amplitude of the transfer function in the mentioned frequency range is higher for measurement points closer to the coupler's orifice, where the volume of the coupler is reduced (by the bigger blocking tips). The pattern is not completely regular for all measurement depths and earphones mostly due to leakage, which may differ from measurement to measurement. This tendency is not prominent for the circum-aural earphone (Sennheiser HDA200), where the differences are relatively small below 5 kHz. This could be due to the bigger volume between the coupler and the cushion of the circum-aural earphone versus the corresponding volume with a supra-aural earphone. Thus a change of the volume by the blocking tip has less relative importance for the circum-aural earphone.

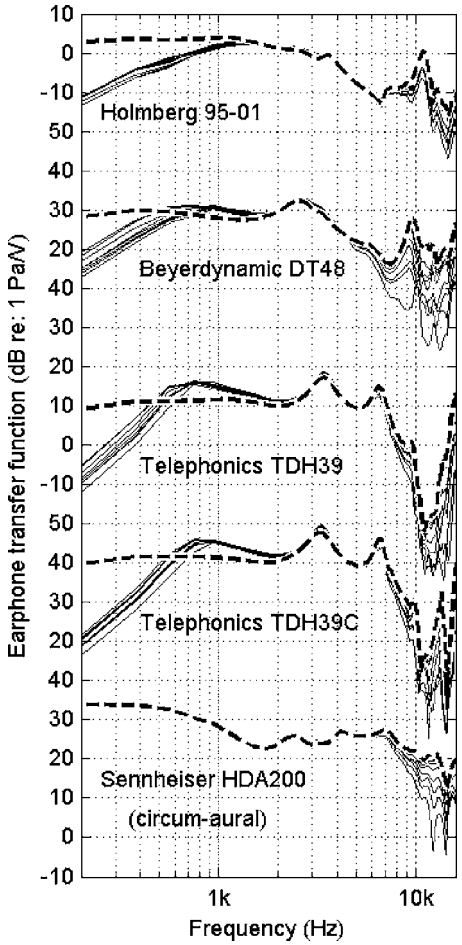


FIG. 11. Earphone transfer functions measured by probe microphone at eight depths in blocked coupler (thin solid lines), from measurement plane 1 (coupler's microphone surface) to plane 8 (just below the coupler's orifice), together with the function measured by coupler's microphone (thick dashed line) for all earphones.

The pattern at higher frequencies (above 5 or 8 kHz) is opposite to that at low and midfrequencies. Thus, the shortest tip yields the highest amplitudes of the transfer functions here. The pattern is not as regular for most of the supra-aural earphones, as for the circum-aural earphone.

The earphone transfer functions measured at different depths in the open coupler are shown in Fig. 12 for two earphones. For the supra-aural earphone (Telephonics TDH39), there are differences between the transfer functions at low frequencies. These are, however, not considered to relate to the different measurement positions (depths) but to leakage, which may vary slightly from measurement to measurement. For the circum-aural earphone (Sennheiser HDA200), where there is almost no leakage, the transfer functions for different depths coincide. At low and midfrequencies, the wavelength of sound is great compared to the dimensions of the coupler, and the sound pressure is uniform inside the coupler.

The dependence on measurement position (depth) is prominent above 5–7 kHz depending on earphone. The highest similarity to the function measured by the coupler's microphone is for the measurement position closest to that

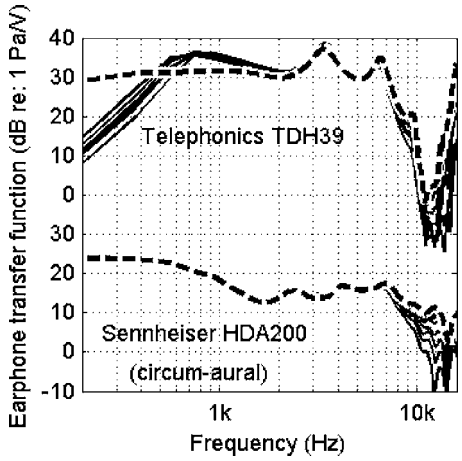


FIG. 12. Earphone transfer functions measured by probe microphone at eight depths in open coupler (from measurement plane 1 to 8) (thin solid lines) together with the function measured by coupler's microphone (thick dashed line) for two earphones.

microphone. The difference at high frequencies increases for positions closer to the coupler's orifice. This is similar to the pattern found for the blocked coupler.

D. Pressure division in human ears

The individual pressure divisions (PDs), obtained by complex division of the average earphone transfer functions PTF_{OE} and PTF_{BE} for each earphone and each subject, and average PDs taken across all subjects, are presented in Fig. 13, and compared to the PDs for free-field exposure. A very similar pattern exists for all subjects and each earphone. Some differences appear for some subjects, especially at the first peak, which is almost negligible in some cases. The patterns for different earphones are also similar, also compared to free-field PD. The PD is equal to or very close to 0 dB at low frequencies. There is a peak between 2 and 3 kHz with an average value of a few decibels. The next two dips and a peak are also common for the pattern for the PD, where the first dip appears around 4 kHz and the second one around 10 kHz, while the peak appears between 7 and 8 kHz.

The standard deviation for PDs is—in comparison to the deviation for PTF_{BE}s—somewhat smaller at lower frequencies and somewhat greater above approximately 3 kHz for most of the earphones.

E. Pressure division ratio

The pressure division ratios (PDRs) (ratio between the PDs for free-field and earphone exposure) are determined only for human ears, since measurements with free-field exposure were not made with the coupler. The individual PDRs presented in Fig. 14 for all earphones and subjects show that intersubject variation exists in the PDRs too. Therefore, although the average PDR taken across all subjects for each earphone is close to 0 dB (with some small fluctuations around this value above 1 or 2 kHz), individual PDRs can deviate from this zero value considerably. At frequencies up to 2 or 3 kHz depending on earphone, this deviation is relatively small (in the order of few decibels). However, at

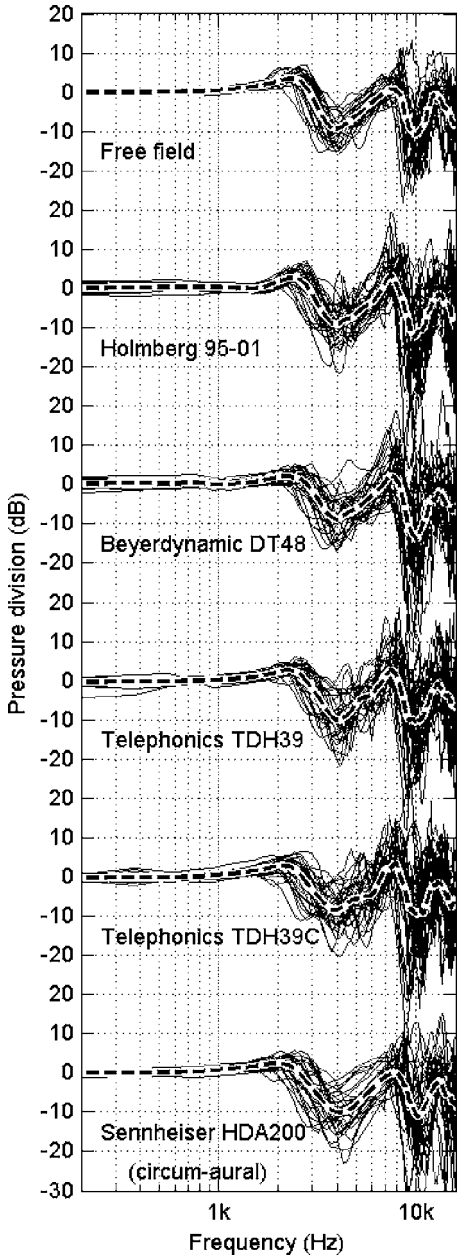


FIG. 13. Individual PDs for all subjects (—) and corresponding average of these individual PDs (- - -) for both free-field and earphone (left capsule) exposures.

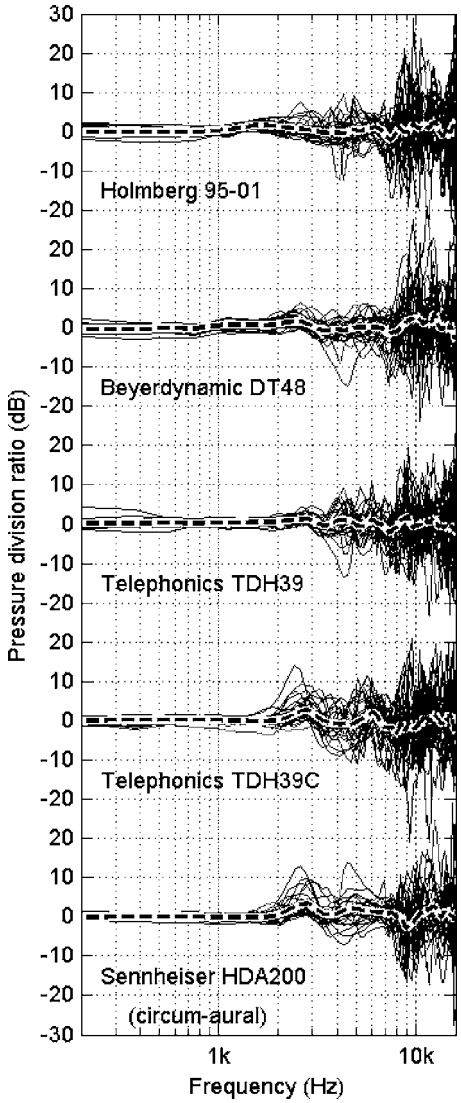


FIG. 14. Individual PDRs for all subjects (—) and corresponding average of these individual PDRs (- - -) for the left capsule of all earphones.

higher frequencies, it has significant values that increase with an increase in frequency. An individual PDR can have the values of even ±20 dB or even more.

The standard deviation for PDRs is similar to that for PDs, except that the values above 2 kHz are somewhat smaller here. Also, there is no substantial difference in the standard deviation for PDRs among the earphones, as, e.g., in the deviation for PTF_{BE}s.

F. Pressure division in the coupler

The differences between the earphone transfer functions measured in the open and blocked coupler are not the same as those for human ears. This is reflected in the PDs for different measurement positions (depths) in the coupler presented in Fig. 15. The pattern seen in PDs for human ears is not well resembled by the coupler PDs at any depth. Different depths affect the PD so that the closer the coupler orifice, the lower the amplitude. This is especially apparent at mid- and high frequencies for the supra-aural earphones, and at high frequencies for the circum-aural earphone (Sennheiser HDA200).

IV. DISCUSSION

A. Assessment of measurements in the coupler

Since the earphone transfer functions in the open coupler were measured at approximately the same measurement plane (plane 1) using both the coupler's microphone and the probe microphone, the possible influence of the probe microphone can be determined.

The transfer functions for the supra-aural earphones measured by the probe microphone have smaller values at low frequencies than the corresponding ones measured by the coupler's microphone [an example is given for Telephonics TDH39 in Fig. 16(a)]. For some earphones, there is also a small bump, i.e., an increase in amplitude. These differ-

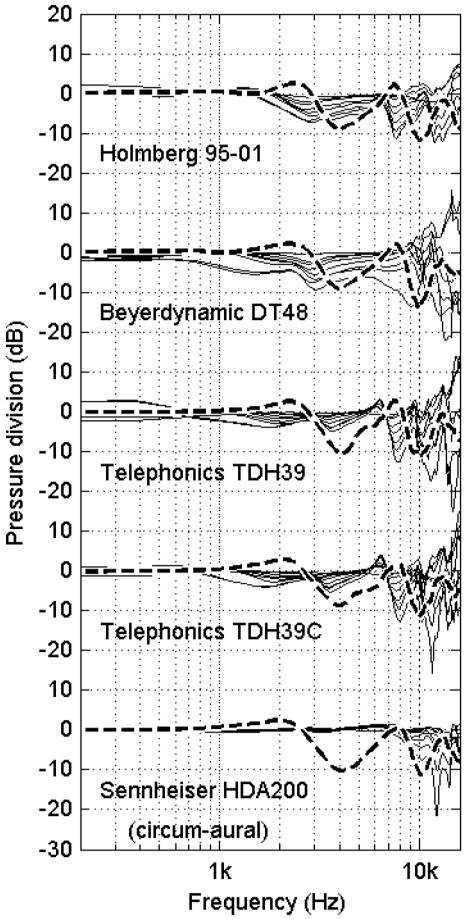


FIG. 15. PDs for eight depths in the coupler (thin solid lines), from measurement plane 1 to plane 8, together with mean PDs for human ears calculated across all subjects and both capsules of each earphone (thick dashed line).

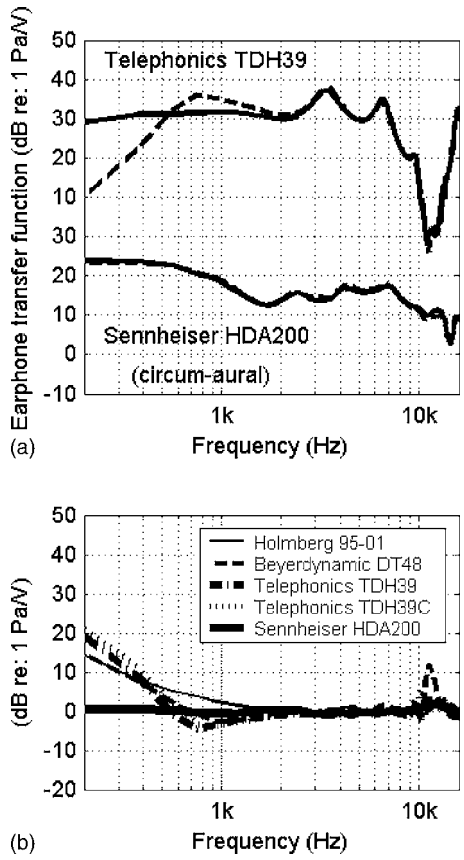


FIG. 16. Earphone transfer functions measured in the coupler by coupler's microphone (—) and mean transfer functions (taken across three placements of both capsules) measured at plane 1 in the coupler by probe microphone (- - -) (a), and the difference between transfer functions measured by coupler's and probe microphone (b).

ences can be seen in Fig. 16(b), and are presumably caused by the metal tube of the probe microphone placed at the edge of the coupler's orifice. This tube impedes perfect contact of earphone and coupler, and may thus introduce leakage. Small difference can also appear at higher frequencies for some supra-aural earphones. For the Beyerdynamic DT48 earphone (only) there is a prominent peak of the difference somewhat above 10 kHz. Nevertheless, generally speaking, the transfer functions coincide well at mid- and high frequencies.

The difference between the transfer functions for the circum-aural earphone (Sennheiser HDA200) is less than 1 dB up to 10 kHz, and less than 2 dB above that frequency, Fig. 16. The probe microphone tube apparently does not introduce any leakage for this earphone.

The possible influence of the probe microphone tube on the transfer function is approximately the same in the two measurement conditions, blocked and open coupler, since the position of the tube was the same. Thus, it is assumed that the PDs for the coupler are reliably determined. The tube does, however, influence the Thévenin impedance, and introduce in principle an error. Since leakage exists also when supra-aural earphones are placed on human ears, the error will—if anything—most likely systematically influence the measured PDs for the coupler to be more comparable to the PDs for human ears. Thus, the true dissimilarity can, in principle, be slightly worse.

B. Variations in earphone transfer functions

The main contributor of the intrasubject variation of PTF_{BE}s and PTF_{OE}s at low frequencies—caused by repositioning of the earphone on the human ear—is the leakage, which is more prominent for the supra-aural earphones than for the circum-aural. Because of the inaccurate fit of the earphone cushion to the complex geometry of the ear, the leakage differs from one trial to another. The transducer could also be differently inclined from measurement to measurement, which would cause differences in transfer functions in wider frequency ranges, including higher frequencies.

The intrasubject variations obtained are in the range of already presented variations for the audiometric earphones, but also for other headphones, though the variations differ somewhat between studies. Larger standard deviations (of the order of about 9 dB) at some frequencies above 9 kHz, and considerably smaller deviations (approximately up to 2 dB) below 8 or 9 kHz have been reported for headphone-to-ear-canal-transfer-functions associated with 20 placements of a headphone Sennheiser HD520 on an acoustic manikin,[7] that is, on both human ears and a manikin.[4] Somewhat dif-

ferent deviations have been reported in some other publications, e.g., standard deviation below 3 dB at frequencies in the range from 1 to 12 kHz for a circum-aural headphone Sennheiser Linear 250 I, but larger deviation for a supra-aural headphone Realistic Nova 17, especially at higher frequencies.[8]

The presented results reveal that the intersubject variation of earphone transfer functions is generally considerably larger than the intrasubject variation.

The intersubject variation found for all earphones stresses the differences of the sound pressure level in the ears of different subjects, when fed a given, calibrated input voltage to the earphone. The difference between mean sound pressure level (associated with an average human ear) and the level for a particular individual could be in the order of 5 or 7 dB at lower frequencies and in the order of 20 dB at higher frequencies for the supra-aural earphones. At the same time, the difference between the sound pressure levels in the ears of two individuals for the same input voltage fed to a particular earphone could be even greater than 35 dB at higher frequencies in extreme cases. The difference is somewhat smaller for the circum-aural earphone, but can still be fairly large. This individual difference is of direct consequence for the accuracy of the hearing thresholds determined.

Earphone transfer functions measured in human ears have been compared to the corresponding functions measured in different couplers in several studies.[5,6,13,14,17,18] In all of these studies, it has been shown that these functions are different, especially at low (for some earphones) and high frequencies (for almost all earphones). Usually, the level reduction caused by leakage seen in the functions measured in human ears for the supra-aural earphones does not exist in the transfer functions measured in the coupler, which is confirmed in the present investigation. The best agreement between the transfer functions found here is at midfrequencies, Fig. 9. Concerning the circum-aural earphone, the difference between the functions measured in human ears and in the coupler is within a similar range of values as the differences for the supra-aural earphones, but the trends are different for this earphone. There is almost no difference at low frequencies since the leakage does rarely exist for circum-aural earphones.

C. Pressure divisions

The fact that the mean PDs in human ears (calculated across all subjects) for all earphones investigated are almost equal to 0 dB at low frequencies (below 1 kHz) indicates that the Thévenin impedance of each of the earphones Z_{EP} is considerably smaller here than the impedance of the ear canal Z_{EC} according to Eq. (1). For such a ratio of impedances, an earphone could be considered as a nearly ideal sound pressure source. The influence of individual differences is significantly reduced for such a source so that similar sound pressures could be obtained in the ears of different individuals.

At frequencies above 1 kHz, the mean PD fluctuates showing some prominent peaks and dips. In that region, the relation between the impedances of the earphone and ear canal is complex. Since the mean PDs for all earphones agree well, then the ratios of the earphone impedances and the impedance of the average human ear canal are very similar for all earphones investigated. These earphones could be considered to behave very similar in respect to the mentioned ratio of impedances. On the other hand, the similarity among the PDs for different earphones is greater than the one among the PTF_{OE}s and among the PTF_{BE}s on both mean and individual basis. Thus, these earphones are more similar to each other in respect to the ratio of the earphone and ear canal impedance than to the sound pressures generated in the ear.

Since the mean PDs for both free-field and earphone exposure together with the mean PDRs for all earphones are close to 0 dB at low frequencies (below 1 or 2 kHz), according to Eqs. (1) to (3) it seems that the Thévenin impedances for both exposures (Z_R and Z_{EP}) are considerably smaller than the impedance of the ear canal (Z_{EC}). The deviations of the PDRs from zero line above 1 or 2 kHz could indicate a greater influence of the Thévenin impedances at those frequencies.

The general patterns of the PDs and PDRs compare well with PDs and PDRs for other commercially available headphones investigated in Ref. 21, although there are some distinct patterns for some headphones.

When the PDs for human ears are compared to the PDs for the coupler, it is difficult to find any similarity between them. This indicates that the couplings in two situations differ, with a possible lack of control of the ear stimulus level as a consequence. The difference is especially important at frequencies around the ear canal resonance, where the first peak appears in the PDs for human ears. In this frequency range noise induced hearing loss typically occurs, why accurate calibration and threshold determination is obviously essential.

V. CONCLUSION

The intrasubject and intersubject variations of the earphone transfer functions confirm that the sound pressure level generated in an ear of a particular subject by the audiometric earphone could differ considerably from the defined value. In the extreme case, this difference could be in the order of 15 dB or even more. Besides, the earphone transfer functions measured in human ears differ from the corresponding ones measured in the coupler, especially when individual functions are analyzed.

The similarity of the mean PDs for all earphones indicates that these earphones could be considered to behave very similar on a mean basis in respect to the ratio of the earphone impedance and the impedance of the average human ear canal.

According to the mean PDs and PDRs, it seems that the Thévenin impedances for free-field and earphone exposure (the radiation and earphone impedance) are considerably smaller than the impedance of the ear canal at low frequencies. Also, it could be assumed that these impedances behave

similar on a mean basis when added to the ear canal impedance, especially at some frequencies where the PDR is very close to the zero line.

The differences between human ears and the standardized coupler are further emphasized, when considering the PD as relevant quantity for comparison. Even some similarities that exist in earphone transfer functions measured in ears and the coupler disappear in PDs. This indicates that some important effects of coupling of audiometric earphones to human ears do not exist in coupling of the earphones to the coupler. The influence of these differences between the ears and the coupler as well as their consequences on the results of audiometric tests are yet to be determined.

ACKNOWLEDGEMENTS

Financial support from the Danish Technical Research Council and Aalborg University is greatly acknowledged. The authors would like to thank H. Møller for initiating ideas and suggestions, and K. Meesawat for his invaluable help with the practical work. Also, the authors would like to thank all our subjects for participation with the earplugs and microphones in their ears. Special thanks to Dr. Steven Colburn and the anonymous reviewer for their constructive reviews and many helpful comments.

[1]J. S. Russoti, T. P. Santoro, and G. B. Haskell, "Proposed technique for earphone calibration," J. Audio Eng. Soc., **36**(9), 643–650 (1988).
[2]J. Zwislocki, B. Kruger, J. D. Miller, A. F. Niemoeller, E. Shaw, and G. Studebaker, "Earphones in audiometry," J. Acoust. Soc. Am. **83**(4), 1688–1689 (1988).
[3]S. E. Voss, J. J. Rosowski, C. A. Shera, and W. T. Peake, "Acoustic mechanisms that determine the ear-canal sound pressure generated by earphones," J. Acoust. Soc. Am. **107**(3), 1548–1565 (2000).
[4]K. I. McAnally and R. L. Martin, "Variability in the headphone-to-ear-canal transfer function," J. Audio Eng. Soc. **50**(4), 263–266 (2002).
[5]H. Takeshima, T. Hiraoka, Y. Suzuki, M. Kumagai, and T. Sone, "Reference equivalent threshold sound pressure levels for new earphones," *Proc. ICA 95*, Trondheim, Norway, Vol. III, (15th International Congress of Acoustics, Trondheim, Norway, June 26–30, 1995), pp. 297–301.
[6]M. D. Burkhard and E. R. L. Corliss, "The response of earphones in ears and couplers," J. Acoust. Soc. Am. **26**(5), 679–685 (1954).
[7]A. Kulkarni and H. S. Colburn, "Variability in the characterization of the headphone transfer-function," J. Acoust. Soc. Am. **107**(2), 1071–1074 (2000).
[8]D. Pralong and S. Carlile, "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space," J. Acoust. Soc. Am. **100**(6), 3785–3793 (1996).
[9]S. E. Voss, J. J. Rosowski, S. N. Merchant, A. R. Thorton, C. A. Shera, and W. T. Peake, "Middle ear pathology can affect the ear-canal sound pressure generated by audiological earphones," Ear Hear. **12**(4), 265–274 (2000).
[10]ISO 389-1: Acoustics – Reference zero for the calibration of audiometric equipment – Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones. International standard ISO 389-1, International Standards Organization, 1998.
[11]ISO/TR 389-8: Acoustics – Reference zero for the calibration of audiometric equipment – Part 8: Reference equivalent threshold sound pressure levels for pure tones and circum-aural earphones. Draft international standard ISO/TR 389-8, International Standards Organization, 2001.
[12]M. E. Delany, L. S. Whitlle, J. P. Cook, and V. Scott, "Performance studies on a new artificial ear," Acustica **18**, 231–237 (1967).
[13]P. V. Brüel and E. Frederiksen, "Artificial ears for the calibration of earphones of the external type-part I," Technical Review B & K. **4**, 1–27 (1961).
[14]R. M. Cox, "NBS-9A coupler-to-eardrum transformation: TDH-39 and TDH-49 earphones," J. Acoust. Soc. Am. **79**(1), 120–123 (1986).
[15]K. K. Sharan, J. R. Cox, and A. Niemoeller, "Evaluation of new couplers for circumaural earphones," J. Acoust. Soc. Am. **38**, 945–955 (1965).
[16]A. H. Ithell, E. G. T. Johnson, and R. F. Yates, "The acoustical impedance of human ears and a new artificial ear," Acustica **15**, 109–116 (1965).
[17]N. P. Erber, "Variables that influence sound pressures generated in the ear canal by an audiometric earphone," J. Acoust. Soc. Am. **44**, 555–562 (1968).
[18]E. A. G. Shaw, "Ear canal pressure generated by circumaural and supraaural earphones," J. Acoust. Soc. Am. **39**, 471–479 (1966).
[19]T. Hirahara and K. Ueda, "Investigation of headphones suitable for psychophysical experiments," 199th ASA meeting, Penn State University, USA 1990, Abstract in J. Acoust. Soc. Am. **87**(S1), S142 (1990).
[20]H. Møller, "Fundamentals of binaural technology," Appl. Acoust., **36**(3/4), 171–218 (1992).
[21]H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, "Transfer characteristics of headphones measured on human ears," J. Audio Eng. Soc. **43**(4), 203–217 (1995).
[22]M. Vörlander, "Acoustic load on the ear caused by headphones," J. Acoust. Soc. Am. **107**(4), 2082–2088 (2000).
[23]D. Ćirić, D. Hammershøi, and B. L. Karlsen, "Coupling of earphones to human ear," Proc. 5th European Conf. on Noise Control, Euronoise 2003 (on the CD-ROM), Naples, Italy, 2003, Abstract in Acta Acustica united with Acustica **89**(Suppl. 1), s74 (2003).
[24]D. Ćirić and D. Hammershøi, "Coupling of earphones to human ear and to coupler," 147th Meeting of Acoust. Soc. Am., New York, USA, 2004, Abstract in J. Acoust. Soc. Am. **115**(5), 2499 (2004), paper 2pPP1.
[25]IEC: Electroacoustics – Simulators of human head and ear – Part 1: Ear simulator for the calibration of supra-aural earphones. International standard IEC 60318-1, International Electrotechnical Commission, 1998.
[26]IEC: Electroacoustics – Simulators of human head and ear – Part 2: An interim acoustic coupler for the calibration of audiometric earphones in the extended high-frequency range. International standard IEC 60318-2, International Electrotechnical Commission, 1998.
[27]D. Hammershøi and H. Møller, "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**(1), 408–427 (1996).

Mechanisms of generation of the 2f2–f1 distortion product otoacoustic emission in humans

Hanna K. Wilson[a] and Mark E. Lutman
Institute of Sound and Vibration Research, University of Southampton, SO17 1BJ, United Kingdom

(Received 20 March 2006; revised 12 July 2006; accepted 14 July 2006)

The 2f1–f2 distortion product otoacoustic emission (DPOAE) is considered to consist of two components in normally hearing ears, one having constant phase with changing DP frequency (wave fixed) and one having an increasing phase lag with increasing frequency (place fixed). The aim was to identify the wave-fixed and place-fixed components of both 2f1–f2 and 2f2–f1 DPs, and, in particular, to show whether a wave-fixed 2f2–f1 DP exists in normally hearing adults. DPOAE recordings were made in 20 ears of normally hearing young adults. Four frequency ratios were used and recording entailed fixed frequency-ratio sweeps. A separation into wave-fixed and place-fixed components was carried out using a time-window separation method. A method for estimating the noise floor after data processing was developed. Results confirmed the presence of wave-fixed and place-fixed components for 2f1–f2, consistent with previous studies. Both components were also present for 2f2–f1 in virtually all subjects. This latter finding conflicts with current models of DPOAE generation, and so a modified model is proposed. Unlike the 2f1–f2 emission, which has a wave-fixed component that is strongly dependent on the frequency ratio, neither component of the 2f2–f1 emission showed such a dependence. The proposed model explains these findings in terms of the overlap of the primary frequency traveling waves.
[DOI: 10.1121/1.2335421]

PACS number(s): 43.64.Jb, 43.64.Bt [BLM] Pages: 2108–2115

I. INTRODUCTION

Since their discovery by Kemp (1978), otoacoustic emissions (OAEs) have been extensively studied, but their mechanisms are not yet fully understood. Understanding how OAEs relate to the functions of the normal cochlea will increase their value as a clinical tool. An important form of OAE is the distortion product (DPOAE), which is evident at intermodulation frequencies of two primary tones at frequencies symbolized by f1 and f2 ($f1 < f2$). The largest and most widely studied component is the cubic DPOAE at 2f1–f2. The alternative cubic DPOAE at 2f2–f1 is smaller and is less commonly recorded.

It is widely accepted that there are multiple sources of DPOAEs, as reviewed by Shaffer *et al.* (2003). Experiments by Martin *et al.* (1987) using suppression, interfering tones, and temporary threshold shift in the rabbit ear showed that the likely regions of generation of the 2f1–f2 DP emission were the region of maximum primary wave overlap and the primary frequency regions. For the 2f2–f1 DP emission, however, the region close to the DP frequency place appeared to have more influence on the DP amplitude. Martin *et al.* (1998) showed by similar techniques that these findings were supported in human ears as well.

Different terminology has been used to describe the DPOAE sources in the cochlea. Kemp (1986) and Knight and Kemp (1999, 2000, 2001) use the terms "place-fixed" and "wave-fixed" to distinguish between sources that arise from a particular place on the basilar membrane or location on the traveling wave, respectively. Shera and Guinan (1999) and Kalluri and Shera (2001) use the terms "distortion emission" and "reflection emission," placing the emphasis on different mechanisms.

Zweig and Shera (1995) presented a model for the generation of stimulus frequency OAEs that involves waves reflecting off random irregularities in the organ of Corti. These irregularities may be thought of as variation in the number and spacing of outer hair cells or variations in the gain of the active process along the basilar membrane. If an irregularity causes reflections to occur at a specific place on the basilar membrane then this mechanism may also explain the generation of place-fixed DPOAEs.

Current theory suggests that the 2f1–f2 DPOAEs arise originally from nonlinear mixing of the two primary tones at around the place on the basilar membrane, where there is maximum overlap of their traveling wave envelopes (Kim, 1980). This causes a new traveling wave to be generated with components in two directions. The reverse traveling wave carries the initial (wave-fixed, distortion) component back toward the ear canal, while the forward component travels apically to its characteristic frequency place, where it may be reflected to generate a delayed (place-fixed, reflection) component. It is also possible that the 2f1–f2 component might travel directly to the stapes by a "fast" pressure wave in the cochlear fluid (Ren, 2004; Dong and Olson, 2005). For the 2f2–f1 component, the maximum overlap region is at a lower frequency than the DP and hence apical from the characteristic place of the DP. This means that the DP cannot propagate as a traveling wave from the maximum overlap region. (The basilar membrane is stiffness controlled basal to the characteristic place, which enables a propagating

[a]Now at the Department of Audiology, Royal Gwent Hospital, Cardiff Road, Newport NP20 2UB

traveling wave. At frequencies above the characteristic place, it becomes mass controlled and wave motion becomes evanescent, diminishing rapidly). Therefore the 2f2–f1 DPOAE must either be generated at a more basal region, where it can propagate as a traveling wave, or the return path to the ear canal must be via a fast pressure wave in the cochlear fluids.

Studies of the effects of stimulus frequency on DPOAE amplitude use one of three methods. f1 can be held constant while f2 is varied (f2 sweep), f2 can be held constant while f1 is varied (f1 sweep), or both frequencies can be varied while the frequency ratio is held constant. Knight and Kemp (2000) used all three methods to make a detailed study of DPOAE behavior. They presented their results on a "map" of DP amplitude and phase against coordinates of DP frequency and f2/f1 ratio. The f2/f1 ratio axis is numbered in both vertical directions from the center ($f2/f1=1$), with the 2f1–f2 DP in the upper half of the map and the 2f2–f1 DP in the lower half. The continuity between these two halves can be thought of in terms of place on the basilar membrane (i.e., the smaller the f2/f1 ratio, the closer the regions of DP generation). The phase maps obtained could clearly be divided into two regions, with the transition close to $f2/f1=1.1$ in the 2f1–f2 region. This provided evidence that the mechanisms of generation for 2f2–f1 and for 2f1–f2 at small frequency ratios are probably the same, while it seems likely that there is a different source for the 2f1–f2 DP at larger frequency ratios. Knight and Kemp suggested that there may be a crossover region in which both sources operate. Another paper from the same year gave a theoretical analysis of f1 and f2 sweeps (Prijs *et al.*, 2000). They found that with the parameters they used, the 2f1–f2 emission could be effectively explained as a single-source wave-fixed emission, but that the 2f2–f1 could not be assigned to a single-source mechanism.

Konrad-Martin *et al.* (2001) and others have shown that it is possible to suppress one source (reflection) for the 2f1–f2 using a third tone played into the ear. The rationale for this mechanism relies on the two source locations being physically remote and has the limitation that the suppression may also affect the source being studied. Separation of the 2f1–f2 DPOAE by latency (time-window separation) has also shown two components with very different dependencies on the frequency ratio (Knight and Kemp, 2001). It may be inferred from these studies that the two mechanisms of generation of the DPOAE occur in parallel, with the dominant mechanism depending on the recording conditions. A detailed study of the parameters that determine which mechanism contributes most to the DP amplitude has since been carried out by Dhar *et al.* (2005).

Suppression and time-window separation have been shown to give comparable results at moderate primary frequencies and for $f2/f1=1.2$ (Kalluri and Shera, 2001; Konrad-Martin *et al.*, 2001). This implies that the two different mechanisms of 2f1–f2 DP generation do, in fact, arise from different locations (Kalluri and Shera, 2001).

A surprising finding in the Knight and Kemp analysis (2001) was that in one of their two subjects they observed a significant wave-fixed component in the 2f2–f1 DP. This observation has not been made in humans in other published studies (although it has been noted in the guinea pig by Withnell *et al.*, 2003). Both Knight and Kemp (2001) and Withnell *et al.* (2003) suggest that the emission must be from the DP frequency region of the cochlea, rather than from the region of maximum primary-frequency wave overlap (which is the current proposal for the 2f1–f2 DP). However, both 2f1–f2 and 2f2–f1 wave-fixed DPs are thought to arise from interactions between the primary frequency waves, which occur along their entire length on the basilar membrane. The wave-fixed 2f2–f1 component has only been recorded in one human subject, which raises questions about the generality of the finding. The second subject in the Knight and Kemp study did not appear to have this component, although the noise floor of the recording was rather high and the second subject may have had a hidden wave-fixed DP at 2f2–f1.

In this study we aim to address the question of whether the wave-fixed 2f2–f1 DPOAE does occur more generally by testing more subjects in better recording conditions. If it does generally occur, this study also aims to establish the magnitude of the wave-fixed component and its pattern with respect to frequency ratio. This may suggest a revised theory of its generation.

II. METHOD

All tests were carried out in an audiometric booth, which complied with ISO 8253-1 for the measurement of hearing threshold level down to 0 dB HL. In addition, the background noise was always below 30 dB(A), the minimum level measurable with the sound level meter used for monitoring purposes.

A. Subjects

Here 20 subjects were tested (14 female and 6 male). The ages of the subjects ranged from 22 to 30 years (mean 25.9 years, SD 2.3 years). Hearing thresholds were at or below 20 dB HL at the common test frequencies between 250 and 8000 Hz. The ear with lower pure tone thresholds was tested—or the one with smaller SOAEs if thresholds were the same. Subjects were screened using otoscopy, tympanometry, and a questionnaire. Subjects were excluded if they had family history of congenital hearing loss at a young age, a history of excessive noise exposure, ototoxic medication, ear disorder, ear operations, fluctuating hearing, or tinnitus.

B. Equipment

DPOAEs were measured using custom apparatus described previously by Parazzini *et al.* (2005). A digital signal-processing (DSP) card in a PC was attached to an external analog-digital (A/D) and digital-analog (D/A) converter unit (Institute of Hearing Research DSP remote converter module), running at a sample rate of 32.768 kHz. The stimuli were delivered from this system by two Etymotic Research ER-2 insert earphones, connected to a common probe (Etymotic Research ER-10B+). The probe was sealed into the ear canal of the subject using a standard tympanometry probe tip. The probe also contained a low noise microphone, which recorded the sound pressure in the ear canal. The microphone output was amplified by 40 dB by the

ER-10B+ system, fed to the A/D converter and collected by the DSP card. Epochs of 62.5 ms were collected and converted to the frequency domain by FFT. The averaged results were displayed on the PC screen. The phase and magnitude of the DPOAEs were calculated by averaging a number of the epochs up to a maximum of 100. The equipment recorded the phase of the primaries and each DP in degrees and the amplitudes in dB.

The calibration was carried out using an audio frequency spectrometer and IEC-711 ear simulator. The OAE probe was coupled to the ear simulator by a tympanometry tip. The insert earphones were calibrated individually with a resolution of ±0.5 dB, at frequency intervals of 256 Hz from 256 to 10 240 Hz. The calibration corrections obtained were subsequently applied automatically to stimulus levels produced by the earphones (interpolated linearly for intermediate frequencies).

The probe microphone was calibrated at 1024 Hz only. This was to avoid the effects of standing waves at higher frequencies, which could give a variable response depending on the position of the microphone within the ear simulator. The specified microphone frequency response is approximately linear over the range of frequencies to be used and so the single correction factor was applied to the entire range.

C. Parameters

Primary frequencies are denoted by f1 and f2 and the primary levels by L1 and L2. Parameters were chosen so that subjects would only be required for one test session of no more than 45 min. Four frequency ratios were used: f2/f1 =1.05, 1.10, 1.22, and 1.32. The range of frequencies for f2 was approximately 1.0–2.5 kHz. Fixed frequency ratio sweeps were collected and plotted as a function of f2, with data collected every 16 Hz (i.e., with the frequency ratio held constant, the frequency of f2 was increased from 1.0 to 2.5 kHz in 16 Hz increments). The levels used were L1 =65 dB, L2=60 dB. This is thought to be the region of compressive nonlinearity described by Dorn *et al.* (2001) and therefore probes normal cochlear function, where it is most distinct from the nonfunctioning cochlea.

D. Data processing

Existing program from a previous study by Parazzini *et al.* (2005) were modified to be suitable for the data collected in this study.

1. Program 1: Data formatting and phase unwrapping

This program allowed data to be imported from the DPOAE recording software. The phase data, which were originally constrained to a 360° range, were unwrapped. Selected parts of the data were stored in a format suitable for further processing. Note that DP phase data were referenced to the phase expected for an instantaneous cubic nonlinearity, $2\phi_1$-ϕ_2 and $2\phi_2$-ϕ_1 for the DPs at 2f1–f2 and 2f2–f1, respectively, where ϕ_1 and ϕ_2 are the phases of the primaries measured by the probe microphone.

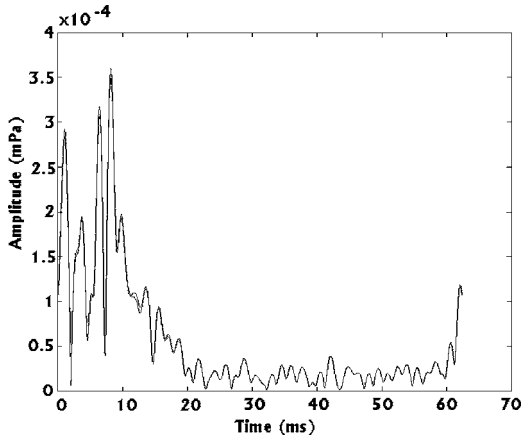


FIG. 1. Example time-domain representation of a 2f2–f1 DPOAE, showing a single wave-fixed (short latency) component at <2 ms (first peak) and a series of place-fixed (long latency) emissions.

2. Program 2: Unmixing

This program carried out time-window separation according to the method of Withnell *et al.* (2003). It produced a time-domain representation of the DPOAE (see Fig. 1) and then separated this into short latency (wave-fixed) and long latency (place-fixed) parts. The cutoff time between the short and long latency parts was set at 2 ms, based on an observation of the time domain representation, which appears to fall to a minimum at around 3–4 ms. A tenth order recursive exponential filter was used to separate the two components; since this has been effective in previous studies (for example, Kalluri and Shera, 2001).

The program output includes plots of the phase and amplitude of both the wave-fixed and place-fixed components in the frequency domain (see Fig. 2).

3. Program 3: Averaging

In order to obtain a single amplitude value for each recording condition, the data were power averaged across frequency for each fixed frequency-ratio sweep.

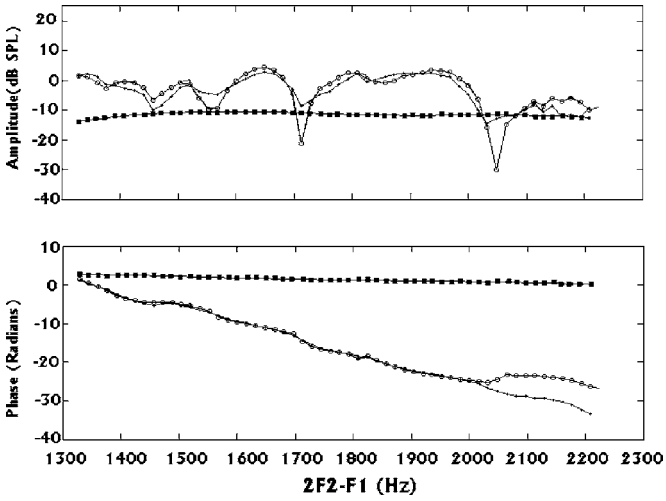


FIG. 2. Example amplitude and phase plots for a 2f2–f1 DPOAE sweep after separation into wave-fixed (distortion) and place-fixed (reflection). The open circles represent the original data before separation. The squares correspond to the wave-fixed emission and the small crosses to the place-fixed emission.

E. Noise floor estimation

The noise floor is discussed very little in the literature on DPOAE separation. Knight and Kemp (2001) made a single estimate of the noise floor for each 2f1–f2 and 2f2–f1 emission. This does not take into account the data processing carried out on the signal and may therefore be unnecessarily high. As there is no standard method for noise floor estimation, the following method was developed using a number of basic assumptions about the noise.

(1) The phase of the noise is random with respect to frequency.

(2) The noise frequency spectrum is approximately flat for the frequency range of interest. This was verified by performing an IFFT on noise recordings with a random phase assigned. It was found that the resulting time-domain noise was approximately stationary (i.e., the average level is the same in the short latency and long latency parts).

(3) There is no interaction between emission and noise; hence the emission observed is simply made up of signal +noise.

(4) The noise component within the signal is represented by the noise in the neighboring frequencies.

Measurements of the amplitude of the noise at frequencies close to the DP frequencies were made during the DPOAE recording. The recorded amplitudes were assigned a random phase. The amplitude and phase data were then treated in exactly the same way as the DPOAE data and averaged, as above. This resulted in separate estimates of the noise levels in the two latency periods for every recording condition for each subject. Thus, the emission and noise amplitudes could be compared directly to assess whether the emissions were significantly above the noise.

Data processing reduces the noise within each individual measurement, compared to the actual background noise level within the recording booth, for a several reasons. First, the recording is carried out over a limited range of frequencies, so noise outside this range does not form part of the data. Second, the unmixing program includes the summing of individual windows, which will have an effect similar to signal averaging. Third, the noise resulting from the unmixing is reduced as a result of the size of the time window used; the noise in the short window appears lower because it has a shorter duration. Exactly the same effect occurs with the emissions themselves.

F. Statistical analysis

A student's t test was used to compare the emission and noise within each recording condition, and hence to determine whether the emission amplitude was significantly above the noise. Confidence intervals for the signal-to-noise ratios were also obtained.

A repeated analysis of variance (ANOVA) measure was used to assess differences across the frequency ratio within each type of emission (wave-fixed and place-fixed, 2f1–f2 and 2f2–f1).

Pearson's correlation coefficient was calculated to look for a significant correlation between the 2f1–f2 and 2f2–f1 DPs at each frequency ratio. This assessed whether an individual with a high-level 2f1–f2 emission also tended to have a high-level 2f2–f1 emission under the same conditions. This would be an indication that the two emission types have a common element in their mechanism of generation. Pearson's correlation coefficient was also used to test for a significant correlation between the wave-fixed and place-fixed emissions at each frequency ratio. This assessed whether an individual with a high level wave-fixed emission also tended to have a high level place-fixed emission under the same conditions. This could indicate that the two types of component are related in their generation.

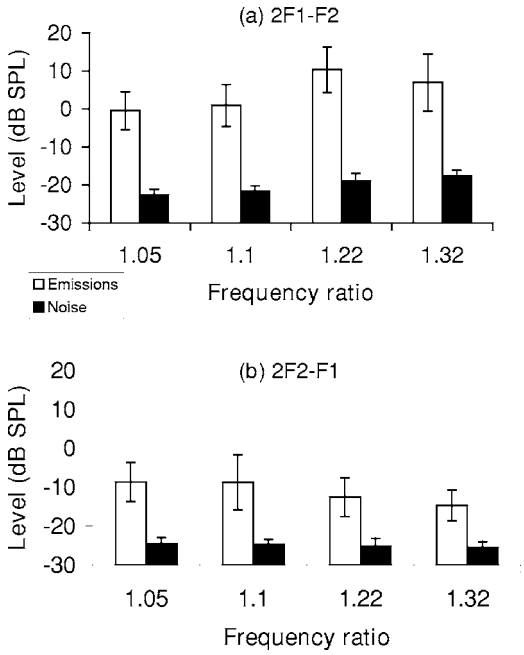


FIG. 3. Emissions (white bars) and adjacent noise levels (black bars), averaged across subjects for (a) 2f1–f2 emissions and (b) 2f2–f1 emissions. Error bars are one standard deviation.

III. RESULTS

A. Individual analysis

It is important to examine the emissions from each subject individually to assess whether they appear genuinely to be present. A "present" emission was considered to be any emission level more than one standard deviation above the noise recording to be reasonably certain that the measurement was not a chance occurrence within the noise. Emissions were present in all the 2f1–f2 and wave-fixed 2f2–f1 DP recordings and were present in 18 out of 20 place-fixed 2f2–f1 DP recordings. From observation, emissions did not meet the "present" criterion if there was a greater level of background noise in the recording and the emission was at a relatively low level.

Figure 3 shows the mean levels for each condition with the corresponding mean noise level. The emissions are well separated from the noise in every condition.

B. Emissions after component separation

The data recorded were processed by time-window separation. An example time domain record after IFFT is shown

in Fig. 1. The amplitude is typical of the time domain representations obtained for both 2f2–f1 and 2f1–f2 DPOAEs. The time domain records generally show a peak before 2 ms and a series of peaks in the range 3–20 ms. From observation of these windows, a cutoff time of 2 ms to separate wave-fixed and place-fixed emissions (based on Knight and Kemp, 2001) appears to be appropriate. Typical examples of amplitude and phase plots in the frequency domain after separation are shown in Fig. 2. It is clear from these plots that the emissions can be split into two components, one with a steady phase across frequency and one with a sloping phase across frequency.

The separated data were power averaged in the same way as the unseparated data in order to obtain single amplitude measurements across all frequencies for each recording condition for each subject. The corresponding noise recordings were averaged in the same way. For each recording condition the mean and standard deviation of the data across subjects were calculated. Just as in the data before separation, the emissions are well separated from the noise.

C. Statistical analysis

1. Differences between emissions and noise

Paired student's t tests were performed on each emissions sample with its corresponding noise sample. The emissions were significantly above the noise ($p<0.01$) in every case. The t tests were also used to obtain confidence intervals for the difference (i.e., the signal-to-noise ratio, SNR). At the lower limit of the confidence intervals, this ranged from 7.7 to 42.3 dB, with wave-fixed emissions generally having a higher SNR than place-fixed emissions.

Pearson's correlation coefficient (r) was used to look for a correlation between the signal and noise (i.e., whether the highest emission levels were from the same subjects as the highest levels of noise). The emissions and noise were not correlated.

2. Differences between frequency ratios

A repeated-measures analysis of variance was used to show whether the pattern of DPOAE amplitudes across the frequency ratio (as in Fig. 3) was consistent among the group of subjects. The amplitudes were found to be significantly different ($p<0.05$) for different frequency ratios in all except two cases, indicating that there is a consistent pattern.

3. Correlation across DP type

Pearson's correlation coefficient was used to look for a significant correlation between the 2f1–f2 and 2f2–f1 DPs at each frequency ratio. There was significant correlation between every pair ($r>0.46$, $p<0.05$) indicating that when an individual has a high-level 2f1–f2 emission, they also tend to have a high-level 2f2–f1 emission under the same conditions.

4. Correlation across component type

Pearson's correlation coefficient was also used to look for a significant correlation between the wave-fixed and place-fixed emissions at each frequency ratio. Seven out of eight comparisons were significant ($r>0.49$, $p<0.05$), indicating that generally an individual with a high-level wave-fixed emission is also likely to have a high-level place-fixed emission under the same conditions.

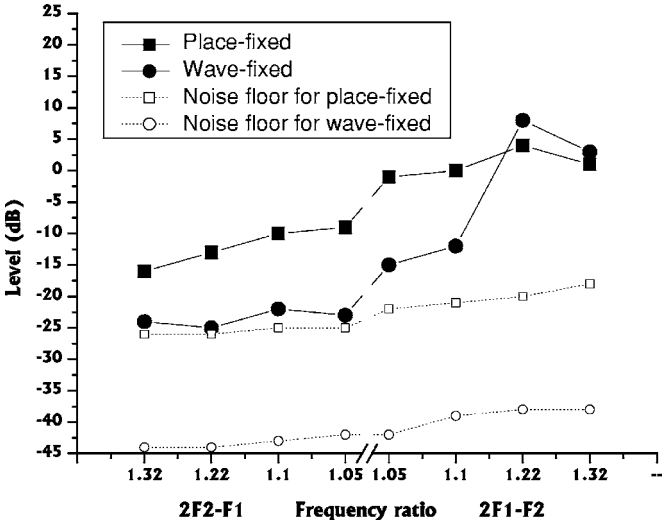


FIG. 4. Average emission and noise levels plotted for both 2f1–f2 (the right half of the graph) and 2f2–f1 (the left half). The solid squares represent the place-fixed emissions and the solid circles represent the wave-fixed emissions. The open squares represent the estimated noise floor for the place-fixed emission and the open circles for the wave-fixed emission.

IV. DISCUSSION

A. Comparison with literature

In general, the results of this study are in good agreement with the previously published literature. The 2f1–f2 DP has been investigated in other studies in far more detail than in the present study (for example, Dhar *et al.*, 2005). The fact that the 2f1–f2 DP emission data in the present study are in agreement with previous findings supports the methodology used and gives an indication that the findings made regarding the 2f2–f1 DP emissions are robust.

It has been suggested on an empirical basis, following animal studies (Withnell *et al.*, 2003) and human studies (Knight and Kemp, 2001), that the 2f2–f1 might be separated into wave-fixed and place-fixed components, and this study has confirmed this proposal.

In their 2001 paper, Knight and Kemp summarized their results in a graph similar to Fig. 4. Frequency ratio is plotted outward in both directions from $f2/f1=1$ in the center. The left hand part is for 2f2–f1 and the right hand part is for 2f1–f2. Figure 4 also shows the corresponding noise floor for each emission source, calculated by the method described above. The general trend of these results is comparable with those of Knight and Kemp. As reported by Knight and Kemp, the 2f1–f2 emission has a wave-fixed component that rises to a maximum at around $f2/f1=1.22$ and a place-fixed component that remains at a more constant level across frequency ratio. The 2f2–f1 emission appears to have a dominant place-fixed component and a weaker wave-fixed component, neither of which have a strong dependence on the frequency ratio. In the Knight and Kemp study the transition in the 2f1–f2 DP from predominantly place-fixed to predominantly wave-fixed occurs in a region around $f2/f1=1.15$,

where the wave-fixed component increases steeply with an increasing frequency ratio. In the present study, it can be seen that the transition does occur, although there are insufficient data points to determine exactly at what frequency ratio.

Knight and Kemp tested only two subjects in their study, one of whom showed relatively high-level emissions under every condition, compared to the other. This has been found to be a typical pattern in the present study, with the emission levels correlating well between different conditions and emission types.

A recent paper by Mauermann and Kollmeier (2004) found that wave-fixed (distortion) emissions show smooth trends in the frequency domain. This is consistent with the patterns found in this study and can be seen clearly in Fig. 2. They have proposed the use of distortion-only emissions to eliminate the fine structure normally associated with DPOAEs, thus increasing their reliability in predicting hearing thresholds. The finding of smooth functions is also consistent with the proposed wave-fixed generation mechanism, where there is minimal variation in relative traveling wave shapes with frequency, consistent with traveling wave functions that scale with log frequency.

B. Current understanding of the generation of the 2f2–f1 DP

There is little published literature on the mechanism of generation of the 2f2–f1 emission, although several authors provide evidence that the source is located basal to the DP frequency place on the basilar membrane. For example, Martin *et al.* (1987) showed that temporary threshold shifts influence the amplitude of the 2f2–f1 DP emission to the greatest extent close to the DP frequency. Martin *et al.* (1998) showed that 2f2–f1 suppression tuning curves peak above f2, which shows that at least the dominant generation source is basal to the f2 frequency place. These findings have since been evaluated by theoretical treatments of DPOAE sources, such as Prijs *et al.* (2000). This analysis also showed that the 2f2–f1 DP emission could not be entirely either wave-fixed or place-fixed, which indicates either a combination of mechanisms or a mechanism that is entirely different than those proposed for the 2f1–f2 DP.

Shera and Guinan (1999) argue that the location of the DPOAE source is unimportant for the time-window separation of DPOAEs. They consider that the separation technique works because the emission components have fundamentally different mechanisms that lead to different latencies. If Shera and Guinan are correct, then time-window separation should be effective on the 2f2–f1 DP, even if two mechanisms occur at the same location, whereas separation using suppression would not.

Knight and Kemp (2001) and Withnell *et al.* (2003) have used time-window separation to show that wave-fixed and place-fixed mechanisms occur simultaneously. It is possible that the two mechanisms are both in operation at the 2f2–f1 frequency-place, although it is also feasible that nonlinear (wave-fixed) mixing may take place in a range of locations basal to the 2f2–f1 place (or all along the region of primary wave interaction). Like the 2f1–f2 DP, the 2f2–f1 DP consists of a number of peaks when expressed in the time domain. One interpretation for this is that there are multiple reflections within the cochlea, which are driven by the activity of the outer hair cells.

Knight and Kemp (2001), Withnell *et al.* (2003), and Meinke *et al.* (2006) have provided experimental evidence that the place-fixed, or reflection, component is the dominant source of the 2f2–f1 DP. Meinke *et al.* (2006) expanded upon the work of Knight and Kemp (2000) using a larger subject group, creating "maps" of the dominant mechanism of the generation of distortion products, but without carrying out any separation. Reflection emissions are generally interpreted as arising from imperfections on the basilar membrane, combined with a "tall and broad" filter mechanism. These imperfections must be basal to the site of reflection because the waves arrive at the site of reflection at the distortion product frequency. Unlike distortion emissions, they are therefore slowed by the highly tuned mechanics of the basilar membrane, and this causes their characteristic long latency.

Dong and Olson (2005) found that the 2f2–f1 shows phase behavior that does not appear to be comparable to either place-fixed or wave-fixed 2f1–f2 emissions (as predicted by Prijs *et al.*, 2000). It is possible that the 2f2–f1 is produced by a different mechanism altogether, but that two components with different latencies are seen due to the way the waves propagate back to the stapes (i.e., a traveling wave and a fluid wave).

C. Proposed mechanism of generation for the 2f2–f1 DP

The wave-fixed component of the 2f2–f1 DP was not proposed until recently (Knight and Kemp, 2001) and has not been studied in a group of more than two human subjects until the present study. The mechanism producing this emission may be similar to the generation of wave-fixed 2f1–f2 emissions, both having a characteristic very short latency.

There is a correlation between the 2f2–f1 and 2f1–f2 DP amplitudes, as well as between the place-fixed and wave-fixed emission amplitudes in the group of subjects included in the present study. This seems to indicate that there is some common factor in the generation of these types of DPOAE. The significance of this finding is supported by the fact that the noise was not correlated with any of the emissions.

Figure 5 shows a mechanism proposed for the generation of the 2f2–f1 (based on Kalluri and Shera, 2001). The proposal is that waves of the frequency 2f2–f1 are generated by a simple nonlinear mechanism basal to the 2f2–f1 frequency place. They are then propagated forward to their characteristic place, where they are slowed by the mechanics of the basilar membrane, causing the delay typical of reflection emissions. Some energy is reflected back via a reverse traveling wave to the base of the cochlea and emitted into the ear canal. It is possible that there are a number of reflection sites (at the characteristic DP place and any imperfections basal to it). In addition to this, the DP frequency wave may arise from anywhere basal to the DP frequency place, giving the spread of place-fixed latencies seen in Fig. 1. Waves at the fundamental frequencies f1 and f2 passing the 2f2–f1 DP

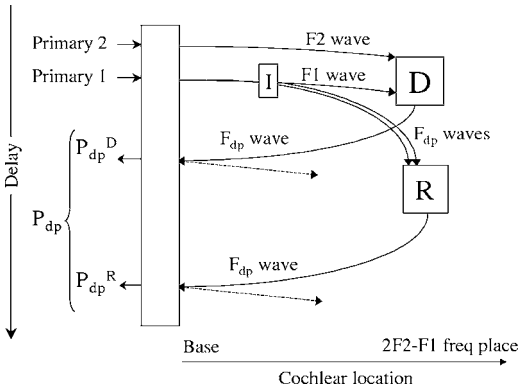


FIG. 5. Proposed mechanism for the generation of 2f2–f1 DPOAEs. Distortion (D) and reflection (R) both occur at the 2f2–f1 frequency place on the basilar membrane. Distortion (D) creates a DP frequency wave that propagates directly back to the base of the cochlea. Waves of the DP frequency, created from imperfections within the basilar membrane (I), slow as they arrive at the DP frequency place and are reflected (R) back to the base of the cochlea.

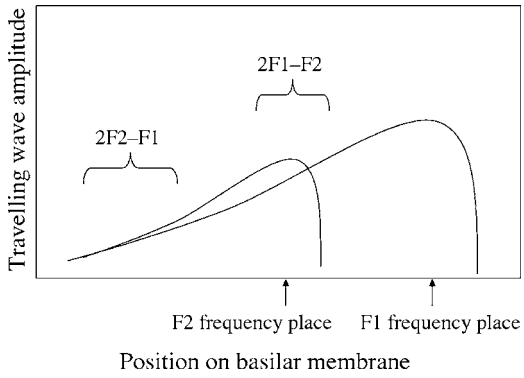


FIG. 6. The proposed region of generation of the wave-fixed 2f1–f2 and 2f2–f1 emissions. The amplitude of the waves is low at the site of 2f2–f1 DP generation, which could explain why the wave-fixed emissions measured are very low.

frequency place are basal to their characteristic frequency and are therefore not yet slowed materially by the mechanics of the cochlea. Nonlinearity at the 2f2–f1 DP frequency place causes energy to be emitted at the DP frequency, and this is propagated back to the ear canal via a reverse traveling wave or possibly more directly via the cochlear fluids. In this way, the wave-fixed source behaves like an actuator, producing waves at the DP frequency. The fact that neither of the primaries has reached its characteristic place at the site where distortion is generated may account for the relatively low level of this component for the 2f2–f1 emission (Fig. 4).

Interpretation of the work of Knight and Kemp (2000) implies that it is possible that the distortion source is in a wave-fixed location, basal to the DP frequency place. In this model, waves of the 2f2–f1 DP frequency are propagated forward to the reflection site more than they are propagated back to the ear canal. This offers another plausible explanation as to why the wave-fixed (distortion) emission is weaker.

In this proposal, it is possible for both place-fixed (reflection) and wave-fixed (distortion) emissions to be in operation at the DP frequency place or basal to it. It is the nature of the two mechanisms that give the long and short latencies.

D. Dependence of emission levels on frequency ratio

The place-fixed (reflection) emission is similar in amplitude across frequency ratios for both 2f1–f2 and 2f2–f1 DPs. This is probably because both arise from random imperfections on the basilar membrane, which will be approximately evenly distributed in a normal subject, and both involve reflection from their characteristic place where full cochlear amplification has occurred.

The wave-fixed (distortion) emission, however, is highly dependent on frequency-ratio for the 2f1–f2 DP, while being much smaller in amplitude and less frequency-ratio dependent for the 2f2–f1 DP. These results can be thought of, schematically, in terms of the amplitude and overlap of the primary frequency traveling waves (Fig. 6).

The amplitude of the 2f1–f2 wave-fixed DP rises to a maximum at around $f2/f1=1.22$ at the primary tone levels tested. Recall that this DP is thought to arise from nonlinearity in the region of overlap, at or close to the f2 frequency place. The maximum amplitude comes from an optimum overlap between the two traveling waves and depends on the levels and frequency ratio of the primaries. This complex relationship has been modeled by Lukashkin and Russell (2001), explaining why the overlap peaks for frequency ratios occurs in the region of $f2/f1=1.2$.

The 2f2–f1 emission arises at, or basal to, the 2f2–f1 characteristic place. Here, the amplitudes of the two primary-frequency traveling waves are small, because they are more basal than their characteristic places. Changing the frequency ratio will have little impact on their overlap (Fig. 6), so the emission amplitude will have little dependence on the frequency ratio.

V. CONCLUSION

As observed in one of two subjects by Knight and Kemp (2001), the 2f2–f1 distortion product otoacoustic emission consists of two components in normally hearing adult human subjects, one of which has a constant phase with changing DP frequency (wave-fixed) and one of which has a sloping phase with changing frequency (place-fixed). It may be assumed that these correspond to reflection and distortion mechanisms, respectively. The two components have been separated by latency—an approach that does not necessarily depend on the two components arising from different locations within the cochlea.

The separation demonstrates the existence of wave-fixed and place-fixed components well above the background noise in almost all the 2f2–f1 DPOAEs recorded. This suggests that the wave-fixed 2f2–f1 emission component is a general feature in normal hearing subjects. The level of the 2f2–f1 DP emission was well below that of the 2f1–f2 DP, irrespective of the frequency ratio.

Neither the wave-fixed nor the place-fixed component of the 2f2–f1 DPOAE appears to depend strongly on the frequency ratio. This contrasts with the strong pattern across frequency ratio seen for the wave-fixed 2f1–f2. This can be

explained in terms of traveling wave overlap on the basilar membrane, in a way analogous to the widely accepted theory for 2f1–f2 DPOAEs.

Dhar, S., Long, G. R., Talmadge, C. L., and Tubis, A. (**2005**). "The effect of stimulus-frequency ratio on distortion product otoacoustic emission components," J. Acoust. Soc. Am. **117**, 3766–3776.

Dong, W., and Olson, E. S. (**2005**). "Two-tone distortion in intracochlear pressure," J. Acoust. Soc. Am. **117**, 2999–3015.

Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., Cyr, E., and Gorga, M. P. (**2001**). "Distortion product otoacoustic emission input/output functions in normal-hearing and hearing-impaired human ears," J. Acoust. Soc. Am. **110**, 3119–3131.

Kalluri, R., and Shera, C. A. (**2001**). "Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation," J. Acoust. Soc. Am. **109**, 622–637.

Kemp, D. T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T. (**1986**). "Otoacoustic emissions, travelling waves and cochlear mechanisms," Hear. Res. **22**, 95–104.

Kim, D. O. (**1980**). "Cochlear mechanics: Implication of electrophysiological and acoustical observations," Hear. Res. **2**, 297–317.

Knight, R. D., and Kemp, D. T. (**1999**). "Relationships between DPOAE and TEOAE amplitude and phase characteristics," J. Acoust. Soc. Am. **106**, 1420–1435.

Knight, R. D., and Kemp, D. T. (**2000**). "Indications of different distortion product otoacoustic emission mechanisms from a detailed f(1), f(2) area study," J. Acoust. Soc. Am. **107**, 457–473.

Knight, R. D., and Kemp, D. T. (**2001**). "Wave and place fixed DPOAE maps of the human ear," J. Acoust. Soc. Am. **109**, 1513–1525.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). "Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears," J. Acoust. Soc. Am. **109**, 2862–2879.

Lukashkin, A. N., and Russell, I. J. (**2001**). "Origin of the bell-like dependence of the DPOAE amplitude on primary frequency ratio," J. Acoust. Soc. Am. **110**, 3097–3106.

Martin, G. K., Lonsbury-Martin, B. L., Probst, R., Scheinin, S. A., and Coats, A. C. (**1987**). "Acoustic distortion products in rabbit ear canal. II. Sites of origin revealed by suppression contours and pure-tone exposures," Hear. Res. **28**, 191–208.

Martin, G. K., Jassir, D., Stagner, B. B., Whitehead, M. L., and Lonsbury-Martin, B. L. (**1998**). "Locus of generation for the 2f(1)-f(2) vs. 2f(2)-f(1) distortion-product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies, and amplitude correlations," J. Acoust. Soc. Am. **103**, 1957–1971.

Mauermann, M., and Kollmeier, B. (**2004**). "Distortion product otoacoustic emission (DPOAE) input/output functions and the influence of the second DPOAE source," J. Acoust. Soc. Am. **116**, 2199–2212.

Meinke, D., Martin, G., and Stagner, B. (**2006**). "A harmonic difference tone component in low-frequency human DPOAEs?" J. Assoc. Res. Otolaryngol. **29**, 57.

Parazzini, M., Bell, S., Thuroczy, G., Molnar, F., Tognola, G., Lutman, M. E., and Ravazzani, P. (**2005**). "Influence on the mechanisms of generation of distortion product otoacoustic emissions of mobile phone exposure," Hear. Res. **208**, 68–78.

Prijs, V. F., Schneider, S., and Schoonhoven, R. (**2000**). "Group delays of distortion product otoacoustic emissions: relating delays measured with f1- and f2-sweep paradigms," J. Acoust. Soc. Am. **107**, 3298–3307.

Ren, T. (**2004**). "Reverse propagation of sound in the gerbil cochlea," Nat. Neurosci. **7**, 333–334.

Shaffer, L. A., Withnell, R. H., Dhar, S., Lilly, D. J., Goodman, S. S., and Harmon, K. M. (**2003**). "Sources and mechanisms of DPOAE generation: Implications for the prediction of auditory sensitivity," Ear Hear. **24**, 367–379.

Shera, C. A., and Guinan, J. J. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.

Withnell, R. H., Shaffer, L. A., and Talmadge, C. L. (**2003**). "Generation of DPOAEs in the guinea pig," Hear. Res. **178**, 106–117.

Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.

In search of basal distortion product generators

Robert H. Withnell[a] and Jill Lodde
Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, Bloomington, Indiana 47405

(Received 8 April 2006; revised 12 July 2006; accepted 26 July 2006)

The $2f_1$-f_2 distortion product otoacoustic emission (DPOAE) is thought to arise primarily from the complex interaction of components that come from two different cochlear locations. Such distortion has its origin in the nonlinear interaction on the basilar membrane of the excitation patterns resulting from the two stimulus tones, f_1 and f_2. Here we examine the spatial extent of initial generation of the $2f_1$-f_2 OAE by acoustically traumatizing the base of the cochlea and so eliminating the contribution of the basal region of the cochlea to the generation of $2f_1$-f_2. Explicitly, amplitude-modulated, or continuously varying in level, stimulus tones with $f_2/f_1=1.2$ and $f_2=8000-8940$ Hz were used to generate the $2f_1$-f_2 DPOAE in guinea pig before and after acoustically traumatizing the basal region of the cochlea (the origin of any basal-to-f_2 distortion product generators). It was found, based on correlation analysis, that there does not appear to be a basal-to-f_2 distortion product generation mechanism contributing significantly to the guinea pig $2f_1$-f_2 OAE up to $L_1=80$ dB sound pressure level (SPL).
 [DOI: 10.1121/1.2338291]

PACS number(s): 43.64.Jb, 43.64.Kc [BLM] Pages: 2116–2123

I. INTRODUCTION

Since distortion product otoacoustic emissions (DPOAEs) were first reported in the literature (Kemp, 1979), an ever-increasing view of the complexity of their origin has developed over the years. Viewed as predominantly arising from the nonlinear interaction in the cochlea of a two stimulus tone input with the distortion propagating back to the stapes from the region of nonlinear interaction, i.e., the f_2 region (Brown and Kemp, 1984; Martin *et al.*, 1987), it was nevertheless recognized that, for the $2f_1$-f_2 DPOAE in particular, the emission measured in the ear canal could represent the sum of distortion generated in the region of f_2 and emission arising from the $2f_1$-f_2 region (Kim, 1980). This view of the $2f_1$-f_2 OAE being the vector sum of contributions from the f_2 region and $2f_1$-f_2 region was confirmed in the mid to late 1990s (Gaskill and Brown, 1996; Heitmann *et al.*, 1998; Mauermann *et al.*, 1999). Subsequent to the development of Kemp's idea (e.g., Kemp, 1986) of OAEs arising from either a place-fixed or a wave-fixed mechanism (Zweig and Shera, 1995; Talmadge *et al.*, 1998; Shera and Guinan, 1999), it was established that the $2f_1$-f_2 OAE not only arises from two discrete sources but that each of these sources has a different predominant mechanism of generation (Talmadge *et al.*, 1999; Kalluri and Shera, 2001). Thus the prevailing view for the origin of the $2f_1$-f_2 OAE (and, by extension, all DPOAEs with a frequency less than f_2) became one of the emissions being generated in two different cochlear locations, each location involving a different predominant mechanism of generation (see Fig. 1). This view for the origin of the $2f_1$-f_2 OAE seems to be applicable to all mammals, but differences in cochlear tuning and cochlear inhomogeneity may provide for the different emphases observed for the two components measured as a vector sum in the ear canal (e.g., Withnell *et al.*, 2003; Schneider *et al.*, 2003).

Additional complexity to the generation of DPOAEs was suggested by Fahey *et al.*, (2000) to provide for observed suppression/enhancement of DPOAEs by a third tone that had a frequency more than an octave above the stimulus tone frequencies. Two mechanisms were suggested to account for this effect: a catalyst mechanism and a harmonic mechanism. The catalyst mechanism involves a concatenation of nonlinear interaction of the three stimulus tone basilar membrane (BM) responses, their harmonics, and intermodulation distortion products. The harmonic mechanism involves the nonlinear interaction of the cochlear harmonic of one of the stimulus tones with the cochlear response to the other stimulus tone. The harmonic mechanism, it is suggested, provides a means for DPOAEs to be produced, in the absence of a third tone, additional to the view outlined in Fig. 1 (Fahey *et al.*, 2000; Martin *et al.*, 2003). This is illustrated schematically in Fig. 2. Because the harmonic mechanism involves nonlinear interaction with a harmonic of either f_1 or f_2, it is presumably a stimulus level dependent effect.

Evidence for a harmonic and/or catalyst mechanism based on DPOAE suppression tuning curves obtained with a third tone has been presented (e.g., Martin *et al.*, 1999). In rabbit, evidence for suppression/enhancement of the $2f_1$-f_2 DPOAE with a suppressor tone frequency associated with a cochlear location remote from that place in the cochlea tuned to the stimulus frequencies is observed concomitant with a cochlea that generates many orders of distortion products (e.g., Fahey *et al.*, 2006). It may be that the rabbit cochlea generates more distortion than other mammals, providing for the conjecture that the suppression/enhancement is an

[a]Author to whom correspondence should be addressed; electronic mail: rwithnel@indiana.edu

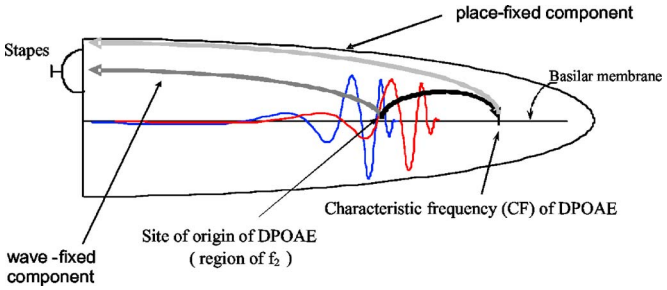


FIG. 1. (Color online) Schematic of origin of $2f_1$-f_2 OAE with the cochlea unfurled and scala media represented by the basilar membrane. The OAE measured in the ear canal is a vector sum of contributions arising from the f_2 and CF regions.

anomalous effect specific to rabbit but the effect has also been reported in guinea pig (Martin *et al.*, 1999) and humans (Martin *et al.*, 2003).

Fahey *et al.* (2000) suggested that the catalyst mechanism would be dominant in the high-frequency region of the cochlea while the harmonic mechanism would be dominant in the low-frequency region of the cochlea. This conclusion was based on the observation that the harmonic mechanism would not be expected to operate at intermediate and high frequencies where the high-frequency cutoff of the basilar membrane response is sharp. In guinea pig, the slope of the high-frequency cutoff of the basilar membrane "near-threshold" response (inferred from measurements of auditory nerve fiber tuning) increases from about 30 dB/octave below 1 kHz to greater than 100 dB/octave above 3 kHz (Evans, 1972). For fundamental frequencies of 3 kHz and higher, any second (or higher) harmonic produced is spatially confined on the basilar membrane to the region from the stapes to approximately its own characteristic place, i.e., no significant second harmonic is generated on the basilar membrane between the characteristic place of the second harmonic and the characteristic place of the fundamental. The second harmonic in such a case is being generated by nonlinear interaction well basal to the peak of the fundamental traveling wave and so is small, i.e., harmonics do not achieve sufficient amplitude at their own characteristic frequency place to generate sufficient nonlinear interaction to produce quadratic distortion that is a combination of the second harmonic of the fundamental with the BM response to the other stimulus tone. With increasing stimulus level, the basilar membrane response broadens with the peak of the response shifting basally (Johnstone *et al.*, 1986) producing a larger amplitude response to the fundamental tone basal to the characteristic place for the second harmonic. For the generation of $2f_1$-f_2, the second harmonic of f_1 generated with increasing stimulus level may achieve sufficient amplitude at its own characteristic place to interact nonlinearly with the BM response to the f_2 tone to produce $2f_1$-f_2 as a quadratic distortion tone.[1]

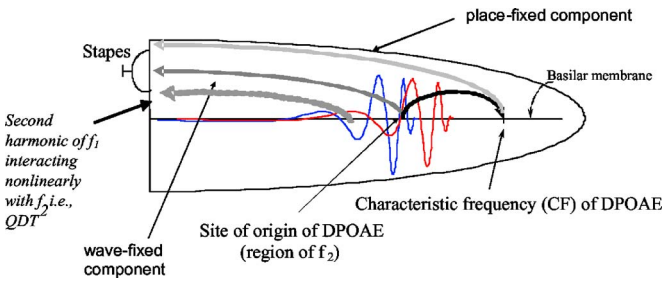


FIG. 2. (Color online) Schematic of origin of $2f_1$-f_2 OAE, including the speculated harmonic distortion mechanism, with the cochlea unfurled and scala media represented by the basilar membrane. The OAE measured in the ear canal is a vector sum of contributions arising from the $2f_1$, f_2, and CF regions.

The catalyst mechanism requires the simultaneous presentation of three (or more) stimulus tones with distortion products arising out of a concatenation of nonlinear interaction, "this heterodyning down…done in such a way that the value of the f_3 frequency is not contained in any final products" (Fahey *et al.*, 2000, p. 1796). For example, the nonlinear interaction in the cochlea of the BM responses to f_3 and f_1 may generate f_3-f_1 and the nonlinear interaction of f_3 and f_2 may generate f_3-f_2, the two intermodulation distortion products interacting nonlinearly in the region of f_3-f_1 in the cochlea to produce f_2-f_1. Fahey *et al.* (2000) give other examples that require the generation of harmonics of the stimulus tones in the cochlea, and so, third order distortion products, for instance (e.g., $2f_1$-f_2), are also dependent on a harmonic mechanism for their generation. Pertinent to the operation of a catalyst mechanism is the study of Kim *et al.* (1997) where distortion product otoacoustic emissions were generated by multiple stimulus tone pairs in humans. To test the hypothesis that regions of the cochlea separated by an octave do not significantly interact nonlinearly, Kim *et al.* (1997) examined $2f_1$-f_2 OAE generated by each of three stimulus tone pairs presented separately and presented simultaneously with stimulus levels of 65 and 50 dB SPL for each of the stimulus pairs, adjacent f_2's an octave apart. It was found that the effect of using three stimulus tone pairs to generate $2f_1$-f_2 OAE produced only a very small difference in $2f_1$-f_2 OAE level (less than 1.3 dB) with a tendency for the OAE to be lower in the three pair stimulus condition than for single stimulus pairs. The effect, while small, may be evidence of suppression effects and/or a catalyst mechanism.

More generally than the basal distortion product generation mechanisms proposed by Fahey *et al.* (2000), the spatial extent of generation of an OAE has not been widely examined, particularly at high stimulus levels. Withnell and Yates (1998) found the growth of the $2f_1$-f_2 OAE[2] at low to moderate stimulus levels to be analogous to the basilar membrane input-output function and so consistent with a locus of origin of f_2 for $f_2/f_1=1.6$. The spatial extent of generation of the $2f_1$-f_2 OAE is clearly dependent though on the stimulus frequency ratio (f_2/f_1). A number of studies have examined the locus of $2f_1$-f_2 OAE generation using iso-suppression tuning curves (e.g., Brown and Kemp, 1984; Kummer *et al.*, 1995; Abdala *et al.*, 1996) and shown it to be near f_2. Martin *et al.*, 1998 examined $2f_1$-f_2 OAE suppression tuning curves for $f_2/f_1=1.2$, showing the locus of generation in humans to be near, but apical, to f_2 for the $2f_1$-f_2 OAE. At high stimulus levels, broadened BM excitation patterns with basal-ward shifts in the maximum BM responses provide a larger region of nonlinear interaction between the BM excitation patterns arising from the two pure tone stimuli.

The spatial extent of generation of the $2f_1$-f_2 OAE can be examined by acoustically traumatizing the base of the cochlea and so eliminating the contribution of the basal region of the cochlea to the generation of $2f_1$-f_2. This experimental paradigm includes examination of a harmonic mechanism, this mechanism not requiring the presence of a third tone to be investigated in terms of its contribution to the generation of DPOAEs, acoustically traumatizing the base of the cochlea eliminating the contribution of a harmonic mechanism to the generation of $2f_1$-f_2. Without the addition of a third tone, it does not examine a catalyst mechanism. In this study, we examine evidence for a basal-to-f_2 distortion product generator in the guinea pig by comparing $2f_1$-f_2 OAE recorded before and after acoustically traumatizing the basal turn using a 12 kHz ~ 100–105 dB SPL tone. To explore the dependence of stimulus level on the spatial extent of generation of the $2f_1$-f_2 OAE, amplitude-modulated stimuli were used to generate an amplitude-modulated $2f_1$-f_2 OAE, i.e., an OAE that varies in amplitude over time in a stimulus-level dependent manner. Stimulus frequencies in the range f_2=8–8.94 kHz were used (f_2/f_1=1.2), a frequency range in the same region of the cochlea as that examined by Martin *et al.* (1999), who found $2f_1$-f_2 OAE suppression/enhancement effects ~ an octave above f_1 in guinea pig for f_2=6.2 kHz with f_2/f_1=1.2.

II. METHOD

A. Animal surgery

Albino guinea pigs (300–550 g) were anesthetized with Nembutal (35 mg/kg i.p.) and Atropine (0.06–0.09 mg i.p.), followed approximately 15 min later by Hypnorm (0.1–0.15 ml i.m.). Neuroleptanaethesia was maintained with supplemental doses of Nembutal and Hypnorm. Guinea pigs were tracheostomized and mechanically ventilated on Carbogen (5% CO_2 in O_2) with body rectal temperature maintained at approximately 38.5 °C. The head was positioned using a custom-made head holder that could be rotated for access to the ear canal. Heart rate was monitored throughout each experiment. The bulla was opened dorso laterally and a silver wire electrode placed on the round window niche for the recording and monitoring of the compound action potential (CAP). CAP thresholds were recorded between 4 and 20 kHz in 2 kHz steps throughout each experiment. Pancuronium (0.15 ml i.m.) may have been administered to reduce physiological noise associated with muscle contractions.

B. Signal generation and data acquisition

The method for stimulus delivery has been described previously (Withnell *et al.*, 1998; Withnell and Yates, 1998). Briefly, the acoustic stimuli were delivered by a Beyer DT48 loudspeaker placed approximately 4 cm from the entrance to the ear canal. Ear canal sound pressure recordings were made by a Sennheiser MKE 2-5 electrostatic microphone fitted with a metal probe tube (1.2 mm long, 1.3 mm i.d., 1500 Ω acoustic resistor) positioned approximately 2 mm into the ear canal. The microphone and probe tube combination was calibrated against a Bruel and Kjaer 1/8 in. microphone. The output from the probe tube microphone was amplified 20 dB, high-pass filtered (0.64 kHz, 4 pole Butterworth) and transmitted as a balanced input to one of the analogue input channels of a Card Deluxe sound card (Digital Audio Labs, www.digitalaudio.com) with an additional 4 dB of gain provided by the sound card. The signal was digitized at a rate of 32 000 Hz.

Signal generation and data acquisition was computer controlled using SYSRES (Neely and Stevenson, 2002). Data were recorded for a total of 65.536 s at each f_2 stimulus frequency. The stimulus complex consisted of two amplitude-modulated pure tone stimuli, digitally generated and output separately on two different output channels, mixed without amplification and buffered by a Tucker-Davis Technologies HB6 loudspeaker buffer-amplifier. Amplitude-modulated, or continuously varying in level, stimulus tones have been used previously to obtain DPOAE input-output functions (Neely *et al.*, 2005), a technique that produces input-output functions in a shorter time period than that obtained for measurements using discrete stimulus levels (Neely *et al.*, 2005) and also provides for OAE amplitude and phase in an L_1-L_2-OAE space (a three-dimensional plot of OAE amplitude or phase versus L_1 and L_2).

C. Procedure

In this study, amplitude-modulated stimuli were used to generate an amplitude-modulated $2f_1$-f_2 OAE, i.e., an OAE that varies in amplitude over time in a stimulus-level dependent manner. This provides for an OAE that can be examined in terms of a basal-to-f_2 distortion product generator by comparing $2f_1$-f_2 OAE recorded before and after acoustically traumatizing the basal turn where any stimulus-level dependent sensitivity in terms of OAE generation is implicit in the OAE. Stimulus-frequency ratio was held constant at f_2/f_1 =1.2 with f_2 fixed in value. Ten separate measurements were made for ten different values of f_2, f_2=8–8.94 kHz with a 94 Hz step size. Amplitude modulation of the two pure tone stimuli provided L_1 and L_2 varying from approximately 40 to 80 dB SPL. Measurement at each of the ten f_2 frequencies was performed twice prior to acoustically traumatizing the base of the cochlea and once after acoustically traumatizing the base of the cochlea.

D. Traumatizing tone

For all experiments reported in this paper, a 12 kHz, ~100–105 dB SPL tone generated by an Agilent 33120A signal generator connected to a Beyer DT48 loudspeaker via a Tucker-Davis Technologies HB6 loudspeaker buffer-amplifier was delivered to the ear for approximately 10 min to traumatize the basal region of the cochlea. The magnitude of the damage produced was quantified by the measurement of CAP thresholds before and after traumatizing the cochlea. CAP threshold was visually determined using the software equivalent of an averaging oscilloscope (number of responses averaged=40).

E. Data analysis

Data analysis was performed in MATLAB. The analysis to extract $2f_1$-f_2 from the ear canal recording was written in MATLAB (source code written by Dr. Stephen Neely), summarized as follows:

i. a discrete fast fourier transform (FFT) was performed on the data (2^{21} samples recorded at a sampling rate of 32 kHz) recorded at each f_2 frequency;
ii. the FFT was subsequently Blackman windowed centered at $2f_1$-f_2 and truncated to a size of 8192 bins centered at $2f_1$-f_2;
iii. an inverse fast Fourier transform (IFFT) was performed on the truncated, windowed FFT to obtain the signal complex amplitude waveform versus time. Truncation of the data set to 8192 bins produces an IFFT that has been down sampled to a sampling rate of 64 Hz;
iv. the IFFT waveform was divided into two equal lengths and then averaged;
v. the Hilbert envelope (the modulus of the IFFT) was then calculated; and
vi. the amplitude was then obtained from the Hilbert envelope according to the formula amplitude $=20^{*}\log[|\mathrm{IFFT}|/(2^{*}10^{-5})]$.

This analysis to extract $2f_1$-f_2 from the ear canal recording is identical to that reported in Neely *et al.* (2005). A microphone probe tube correction was applied after step (vi).

The $2f_1$-f_2 OAE recorded at each f_2 frequency was measured twice pre-trauma and once post-trauma. To examine quantitatively the effect of damage to the basal region of the cochlea on the amplitude of the $2f_1$-f_2 OAE obtained, a correlation coefficient was calculated for

i. the log of the Hilbert transform of the two pre-trauma $2f_1$-f_2 OAE measurements and
ii. the log of the Hilbert transform of the initial pre-trauma measurement versus the log of the Hilbert transform of the $2f_1$-f_2 OAE measurement recorded post-trauma to the base of the cochlea.

The two pre temporary threshold shift (TTS) recordings provided a lower boundary for the correlation coefficient when there has been no change in cochlear status. Measurement of the $2f_1$-f_2 OAE over an extended f_2 frequency range (8–8.94 kHz) provides additional data for comparison of pre versus post TTS measurements. Calculating the correlation coefficient for two wave forms in time incorporates any differences in frequency and/or phase of the Hilbert envelope but not simple scaling changes of the waveform. Alteration in the gain of the active process associated with damage to the outer hair cells (OHCs) produces a linearization of the growth of the $2f_1$-f_2 OAE (e.g., Neely *et al.*, 2003) and so would produce an alteration in the shape of the Hilbert envelope, the Hilbert envelope for an amplitude-modulated OAE varying in amplitude over time in a stimulus-level dependent manner. Therefore, comparison of the correlation coefficient of two OAE responses pre-acoustic trauma to the base of the cochlea versus the correlation coefficient pre-versus post-acoustic trauma should be sensitive to changes in the OAE associated with changes to the region/s of the cochlea responsible for generation of the $2f_1$-f_2 OAE. Expressing the Hilbert envelope of the OAE in logarithmic terms before calculating the correlation coefficient gives greater weighting to OAE obtained at lower stimulus levels or that part of the OAE that will be affected more significantly by changes to the gain of the active process (the log of the Hilbert envelope tends to weight the responses at all stimulus levels more equally).

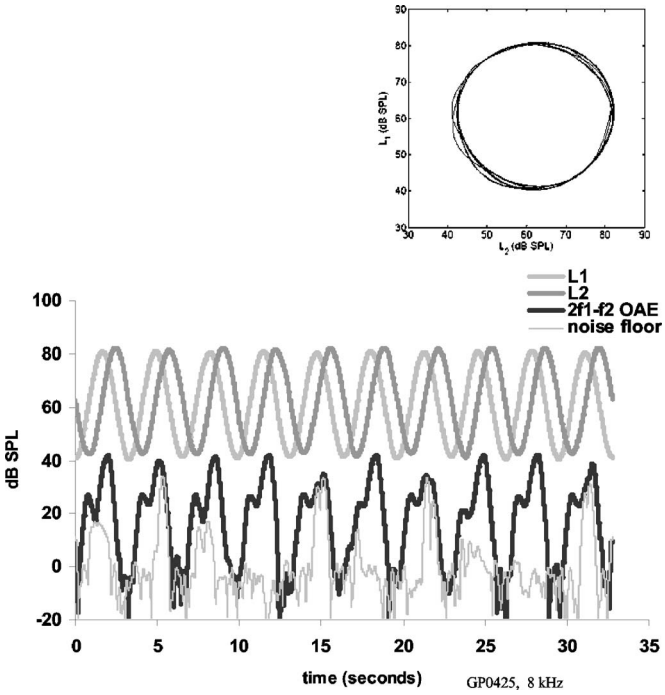


FIG. 3. The log of the Hilbert envelopes of two amplitude-modulated stimulus tones and the $2f_1$-f_2 OAE, obtained from analysis of the ear canal recording. Also shown in the top right corner is the stimulus input-output function.

III. RESULTS

Figure 3 shows an example of the log of the Hilbert envelope of the amplitude-modulated stimulus tones and the $2f_1$-f_2 OAE. Stimulus level in Fig. 3 ranges from ~40 to 80 dB SPL, producing an OAE that varies in amplitude over time in a stimulus-level dependent manner. A stimulus-level range of ~40–80 dB SPL is representative of all the stimuli used in this study. In this particular example, the two stimuli are amplitude modulated at the same rate (~0.3 Hz) with a 90° phase difference or 0.83 s difference between amplitude-modulated stimulus envelopes. An identical amplitude modulation rate with no phase difference is equivalent to co-varying stimulus intensity level. By introducing a 90° phase difference, an input-output function is obtained that has an L_1-L_2 space that is an ellipse (Neely *et al.*, 2005), i.e., for any given L_1 it does not have a wide range of L_2 values (see inset, Fig. 3) but nevertheless examines the $2f_1$-f_2 OAE over a range of stimulus levels. To obtain a $2f_1$-f_2 OAE that arises from a more complete two-dimensional stimulus level space requires that the amplitude modulation rate of the two stimuli differ (see Neely *et al.*,

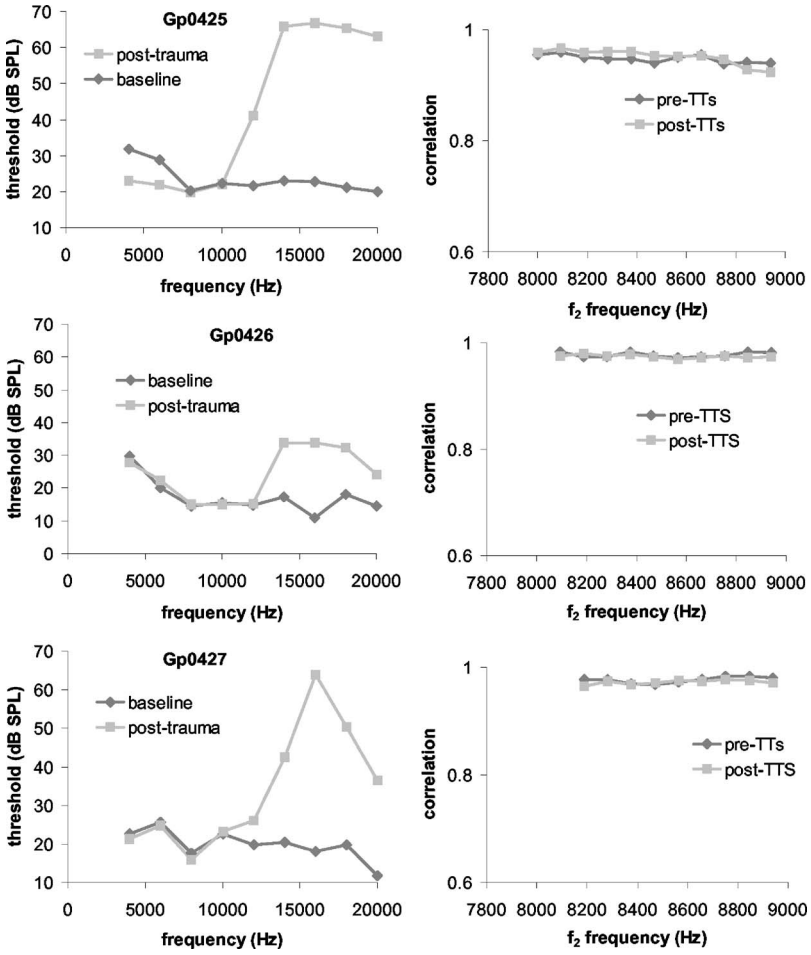


FIG. 4. CAP audiograms pre- vs post-trauma and the correlation of the OAE obtained in two pre-trauma measurements and pre- vs post-trauma when the change in CAP thresholds produced by the acoustically traumatizing tone does not extend into the region of the cochlea tuned below 10 kHz. No change is observed in the correlation coefficient of the two pre-TTS OAEs vs the pre- and post-TTS OAEs for all three animals.

2005). It is evident from Fig. 3 that the log of the Hilbert envelope of the $2f_1$-f_2 OAE varies reasonably systematically over time for an amplitude-modulated stimulus complex that repeats systematically as would be expected from a memoryless or limited memory system on the time scale we are observing the OAE; note that the recorded response was essentially the real time recording without averaging (number of averages=2), the $2f_1$-f_2 OAE extracted from the ear canal recording without the benefit of synchronous averaging by narrow-band windowing the response about the center frequency of interest using a Blackman window. The $2f_1$-f_2 OAE time waveform shows a double peak within each cycle of modulation of the stimuli, such an amplitude variation or notch being consistent with a vector cancellation of two OAE components (Brown, 1987; Whitehead *et al.*, 1992; Mills and Rubel, 1994) or shifts in outer hair cell operating point (Lukashkin *et al.*, 2002). Also shown in Fig. 3 is the noise floor which is calculated as the difference between the first and last half of the recordings (Sec. II E step iv of the Method section but taking the difference rather than the average); surges or peaks in the noise floor are evident coincident with peaks in the OAE response and presumably reflect incomplete cancellation of the OAE rather than noise, implying a recorded response that is not completely stationary over time, although there is no suggestion that such incomplete cancellation is systematic.

To examine the role that the basal region of the cochlea has on the generation of the $2f_1$-f_2 OAE, the base was damaged using an acoustically traumatizing 12 kHz, ~100–105 dB SPL presented for approximately 10 min, the 12 kHz tone presented with the intention of producing damage to the region of the cochlea basal to the initial nonlinear generation region of the $2f_1$-f_2 OAE. Figure 4 shows three examples of results obtained when the change in CAP thresholds produced by the acoustically traumatizing tone did not extend into the region of the cochlea tuned below 10 kHz, where the initial nonlinear (f_2) region of generation of the $2f_1$-f_2 OAE extended from 8 to 8.94 kHz. The left panels show CAP thresholds before and after acoustically traumatizing the base of the cochlea. The right panels show correlation coefficients obtained at each f_2 stimulus frequency before and after the TTS. CAP thresholds are unaltered up to 10 kHz but significant change in CAP thresholds above 10 kHz is apparent, including the 13.3–14.9 kHz region from whence a harmonic distortion mechanism would produce $2f_1$-f_2 as a quadratic distortion tone, i.e., where the second harmonic of the BM response to the f_1 stimulus, $2f_1$, interacts nonlinearly with the BM response to the f_2 stimulus to generate $2f_1$-f_2. No change is observed in the correlation of the two pre-TTS OAEs versus the pre- and post-TTS OAEs for all three animals, suggesting no change in $2f_1$-f_2 OAE amplitude for L_1 and L_2 varying from approximately 40–80 dB SPL, after damaging the basal region of the cochlea. This analysis does not support the region of the cochlea basal to the initial nonlinear generation region, f_2, contributing to the generation of the $2f_1$-f_2 OAE.

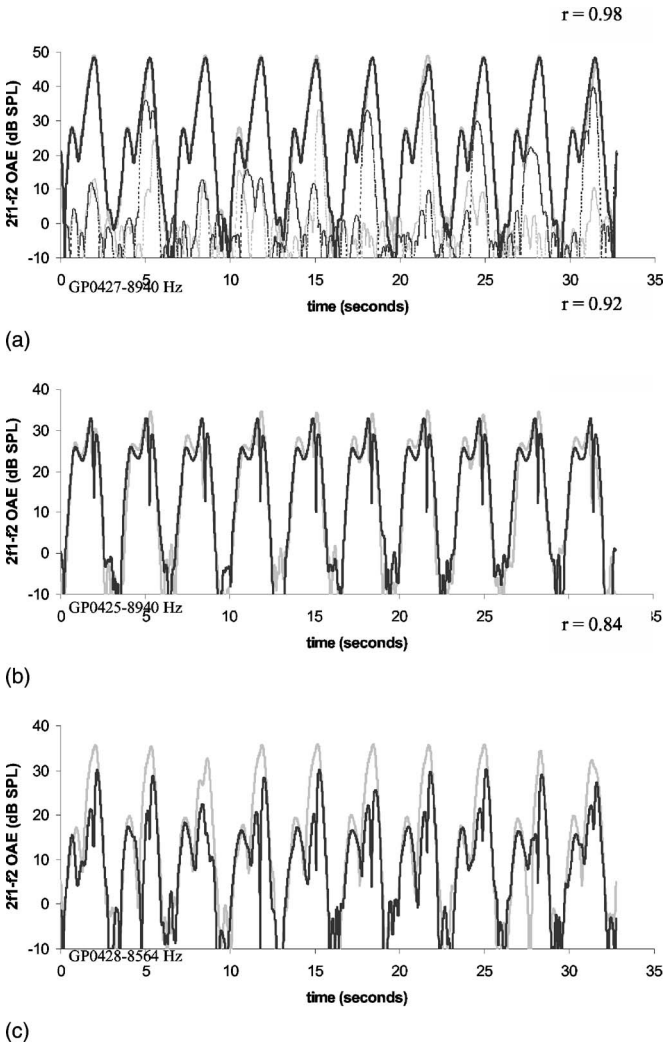


FIG. 5. Examples of $2f_1$-f_2 OAE Hilbert envelopes with correlation coefficients of 0.98, 0.92, and 0.84. The waveform envelopes in panel (a) for a correlation of 0.98 are basically identical, in panel (b) for a correlation of 0.92 the discrepancy is small and confined to the peak of the responses, and in panel (c) for a correlation of 0.84 the discrepancy is larger than in (b) although the wave forms retain somewhat similar shapes.

To illustrate the differences in waveform morphology that produce various correlation coefficients, Fig. 5 shows examples of $2f_1$-f_2 OAE Hilbert envelopes with correlation coefficients of 0.98, 0.92, and 0.84. The waveform envelopes in panel (a) for a correlation of 0.98 are basically identical, in panel (b) for a correlation of 0.92 the discrepancy is small and confined to the peak of the responses, and in panel (c) for a correlation of 0.84 the discrepancy is larger than in (b) although the general waveform envelope shape is retained.

Figure 6 shows two examples of results obtained when the change in CAP thresholds produced by the acoustically traumatizing tone *does* extend into the region of the cochlea tuned below 10 kHz. The left panels show CAP thresholds to have been affected within the f_2 region that generates $2f_1$-f_2 as a cubic distortion tone subsequent to exposure to the acoustically traumatizing tone. The OAE correlations corresponding to the left panels show changes in the OAE occurred post TTS, correlation coefficients reducing after acoustically induced damage to the cochlea.

Figure 4 illustrates that $2f_1$-f_2 OAEs before and after damage induced basal to the initial nonlinear generation region of the $2f_1$-f_2 OAE are highly correlated while Fig. 6 illustrates that when such damage extends to include the initial nonlinear generation region of the $2f_1$-f_2 OAE, changes to the OAE are quantifiable in terms of changes in waveform morphology expressed in terms of a correlation coefficient. But does this discount a significant source of $2f_1$-f_2 basal to f_2, either through a harmonic mechanism or simply a basalward shift in the region of generation, as might occur with increasing stimulus level, contributing to the generation of the $2f_1$-f_2 OAE? To explore this possibility, the $2f_1$-f_2 OAE was examined during that part of the time domain recording where $L_1 > 60$ dB SPL, based on the premise that L_1 stimulus level dependence is greater than L_2 stimulus level dependence because the amplitude of the $2f_1$-f_2 OAE is determined by the square of the BM response to the L_1 tone multiplied by the BM response to the L_2 tone (Withnell and Yates, 1998). Figure 3 seems to support this assumption,

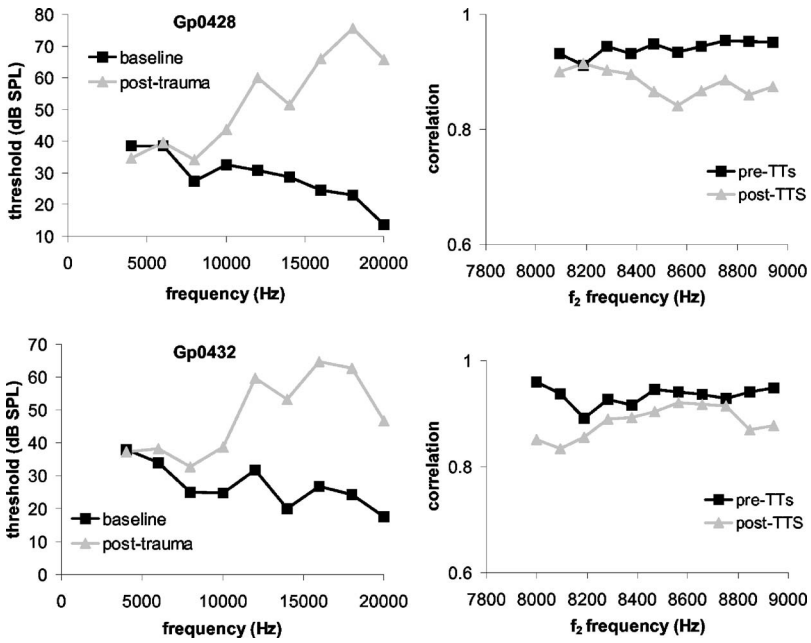


FIG. 6. CAP audiograms pre- vs post-trauma and the correlation of the OAE obtained in two pre-trauma measurements and pre- vs post-trauma when the change in CAP thresholds produced by the acoustically traumatizing tone *does* extend into the region of the cochlea tuned below 10 kHz. The OAE correlation coefficients show changes in the OAE occurred post TTS, correlation coefficients reducing after acoustically induced damage to the cochlea.

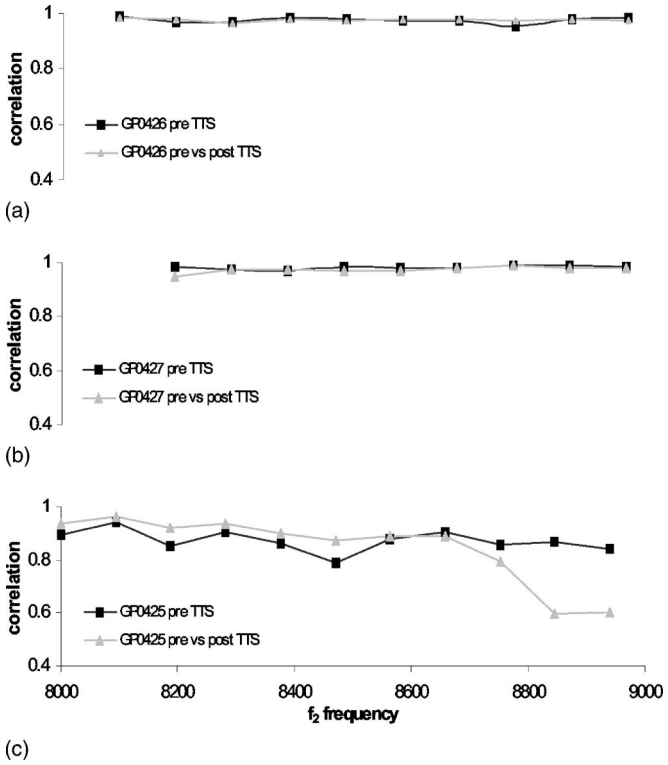


FIG. 7. A comparison of the correlation coefficients of the two pre-TTS OAEs vs the pre- and post-TTS OAEs obtained when $L_1 > 60$ dB SPL (from the data set reported in Fig. 4). It is evident that the correlation coefficients do not differ except above 8.7 kHz in panel (c), arguing against a significant source of $2f_1$-f_2 basal to f_2 contributing to the generation of the $2f_1$-f_2 OAE when $L_1 > 60$ dB SPL.

with OAE amplitude a maximum when $L_1 = L_2$[3] and then rolling off rapidly as L_1 decreases, even though L_2 continues to increase in amplitude for approximately 0.5 s after $L_1 = L_2$. Figure 7 compares the correlation coefficients of the two pre-TTS OAEs versus the pre- and post-TTS OAEs obtained when $L_1 > 60$ dB SPL from the data set reported in Fig. 4. It is evident that the correlation coefficients do not differ except above 8.7 kHz in panel (c), arguing against a significant source of $2f_1$-f_2 basal to f_2 contributing to the generation of the $2f_1$-f_2 OAE when $L_1 > 60$ dB SPL.

Measurement of the $2f_1$-f_2 OAE using amplitude-modulated stimuli provides for the generation of an L_1-L_2 cubic distortion tone (CDT) space input-output function that should be a very sensitive measure of the effects of induced cochlear damage. The data in Figs. 4 and 6 suggest that the amplitude of the $2f_1$-f_2 OAE is only altered when cochlear damage extends into the f_2 region that generates $2f_1$-f_2 as a cubic distortion tone. It does not support $2f_1$-f_2 being generated as a quadratic distortion tone at the cochlear $2f_1$ location or as a result of broadened stimulus-related BM excitation patterns with basal-ward shifts in the maximum BM responses extending the region of significant nonlinear interaction basal to f_2.

IV. DISCUSSION

The traditional notion for the generation of intermodulation distortion product OAEs, a two source interference model (e.g., Talmadge *et al.*, 1999; see Fig. 1), was brought into question with the findings of Martin *et al.* (1999) that showed, with the addition of a third (suppressor) tone, iso-suppression/enhancement contours with changes in the $2f_1$-f_2 OAE for a suppressor tone more than one octave above the stimulus tone frequencies. That is, the suppressor tone, at a frequency an octave above f_2, produced a significant reduction in $2f_1$-f_2 OAE amplitude. The findings of Martin *et al.* included results to an experiment analogous to that reported in this paper, i.e., the use of a traumatizing tone to damage the cochlea basal to the f_2 region in a rabbit [see Fig. 8 of Martin *et al.* (1999)); after exposure, the alteration to the $2f_1$-f_2 OAE amplitude observed with addition of the suppressor tone was removed. Martin *et al.* (1999) interpreted findings such as this to be consistent with the acoustically traumatizing tone having affected source/s of $2f_1$-f_2 OAE basal to f_2. Fahey *et al.* (2000) posited that one of two mechanisms could account for this phenomenon, a catalyst mechanism and/or a harmonic distortion mechanism. The harmonic mechanism, as described by Fahey *et al.* (2000), would not be expected to operate at intermediate and high frequencies where the high-frequency cutoff of the basilar membrane response is sharp, i.e., harmonics do not achieve sufficient amplitude at their own characteristic frequency place to generate sufficient nonlinear interaction quadratic distortion tone (QDT) to produce $2f_1$-f_2. Martin *et al.* (1999) used an f_1 of 2.53 kHz for this experiment, representing the mid or intermediate frequency region where the high-frequency cutoff is steep (Borg, 1988) and so a harmonic mechanism would not be expected to play a role as a basal to f_2 source of $2f_1$-f_2. In contrast, the catalyst mechanism, where $2f_1$-f_2 OAE arises out of a concatenation of nonlinear interaction of the three stimulus tone BM responses, their harmonics, and intermodulation distortion products is predicted to generate the response pattern observed (Fahey *et al.*, 2000). The absence in this paper of evidence for a harmonic mechanism in the generation of the $2f_1$-f_2 OAE for 8 kHz $\leqslant f_2 \leqslant$ 8.94 kHz (at least up to a stimulus level of 80 dB SPL) supports the prediction of Fahey *et al.* (2000).

A. Origin of the $2f_1$-f_2 OAE

Damage to the cochlea basal to the f_2 place does not alter the $2f_1$-f_2 OAE, a finding that establishes the initial generation site of the emission as being in the region of f_2 for the stimulus levels used. That is, while the region that generates the distortion is expected to be distributed over some length of the cochlea, it predominantly arises initially from the locus of f_2 for stimulus levels up to ~80 dB SPL. Acoustically traumatizing the base of the cochlea or preexisting basal damage provides an alternate means of examining the spatial extent of generation of the $2f_1$-f_2 OAE without the addition of a third (suppressor) tone that, in a nonlinear, active cochlea, may complicate examination of the origin of the $2f_1$-f_2 OAE.

The recent development of a scanning laser interferometer that provides for basilar membrane vibration to be measured as a function of longitudinal location (Ren, 2004) affords the possibility that the spatial extent of generation of $2f_1$-f_2 intermodulation distortion on the basilar membrane

can be examined directly in the future. It is evident though from this study that, in the absence of a third tone, the $2f_1$-f_2 OAE generated by a two-tone stimulus complex predominantly has an initial generation site solely in the region of f_2.

ACKNOWLEDGMENTS

We wish to thank Dr. Christopher Shera and Dr. Bill Shofner and two anonymous reviewers for providing valuable feedback and advice on earlier versions of this manuscript, Dr. Glen Martin and Dr. Barden Stagner for edifying discussions on basal sources of DPs, and Dr. Stephen Neely for generously providing the software to perform the experiments and extract the OAE responses from the ear canal recordings.

[1]Note that apical to the resonant place for $2f_1$, the BM response at the frequency $2f_1$ will be negligible, regardless of the stimulus level of f_1 (Lighthill, 1991).

[2]The experimental paradigm involved varying only L_2 with $f_2/f_1=1.6$

[3]It is interesting to note from Fig. 3 that the $2f_1$-f_2 OAE has a maximum amplitude when $L_1=L_2$ in guinea pig for $f_2/f_1=1.2$, in contrast to previous findings in humans and not supportive of the contention that the amplitude of the OAE is maximized when the BM responses to the two stimulus tones at the f_2 place are equal (when $L_2/L_1 \approx 10$ dB).

Abdala, C., Sininger, Y. S., Ekelid, M., and Zeng, F. G. (**1996**). "Distortion product otoacoustic emission suppression tuning curves in human adults and neonates," Hear. Res. **98**, 38–53.

Borg, E., Engstrom, B., Linde, G., and Marklund, K. (**1988**). "Eighth nerve fiber firing features in normal-hearing rabbits," Hear. Res. **36**, 191–202.

Brown, A. M. (**1987**). "Acoustic distortion from rodent ears: A comparison of responses from rats, guinea pigs and gerbils," Hear. Res. **31**, 25–38.

Brown, A. M., and Kemp, D. T. (**1984**). "Suppressibility of the $2f_1$-f_2 stimulated acoustic emissions in gerbil and man," Hear. Res. **13**, 29–37.

Evans, E. F. (**1972**). "The frequency response and other properties of single fibers in the guinea pig cochlear nerve," J. Physiol. (London) **226**, 263–287.

Fahey, P. F., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**2000**). "Nonlinear interactions that could explain distortion product interference response areas," J. Acoust. Soc. Am. **108**, 1786–1802.

Fahey, P. F., Stagner, B. B., and Martin, G. K. (**2006**). "Mechanism for bandpass frequency characteristic in distortion product otoacoustic emission generation," J. Acoust. Soc. Am. **119**, 991–996.

Gaskill, S. A., and Brown, A. M. (**1996**). "Suppression of human acoustic distortion product: Dual origin of $2f_1$-f_2," J. Acoust. Soc. Am. **100**, 3268–3274.

Johnstone, B. M., Patuzzi, R., and Yates, G. K. (**1986**). "Basilar membrane measurements and the travelling wave," Hear. Res. **22**, 147–153.

Heitmann, J., Waldmann, B., Schnitzler, H., Plinkert, P. K., and Zenner, H. (**1998**). "Suppression of distortion product otoacoustic emissions (DPOAE) near $2f_1$-f_2 removes DP-gram fine structure—evidence for a secondary generator," J. Acoust. Soc. Am. **103**, 1527–1531.

Kalluri, R., and Shera, C. A. (**2001**). "Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation," J. Acoust. Soc. Am. **109**, 622–637.

Kemp, D. T. (**1979**). "Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea," Arch. Otorhinolaryngol. **224**, 37–45.

Kemp, D. T. (**1986**). "Otoacoustic emissions, traveling waves and cochlear mechanisms," Hear. Res. **22**, 95–104.

Kim, D. O. (**1980**). "Cochlear mechanics: Implications of electrophysiological and acoustical observations," Hear. Res. **2**, 297–317.

Kim, D. O., Sun, X. M., Jung, M. D., and Leonard, G. (**1997**). "A new method of measuring distortion product otoacoustic emissions using multiple tone pairs: Study of human adults," Ear Hear. **18**, 277–285.

Kummer, P., Janssen, T., and Arnold, W. (**1995**). "Suppression tuning characteristics of the $2f_1$-f_2 distortion product emission in humans," J. Acoust. Soc. Am. **98**, 197–210.

Lighthill, J. (**1991**). "Biomechanics of hearing sensitivity," J. Vibr. Acoust. **113**, 1–13.

Lukashkin, A. N., Lukashkina, V. A., and Russell, I. J. (**2002**). "One source for distortion product otoacoustic emissions generated by low- and high-level primaries," J. Acoust. Soc. Am. **111**, 2740–2748.

Martin, G. K., Lonsbury-Martin, B. L., Probst, R., Scheinin, S. A., and Coats, A. C. (**1987**). "Acoustic distortion products in rabbit ear canal. II. Sites of origin revealed by suppression contours and pure-tone exposures," Hear. Res. **28**, 191–208.

Martin, G. K., Jassir, D., Stagner, B. B., Whitehead, M. L., and Lonsbury-Martin, B. L. (**1998**). "Locus of generation for the $2f1$-$f2$ vs $2f2$-$f1$ distortion-product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies, and amplitude correlations," J. Acoust. Soc. Am. **103**, 1957–1971.

Martin, G. K., Stagner, B. B., Jassir, D., Telischi, F. F., and Lonsbury-Martin, B. L. (**1999**). "Suppression and enhancement of distortion-product otoacoustic emissions by interference tones above f(2). I. Basic findings in rabbits," Hear. Res. **136**, 105–123.

Martin, G. K., Villasuso, E. I., Stagner, B. B., and Lonsbury-Martin, B. L. (**2003**). "Suppression and enhancement of distortion-product otoacoustic emissions by interference tones above $f(2)$. II. Findings in humans," Hear. Res. **177**, 111–122.

Mauermann, M., Uppenkamp, S., van Hengel, P. W., and Kollmeier, B. (**1999**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of the frequency ratio $f2/f1$," J. Acoust. Soc. Am. **106**, 3473–3483.

Mills, D. M., and Rubel, E. W. (**1994**). "Variation of distortion product otoacoustic emissions with furosemide injection," Hear. Res. **77**, 183–199.

Neely, S. T., and Stevenson, R. (**2002**). "SysRes," Tech. Memo. 19, Boys Town National Research Hospital, Omaha, NE.

Neely, S. T., Gorga, M. P., and Dorn, P. A. (**2003**). "Cochlear compression estimates from measurements of distortion-product otoacoustic emissions," J. Acoust. Soc. Am. **114**, 1499–1507.

Neely, S. T., Johnson, T. A., and Gorga, M. P. (**2005**) "Distortion-product otoacoustic emission measured with continuously varying stimulus level," J. Acoust. Soc. Am. **117**, 1248–1259.

Ren, T. (**2004**). "Reverse propagation of sound in the gerbil cochlea," Nat. Neurosci. **7**, 333–334.

Schneider, S., Prijs, V. F. *et al.* (**2003**). "Amplitude and phase of distortion product otoacoustic emissions in the guinea pig in an $(f1, f2)$ area study," J. Acoust. Soc. Am. **113**, 3285–3296.

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.

Talmadge, C. L., Long, G. R. *et al.* (**1999**). "Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **105**, 275–292.

Talmadge, C. L., Tubis, A. *et al.* (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.

Whitehead, M. L., Lonsbury-Martin, B. L., and Martin, G. K. (**1992**). "Evidence for two discrete sources of $2f_1$-f_2 distortion-product otoacoustic emission in rabbit. I. Differential dependence on stimulus parameters," J. Acoust. Soc. Am. **91**, 1587–1607.

Withnell, R. H., Kirk, D. L. *et al.* (**1998**). "Otoacoustic emissions measured with a physically open recording system," J. Acoust. Soc. Am. **104**, 350–355.

Withnell, R. H., Shaffer, L. A. *et al.* (**2003**). "Generation of DPOAEs in the guinea pig," Hear. Res. **178**, 106–117.

Withnell, R. H., and Yates, G. K. (**1998**). "Enhancement of the transient-evoked otoacoustic emission produced by the addition of a pure tone in the guinea pig," J. Acoust. Soc. Am. **104**, 344–349.

Withnell, R. H., and Yates, G. K. (**1998**). "Onset of basilar membrane nonlinearity reflected in cubic distortion tone input-output functions," Hear. Res. **123**, 87–96.

Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.

Effect of adaptive psychophysical procedure on loudness matches[a)]

Ikaro Silva[b)]
Institute for Hearing, Speech & Language, and Communications & Digital Signal Processing Center, Electrical & Computer Engineering Department (440 DA), Northeastern University, Boston, Massachusetts 02115

Mary Florentine
Institute for Hearing, Speech & Language, and Department of Speech-Language Pathology & Audiology (106/A FR), Northeastern University, Boston, Massachusetts 02115

(Received 31 January 2006; revised 8 May 2006; accepted 19 July 2006)

Large variability in equal-loudness matches has been observed across studies. The purpose of the present study was to gain insight into the extent to which this variability results from differences in psychophysical procedures and/or differences among listeners. Four adaptive two-interval, two-alternatives-forced-choice procedures were used to obtain equal-loudness matches between 5- and 200-ms 1-kHz tones as a function of level for each of six normal listeners. The procedures differed primarily in the sequence in which the stimuli were presented. The variations tested were the ordering of stimuli by amplitude across blocks of trials (both increasing and decreasing amplitudes), randomizing the order across those blocks, and randomizing the order within blocks. The random-within-block procedure, which sought to randomize any intertrial information, yielded a significantly greater amount of temporal integration than the other three procedures. The results show significant differences in temporal integration measurements at moderate levels for the same listeners across different procedures. Therefore, although there are individual differences among listeners in the amount of temporal integration measured across paradigms, the choice of paradigm also affects the amount of temporal integration measured at moderate levels. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2336747]

PACS number(s): 43.66.Cb, 43.66.Yw, 43.66.Mk [JHG] Pages: 2124–2131

I. INTRODUCTION

In order to understand how loudness is coded in the auditory system, measurements must be made across a wide range of levels. In most loudness experiments, including studies of temporal integration of loudness (and measurements of loudness functions), listeners are typically exposed to a wide range of stimulus levels within a short period of time. The procedural ordering of the presentation levels is variable across experimental paradigms. Experimenters can opt to present the levels of the stimuli in increasing order across blocks of trials (e.g., Moore and Glasberg, 1985), decreasing order across blocks of trials, random order across blocks of trials (e.g., Gockel *et al.*, 2003; Stevens and Hall, 1966; Algom *et al.*, 1989; Florentine *et al.*, 1996; Schneider and Parker, 1990; Lim *et al.*, 1977; Ellermeier *et al.*, 2001; McDermott *et al.*, 1998; Brand and Hohmann, 2002), or in some combination of the above (e.g., Moore, 2004; Buus *et al.*, 1997; McFadden, 1975). Whereas these studies give details on the sequence of stimuli presentations, other studies are not as clear about ordering (e.g., Rankovic *et al.*, 1988; Schlauch *et al.*, 1998; Small *et al.*, 1962; Moore *et al.*, 1999; Mapes-Riordan and Yost, 1999; Poulsen, 1981). In any case, most modern experiments present stimuli using a random across blocks (RAB) paradigm.

Recent data indicate that contextual effects, particularly related to range and ordering of presentations, may profoundly affect loudness measurements (Mapes-Riordan and Yost, 1999; Marks, 1994). The goal of the present study was to examine the effect of sequential ordering on a loudness experiment that is known to yield variable data, encompasses a large range of levels, and has been performed in multiple laboratories. Experiments measuring the temporal integration of loudness fulfill these requirements.

Numerous experiments examining the relationship between loudness and stimulus duration have shown that the auditory system performs temporal integration of loudness (for a review see Florentine *et al.*, 1996). For example, given two sounds with the same intensity and frequency, a 200-ms sound will be judged louder than a 5-ms sound. Temporal integration of loudness—measured as the difference in level required to make a short tone as loud as a long tone—is nonmonotonic with respect to level and is maximum at moderate levels (Pedersen *et al.*, 1977; Florentine *et al.*, 1996; Buus *et al.*, 1999). Florentine *et al.* (1996) compared data from many studies on temporal integration of loudness and observed a large variability in the maximum amount of temporal integration across the studies. At least part of this vari-

[a)]A portion of this work was presented at the midwinter research meeting of the Association for Research in Otolaryngology, 4–9 February 2006, Baltimore, MD.
[b)]Author to whom correspondence should be addressed; electronic mail: silva.i@neu.edu

ability appears to arise from differences in the psychophysical procedures employed (Stephens, 1974; Buus, 2002).

Researchers have long known that confounding variables can affect data obtained from equal-loudness matches; thus the design of a procedure to measure temporal integration of loudness using loudness matches can be a challenging process. For example, Stevens and Greenbaum (1966) found that if one of the stimuli is held fixed in an adjustment procedure, a so-called regression effect could occur due to the preferences that subjects have for listening to sounds at moderate loudness. The consequence of regression effects is that the variable sound will be "over-judged" near threshold values and "under-judged" at high loudness values. If the long sound is the one being varied, this effect will yield lower than normal temporal-integration values at threshold and higher than normal temporal-integration values at high sound levels. On the other hand, if the short sound varied, this effect will yield higher than normal temporal-integration values at threshold and lower than normal temporal-integration values at high sound levels. Florentine *et al.* (1996; 1998) also observed a regression effect in an adaptive two-interval, two-alternative forced-choice (2I, 2AFC) RAB procedure, originally developed by Jesteadt (1980), when measuring temporal integration by equating the loudness of two stimuli of different durations. Later, Buus *et al.* (1997) modified this procedure by combining features of Fletcher and Munson's (1933) forced-choice procedure with Jesteadt's (1980) adaptive procedure. This modification attempted to minimize regression bias toward comfortable loudness ranges, as well as any intertrial information that could bias loudness judgments by forcing the listener to attend only to the loudness of the stimulus. It allowed the level of both stimuli to vary randomly within a block (i.e., an RWB procedure). Listeners were presented with a random sequence of stimuli pairs so that they were unaware of which stimulus was being varied. Whereas the RWB procedure sought to minimize bias effects, it was recently suggested (Schlauch *et al.*, 1997; Buus *et al.*, 1999; Nieder *et al.*, 2003; Epstein and Florentine, 2005) that this procedure might have made listeners *more* susceptible to induced loudness reduction (also known as loudness recalibration), a decrease in the loudness of a sound due to a preceding louder sound.

Nieder *et al.* (2003) show that a preceding sound can reduce the loudness of a subsequent sound if its duration is at least as long as the duration of the subsequent sound. Studies of induced loudness reduction have also suggested that induced loudness reduction is likely to be a sensory effect as opposed to a decisional process (Arieh and Marks, 2001; Nieder *et al.*, 2003). These observations suggest that induced loudness reduction can have an asymmetric effect on loudness matches between tones in the same critical band. This effect is asymmetric with respect to duration in the sense that—given the same intensity—a long tone will influence the loudness of a short tone, but the short tone will not influence the loudness of the long tone. Since in the RWB procedure it is very likely for a higher-level stimulus with a long duration to be presented prior to a lower-level stimulus,

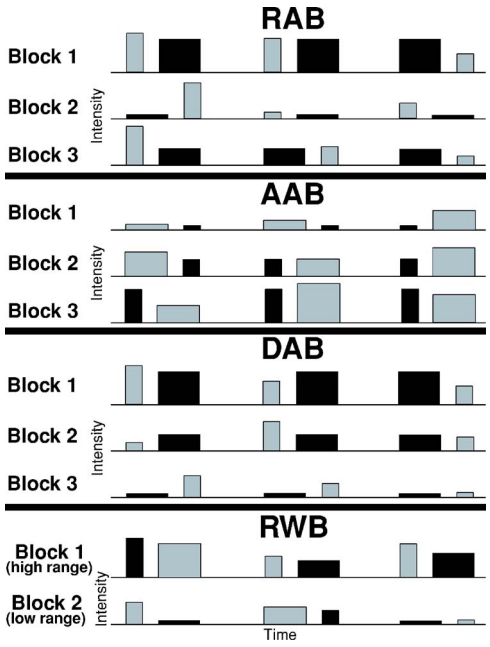


FIG. 1. Schematic diagram of the equal-loudness matching procedures. The random across blocks (RAB), ascending across blocks (AAB), and descending across blocks (DAB) procedures varied the level of the fixed tone (black) in a random, increasing, and decreasing order, respectively, across blocks of trials. The random within blocks (RWB) procedure consisted of only two blocks of trials, where the level of the fixed tone (black) varied randomly across a range of levels.

induced loudness reduction could have significantly increased the amount of temporal integration observed relative to data obtained with the RAB procedure.

In light of recent knowledge regarding induced loudness reduction, an investigation of the effects of procedural factors is necessary because they are likely to explain some of the variability observed in temporal integration measurements. Specifically, this study measures the amounts of temporal integration of loudness within the same set of listeners under four different paradigms that vary principally by sequential ordering of the stimuli. The goal is to determine if sequential order can account for a substantial portion of the variability across psychoacoustic studies of loudness that include a wide range of levels. Although this study was performed using adaptive 2I, 2AFC procedures, because induced loudness reduction can last several seconds and is generated whenever a louder sound precedes another sound at a moderate level and at a near frequency, this effect could be present in any loudness procedure where the intensity of the stimuli varies widely across trials (Arieh *et al.*, 2005; Epstein and Gifford, 2006).

II. METHOD

A schematic of the four different procedures is shown in Fig. 1. The first procedure, shown at the top of the figure, varied the level of the fixed tone RAB of measurements. The original RAB procedure and two variations of it were used. One variation, ascending across blocks (AAB), attempted to minimize induced loudness reduction by presenting the levels of the fixed tones in increasing order across blocks. The other variation, decreasing across blocks (DAB), attempted

to maximize exposure to induced loudness reduction effects by presenting the levels of the fixed tone in descending order across blocks. Unlike the RAB procedure, the procedure of Buus *et al.* (1997; 1999) varied the level and the fixed stimulus duration RWB. The RWB procedure was performed using two independent blocks of trials, one using a set of low levels and the other a set of high levels. Both the RAB and the RWB procedures have been used in at least two different laboratories and yielded similar results (e.g., the Acoustics Laboratory at the Technical University of Denmark and the Communication Research Laboratory at Northeastern University in Boston, see Buus *et al.*, 1997 and 1999).

A. Stimuli

The stimuli were 1-kHz tones with equivalent rectangular durations of 5 and 200 ms. The frequency and durations of the tones were chosen in order to allow comparisons with previously published data. All stimuli had 6.67-ms raised-cosine rise and fall times. Durations measured between the half-amplitude points are 1.67 ms longer than the nominal durations. Thus, the 5-ms stimuli consisted only of the rise and fall, whereas the 200-ms stimuli had a 195-ms steady-state portion. This raised-cosine window ensured that most of the energy was contained within the 160-Hz-wide critical band centered at 1 kHz (Scharf, 1970; Zwicker and Fastl, 1990). Even for the 5-ms tone burst, the energy within the critical band was only 0.3 dB less than the overall energy.

B. Procedure

1. Absolute thresholds

In the first part of the experiment, absolute thresholds for the 5- and 200-ms tones were measured using an adaptive 2I, 2AFC procedure, which was the same as used by Florentine *et al.* (1996) and Buus *et al.* (1999). Each trial contained two observation intervals marked with lights. The pause between the intervals was 500 ms. The signal was presented in either the first or the second observation interval with equal *a priori* probability. The listener's task was to indicate which interval contained the signal by pressing a key on a small computer terminal. Two hundred milliseconds after the listener responded, a 200-ms light indicated the correct answer. Following this feedback, the next trial began after a 500-ms delay.

The level of the signal decreased following three correct responses and increased following one incorrect response. The initial step size was 5 dB until the second reversal, after which it was reduced to 2 dB. Reversals occurred when the signal level changed from increasing to decreasing, or vice versa. This procedure converges on the signal level yielding 79.4% correct responses (Levitt, 1971).

A single threshold measurement was based on three interleaved adaptive tracks. For each trial the track was selected at random from the tracks that had not yet terminated. Termination of tracks occurred after five reversals. The threshold for one track was calculated as the average signal level at the fourth and fifth reversals; a final threshold measurement was taken as the average threshold across the three tracks. At least three such threshold measurements were obtained for each listener and condition (for a total of nine tracks per listener). The average across all measurements was used as the reference point for the sensation level for each listener.

2. Loudness matches

a. Common to all procedures. Loudness matches between 5- and 200-ms tones were obtained with an adaptive procedure in a 2I, 2AFC paradigm. On each trial, the listener heard two tones separated by 500 ms. The fixed tone and the variable tone had equal probability of being presented first, in order to avoid sequential judgment errors. The listener indicated which sound was louder by pressing a key on a computer terminal. The response initiated the next trial after 900 ms for the across-block procedures (RAB, AAB, and DAB), and after 700 ms for the RWB procedure (as used in Buus *et al.*, 1999). The level of the variable sound was changed according to a simple up-down rule. If the listener indicated that the variable sound was louder, its level was reduced; otherwise, it was increased. The step size was 5 dB until the second reversal, after which it was decreased to 2 dB. The track terminated after nine reversals. The equal-loudness level for one track was calculated as the average of the last four reversals. This procedure converges at the level corresponding to the 50% point on the psychometric function (Levitt, 1971). Nine loudness matches were obtained for the fixed tone at nine different levels, ranging from 10 dB SL to 90 dB SL in 10 dB steps. The stimuli were presented at equal SL rather than at equal SPL in order to reduce the variability among listeners' loudness judgments (Hellman and Zwislocki, 1961). Measurements at all nine levels were repeated four times for each listener.

During an experiment session day, listeners were tested first with the AAB procedure, which was only run once per day. The other procedures (RWB, DAB, and RAB) were alternated in a random order. Listeners were rarely tested for more than 2 h on a given day and they were required to take breaks to prevent fatigue.

b. Loudness matches with level of the fixed tone varying across blocks (RAB, AAB, DAB). In the across-blocks procedures, the loudness match at each level was the average of two interleaved tracks. One track started with the variable tone set to 10 dB below an estimated loudness match for that level, and the other track started with the variable tone set to 10 dB above. The stimuli for a given trial in the block were selected randomly from one of these two tracks. Each block yielded one loudness match and lasted for about 2 min.

In the RAB procedure, the level of the fixed tone was varied randomly across blocks (Fig. 1). In the AAB procedure, the level of the fixed tone increased by 10 dB across blocks. For the DAB procedure, the level of the fixed tone decreased by 10 dB across blocks. For all of the across-block procedures, the fixed tone had the same duration across all of the blocks (i.e., within a session).

c. Loudness matches with level of the fixed tone varying randomly within blocks (RWB). The RWB procedure was divided into two blocks: a low-level range block and a high-level range block as shown in Fig. 1. This was done to prevent listener fatigue caused by very long blocks of trials. The low-level block ranged in 10 dB steps from 10 to 50 dB SL. The high-level block ranged in 10 dB steps from 50 to 90 dB SL. There were ten interleaved tracks per block (2 fixed duration $\times$ 5 SLs), and the stimuli at each trial were picked at random from any of the ten tracks that had not yet

been completed. In this procedure, the fixed tones of both durations were presented within a single block randomly using roving-level adaptive procedure in a 2I, 2AFC paradigm. Within a block, the level of the fixed tone was picked at random from five possible values for each trial, which forced the listeners to base their responses only on the loudness of the two sounds presented in a trial. Each block, which lasted about 10 min, yielded five loudness matches for the low range block and five loudness matches for the high range block. For each fixed level, the variable level started 15 dB below the expected value (but not below threshold). This was done to ensure that the listener would hear trials in which the short or the long tone was clearly the louder one (since both tones varied randomly within a block). Also, the initial trial would not be too loud.

C. Apparatus

A PC-compatible computer with a signal processor (TDT AP2) generated the stimuli, sampled the listener's response, and executed the adaptive procedures. The tone bursts were generated digitally with a 50-kHz sample rate and reproduced by a 16-bit digital-to-analog converter (TDT DD1). The output from the D/A was attenuated (TDT PA4), low-pass filtered (TDT FT5, fc=20 kHz, 190 dB/octave), attenuated again (TDT PA4), and sent to a headphone amplifier (TDT HB6), which fed one earpiece of a Sony MDR-V6 headset. The listeners were seated in a sound-attenuating booth.

For routine calibration, the output of the headphone amplifier was fed to a 16-bit A/D converter; the computer sampled the wave form, calculated its spectrum and rms voltage, and displayed the results before each run.

D. Listeners

Six listeners, three females and three males, participated in all parts of the experiment. All listeners had normal hearing and no history of hearing difficulties. Their audiometric thresholds were within 10 dB HL in the test ear, except for listeners L3 and L4, who had thresholds of 15 dB HL at 250 Hz in the test ear (ANSI, 1996). All listeners were paid and they ranged in age from 19 to 26 years.

E. Data analysis

To examine the statistical significance of the difference between the procedures used to make loudness matches, a three-way analysis of variance (ANOVA) (Fixed Tone Duration/Sensation Level × Procedure × Listener) was performed (DATA DESK 6.0). The amount of temporal integration was the dependent variable. The factors and their levels were: Fixed Tone Duration and Sensation Level (D/S, two durations × nine sensation levels = 18 levels), Procedure (Prc, four levels), and Listener (Ls, six levels). The factor Ls was set as a random variable and the other two factors were discrete. Post-hoc Bonferroni analysis was performed on the Prc X D/S interaction in order to further investigate sources of significant interactions. The high- and low-range blocks of the RWB procedure were merged together yielding nine SLs with twice as many points for the 50 SL levels.

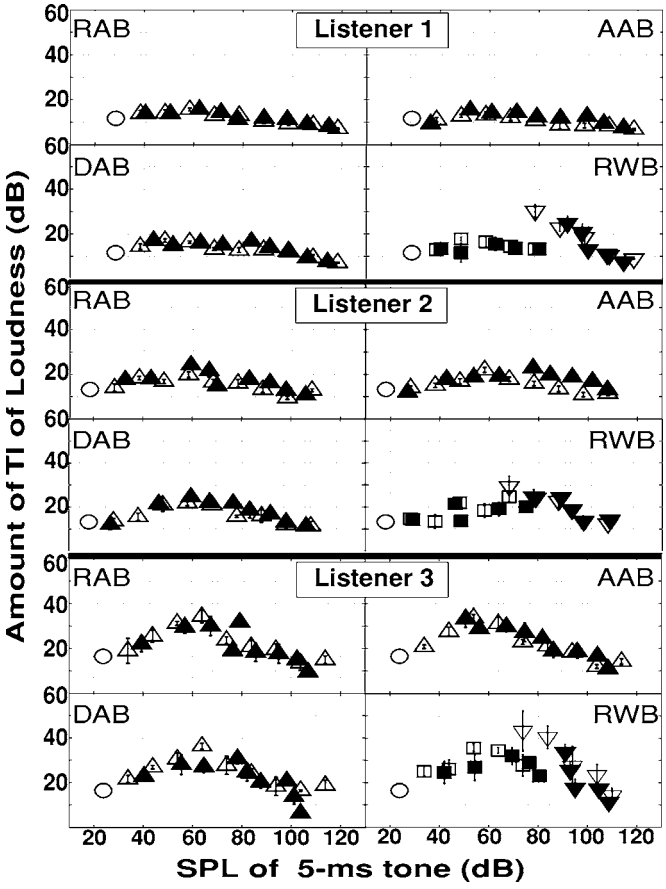


FIG. 2. Amount of temporal integration vs SPL of the 5-ms tone for the first three listeners and all four procedures. The open ovals represent the listeners' thresholds. Closed symbols correspond to when the long tone was held fixed. Open symbols correspond to when the short tone was held fixed. The rectangles correspond to measurements obtained at the low-level range of the fixed tone. The inverted triangles correspond to measurements obtained at the high-level range of the fixed tone. The error bars represent plus and minus one standard error.

III. RESULTS

The loudness matches from the six individual listeners for each of the four procedures are shown in Figs. 2 and 3. All the listeners showed temporal integration of loudness functions with maxima at moderate levels. However, the amount of temporal integration varies greatly across individuals consistent with data in the literature (Florentine *et al.*, 1988; Verhey and Kollmeier, 2002; Epstein and Florentine, 2005; see Florentine *et al.*, 1996 for a review). Listener 1 shows the least amount of temporal integration and Listener 5 shows the most. The temporal integration functions are more similar across listeners than across procedures.

The RAB procedure resulted in values of temporal integration ranging from 16.0 dB (L1) to 48.1 dB (L5). This range is wider than the 12- to 24-dB range of maximum temporal integration reported by Florentine *et al.* (1996) and the 20- to 33-dB range reported by Florentine *et al.* (1998) using the same procedure.

Figure 4 shows the average amount of temporal integration across all listeners when the short and long tones were held fixed (error bars represent one standard error). When the 5-ms tone was held fixed, the RWB showed significantly

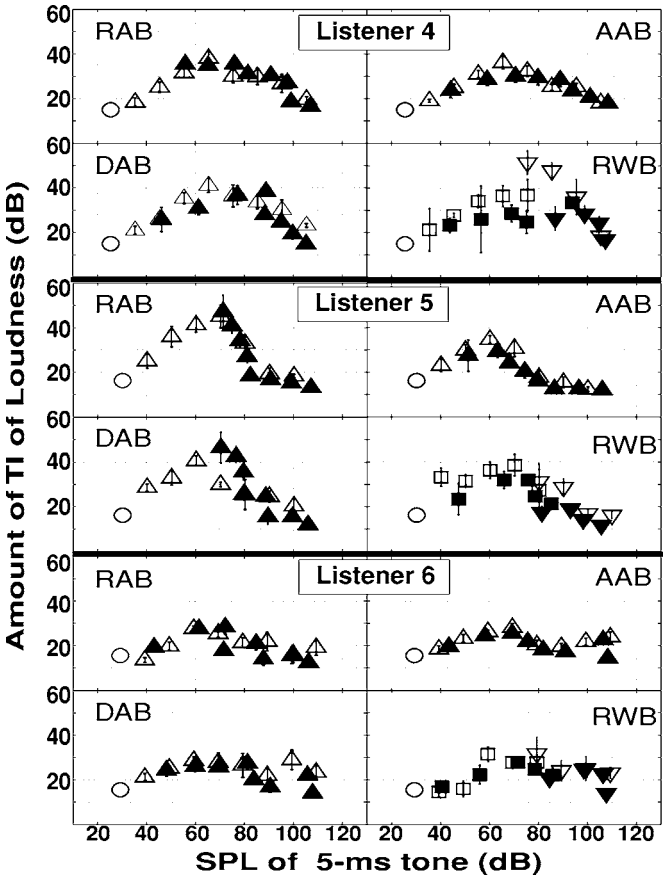


FIG. 3. Amount of temporal integration versus SPL of the 5-ms tone for the last three listeners obtained using all four procedures plotted in the same manner as Fig. 2

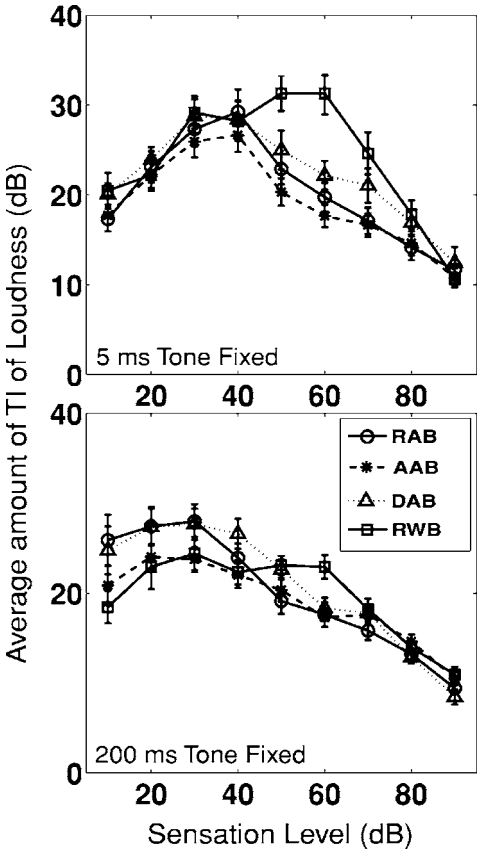


FIG. 4. Average amount of temporal integration of loudness vs Sensation Level across all six listeners for all procedures. The error bars represent plus and minus one standard error.

higher amounts of temporal integration at moderate levels (50, 60, and 70 dB). As shown in Fig. 4 and Table I, the average maximum amount of temporal integration for all listeners for the RWB procedure was 37.3 dB, the highest of all four procedures, compared to 32.1, 28.1, and 32.7 for the RAB, AAB, and DAB procedures, respectively. The average maximum temporal integration using the RWB procedure is 10.3 dB higher than the 27-dB average maximum reported by Buus *et al.* (1999).

The standard errors also changed with procedure. Table I summarizes the standard errors and temporal integration range for each of the four procedures. The average standard error within listeners for the RAB procedure was 2.2 dB, with a range of 0.2–7.2 dB. Unfortunately, few studies report average standard errors for individual listeners using similar procedures. However, the present data can be compared to a couple of existing studies. For example, the average standard error within listeners for the RAB procedure is in reasonable agreement with the results of Florentine *et al.* (1996), who reported an average standard error of 1.3 dB with a range of 0.1–5.3 dB. They are also in good agreement with the Florentine *et al.* (1998) study, who used the same procedure and obtained an average standard error of 1.4 dB with a range of 0.1–4.5 dB. The average standard error for the AAB and DAB procedures was 1.6 and 2.2, respectively. The data for the RWB procedure had an average standard error of 3.2 dB with a range of 0.1–14.9 dB (the highest error and the largest range of all the four procedures). The average standard error for the RWB data is larger than the error published by Buus *et al.* (1999), which was 2.3 dB. Thus, the RWB was the procedure with the highest average standard error of all four procedures.

TABLE I. Standard error and maximum temporal integration statistics for the data from all four procedures.

Procedure	Average std error (dB)	Std error range (dB)	Avg max temporal integration (dB)	Max temporal integration range (dB)
Random Across Blocks (RAB)	2.2	0.2–7.2	32.1	16.0–48.1
Ascending Across Blocks (AAB)	1.6	0.2–7.0	28.1	15.0–35.4
Descending Across Blocks (DAB)	2.2	0.3–6.8	32.7	18.1–44.2
Random Within Blocks (RWB)	3.2	0.1–14.9	37.3	29.3–46.0

The results from the three-way ANOVA showed that the main effect, duration/sensation level of the fixed stimulus, had a significant effect [$F(17,1239)=6.66$, $p<0.01$] on the amount of temporal integration as expected. The interaction between the fixed tone duration/sensation level factor and the procedure factor showed a significant effect [$F(51,1239)=5.13$, $p<0.01$]. Post-hoc Bonferroni analysis indicated that the RWB procedure yielded statistically different amounts of temporal integration from the other three procedures when the short tone was held fixed at 50 and 60 dB SL ($p<0.01$). The RWB procedure also yielded different results from the AAB and RAB procedures when the short tone was held fixed at 70 dB SL ($p<0.01$). When the long tone was held fixed, the only statistical difference occurred at 10 dB SL, between the RWB and RAB procedures ($p=0.04$) with the RAB procedure yielding a higher amount of temporal integration. No significant differences were found among the RAB, AAB, and DAB procedures at any sensation level and fixed tone duration.

IV. DISCUSSION

The present data clearly show a range of consistent, individual differences in equal-loudness matches. Whereas significant differences in temporal integration measurements across listeners using the same procedure have been found (for a review, see Florentine *et al.*, 1998), results from the current experiment also show *significant differences in temporal integration measurements at moderate levels for the same listeners across different procedures*. A possible explanation for obtaining different amounts of temporal integration with different procedures may be induced loudness reduction, as described in Sec. I. It has been suggested that induced loudness reduction can produce a significant order and range effect on the RWB procedure (Nieder *et al.*, 2003). For example, induced loudness reduction can produce significant effects when the inducer precedes the target by at least 200 ms (Arieh *et al.*, 2003a). This effect is maximal when the SPL of the inducer is 60–80 dB SPL, with the inducer level being about 10–20 dB higher than the target level (Mapes-Riordan and Yost, 1999). In addition, recovery from induced loudness reduction is at least on the order of dozens of seconds (Arieh *et al.*, 2005; Epstein and Gifford, 2006). Furthermore, the magnitude of the induced loudness reduction effect depends on the number of previous exposures to inducers (Arieh *et al.*, 2005). If induced loudness reduction has a large effect on the RWB procedure, the difference between the equally loud long and short tones is greater than in the other across-block procedures (i.e., the induced loudness reduction effect is asymmetric with respect to duration). For tones at the same or close SL, a long tone will generally sound louder than a short tone. Since the stimuli intensities were varied at fixed 10-dB-SL steps, the set of tones louder than a specific short tone will include short and long tones at higher intensities as well as long tones at equal or even slightly lower intensities. On the other hand, the set of tones louder than a specific long tone will include only the long and short tones at higher intensities. Thus, the loudness of the short tones will suffer more induced loudness reduction because of a larger set of louder stimuli prior to its presentation. The size and the intensity range of these sets are particularly high for the RWB procedure. Thus, equal-loudness matches for this procedure would show more temporal integration than the other adaptive 2I, 2AFC procedures (i.e., RAB, AAB, and DAB). In fact, differences of up to 14.3 dB were observed between the average maximum of the RWB and the AAB procedures for one listener (see L1, Fig. 2). Although effects such as regression biases could also have contributed to these differences, it is doubtful that regression biases can account for all of these results because regression biases play a minor role at moderate levels (Stevens and Greenbaum, 1966).

In general, the average temporal integration curves obtained when the short tone was held fixed (Fig. 4) are arranged in the following order from lowest to highest: AAB, RAB, DAB, and RWB. Whereas no statistical differences other than the ones aforementioned were found between the procedures, the ordering of these procedures is in agreement with the expected amount of induced loudness reduction that each procedure would generate. The ascending procedure (AAB) should produce the least amount of induced loudness reduction because the softest blocks always preceded the louder ones, preventing the loud sounds from influencing the judgment of the soft blocks. The RAB measurements would be expected to exhibit more ILR than the AAB measurements, but still less than the other procedures because the frequency of loud blocks preceding soft blocks is not as high as in the DAB and RWB procedures. The descending procedure, DAB, should have the largest amount of induced loudness reduction for the across-block procedures (RAB, AAB, and DAB) because exactly the opposite of the ascending procedure (AAB) occurs; in the descending procedure all soft blocks are preceded by louder ones. The high range of the RWB curve should exhibit the most induced loudness reduction because there was no rest period (i.e., half of the trials were presented in a single block). In addition, at such high levels (>80 dB SPL), most trials will cause induced loudness reduction on the moderate-level stimuli (Nieder *et al.*, 2003). At low levels, induced loudness reduction is very unlikely to produce an effect (Botte *et al.*, 1982). However, regression biases could be responsible for the statistical difference between the RWB and the RAB procedure when the long tone was held fixed at 10 SL. From the average curves (Fig. 4), it is apparent that the RWB procedure yields a higher amount of temporal integration than the other procedures when the short tone is held fixed at moderate levels. The maximum of the RWB curve is also shifted to a higher SPL than on the other procedures.

The standard error for the RWB is the largest of all the procedures, which probably results from sequential and range effects. The temporal integration results obtained with the RWB procedure were 10.3 dB higher than those reported by Buus *et al.* (1999), which were also obtained using the RWB procedure. A possible explanation for this is that their procedure was divided into three blocks (low, moderate, and high level ranges), whereas the present RWB procedure was divided into only two blocks (low and high level ranges). Dividing the RWB procedure into three blocks makes the

moderate-level trials less likely to be preceded by intense trials. This decreases the likelihood of the loudness of the short tone being profoundly influenced by other trials, as it probably was in the RWB procedure used in the present study.

The repeated temporal integration measurements obtained at 50 dB SL (i.e., low range versus high range) using the RWB procedure also can be used to gain insight into range effects. The differences between the means at the overlapping points ranged from 3.9 (L5) to 16.9 dB (L1) with an average of 9.9 dB across all listeners. These ranges are also within the range of induced loudness reduction reported by Nieder *et al.* (2003) and Arieh *et al.* (2003b). Nieder *et al.* reported a range from −3 to 20 dB for a 200-ms, 70-dB test tone and 80-dB inducer.

In light of the present data, two implications for future studies are suggested. First, because the present experiment showed significant differences in the equal-loudness matches across the procedures, it is important to clearly describe the sequence of stimulus presentation in studies that use a range of levels and/or durations. Second, it may not be wise to employ the RWB procedure for equal-loudness matches studies until the reasons for the differences between it and the other procedures are better understood. If large differences in stimulus levels across trials is responsible for the discrepancies between RWB and the other three procedures, then increasing the number of blocks in the RWB procedure might result in better agreement (i.e., smaller amounts of temporal integration at moderate levels). In fact, the RWB procedure used by Buus *et al.* (1999) used three blocks and showed somewhat less temporal integration than the current procedure.

V. SUMMARY

The present study compared four different psychophysical procedures for measuring equal-loudness matches between two 1-kHz tones of two durations in each of six listeners. Three of the procedures (AAB, RAB, and DAB) change the fixed tone level across blocks of trials. The RWB is a modified procedure that attempted to reduce inter-trial information by a random presentation of stimuli. The results show:

(1) The equal-loudness matches for the six listeners encompass a wide range of responses, consistent with data from previous studies.
(2) No significant differences were found among the across-block procedures (AAB, DAB, and RAB).
(3) Average data for most listeners show that the RWB data yielded larger amounts of temporal integration than the other three procedures at moderate SL when the short tone was held fixed. Specifically, ANOVA and post hoc Bonferroni analysis show that there is a significant difference between the amount of temporal integration obtained using the RWB procedure and the other three procedures at moderate levels (50–60 dB SL) when the short tone is held fixed. Because this effect was significant at moderate levels, its source cannot be attributed solely to regression biases. A statistical difference between RWB and the RAB procedure was also observed at 10 SL when the long tone was held fixed. It is possible that regression biases could have contributed to this result.
(4) When comparing equal-loudness matches from a wide range of levels across studies in the literature, it is important to consider the stimulus ordering that has been used.

ACKNOWLEDGMENTS

Søren Buus contributed substantially to this project and passed away prior to the completion of this work. The following people were of great help in revising the manuscript: Michael Epstein, Bert Scharf, Eva Wagner, and Jeremy Marozeau. The authors would like to thank the editor and the two anonymous reviewers for their comments. This work was supported by NIH/NIDCD Grant No. R01DC02241.

Algom, D., Rubin, A., and Cohen-Raz, L. (**1989**). "Binaural and temporal integration of the loudness of tones and noises," Percept. Psychophys. **46**(2), 155–166.
ANSI (**1996**). ANSI-S3.6, 1996, "Specifications for audiometers," American National Standards Institute, New York.
Arieh, Y., Kelly, K., and Marks, L. E. (**2005**). "Tracking the time recovery after induced loudness reduction (L)," J. Acoust. Soc. Am. **117**(6), 3381–3384.
Arieh, Y., and Marks, L. E. (**2001**). "Recalibration of loudness: sensory vs. decisional processes," in *Fechner Day 2001*, edited by E. Sommerfeld, R. Kompass, and T. Lachmann (Pabst, Berlin).
Arieh, Y., and Marks, L. E. (**2003a**). "Recalibrating the auditory system: A speed-accuracy analysis of intensity perception," J. Exp. Psychol. Hum. Percept. Perform. **29**, 523–536.
Arieh, Y., and Marks, L. E. (**2003b**). "Time course of loudness recalibration: Implications for loudness enhancement," J. Acoust. Soc. Am. **114**, 1550–1556.
Botte, M.-C., Canévet, G., and Scharf, B. (**1982**). "Loudness adaptation induced by an intermittent tone," J. Acoust. Soc. Am. **72**, 727–739.
Brand, T., and Hohmann, V. (**2002**). "An adaptive procedure for categorical loudness scaling," J. Acoust. Soc. Am. **112**(4), 1597–1604.
Buus, S. (**2002**). "Psychophysical methods and other factors that affect the outcome of psychoacoustic measurements," in *Genetics and the Function of the Auditory System*, edited by L. Tranebjærg, J. Christensen-Dalsgaard, T. Andersen, and T. Poulsen (GN ReSound, Tåstrup, Denmark).
Buus, S., Florentine, M., and Poulsen, T. (**1997**). "Temporal integration of loudness, loudness discrimination, and the form of the loudness function," J. Acoust. Soc. Am. **101**, 669–680.
Buus, S., Florentine, M., and Poulsen, T. (**1999**). "Temporal integration of loudness in listeners with hearing losses of primarily cochlear origin," J. Acoust. Soc. Am. **105**, 3464–3480.
Ellermeier, W., Eigenstetter, M., and Zimmer, K. (**2001**). "Psychoacoustic correlates of individual noise sensitivity," J. Acoust. Soc. Am. **109**(4), 1464–1473.
Epstein, M., and Florentine, M. (**2005**). "A test of the Equal-Loudness-Ratio hypothesis using cross-modality matching functions," J. Acoust. Soc. Am. **118**(2), 907–913.
Epstein, M., and Gifford, E. (**2006**). "A confounding factor in the measurement of induced loudness reduction," J. Acoust. Soc. Am. **120**, 305–309.
Fletcher, H., and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," J. Acoust. Soc. Am. **5**, 82–108.
Florentine, M., Buus, S., and Poulsen, T. (**1996**). "Temporal integration of loudness as a function of level," J. Acoust. Soc. Am. **99**, 1633–1644.
Florentine, M., Buus, S., and Robinson, M. (**1998**). "Temporal integration of loudness under partial masking," J. Acoust. Soc. Am. **104**, 999–1007.
Florentine, M., Fastl, H., and Buus, S. (**1988**). "Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking," J. Acoust. Soc. Am. **84**, 195–203.
Gockel, H., Moore, B., Patterson, R., and Meddis, R. (**2003**). "Louder sounds can produce less forward masking: Effects of component phase in complex tones," J. Acoust. Soc. Am. **114**(2), 978–990.

Hellman, R. P., and Zwislocki, J. J. (**1961**). "Some factors affecting the estimation of loudness," J. Acoust. Soc. Am. **33**, 687–694.
Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," Percept. Psychophys. **28**, 85–88.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Lim, J., Rabinowitz, W., Braida, L., and Durlach, N. (**1977**). "Intensity perception. VIII. Loudness comparisons between different types of stimuli," J. Acoust. Soc. Am. **62**(5), 1256–1267.
Mapes-Riordan, D., and Yost, W. A. (**1999**). "Loudness recalibration as a function of level," J. Acoust. Soc. Am. **106**, 3506–3511.
Marks, L. E. (**1994**). "Recalibrating the auditory system: The perception of loudness," J. Exp. Psychol. Hum. Percept. Perform. **20**, 382–396.
McDermott, H., Lech, M., Kornblum, M., and Irvine, D. (**1998**). "Loudness perception and frequency discrimination in subjects with steeply sloping hearing loss: Possible correlates of neural plasticity," J. Acoust. Soc. Am. **104**(4), 2314–2325.
McFadden, M. (**1975**). "Duration-intensity reciprocity for equal loudness," J. Acoust. Soc. Am. **57**(3), 702–704.
Moore, B. (**2004**). "Testing the concept of softness imperception: Loudness near threshold for hearing impaired ears," J. Acoust. Soc. Am. **115**(6), 3103–3111.
Moore, B., and Glasberg, B. (**1985**). "Effects of flanking noise bands on the rate of growth of loudness of tones in normal and recruiting ears," J. Acoust. Soc. Am. **77**(4), 1505–1513.
Moore, B., Vickers, D., and Baer, T. (**1999**). "Factors affecting the loudness of modulated sounds," J. Acoust. Soc. Am. **105**(5), 2757–2772.
Nieder, B., Buus, S., Florentine, M., and Scharf, B. (**2003**). "Interactions between test- and inducer-tone durations in induced loudness reduction," J. Acoust. Soc. Am. **114**, 2846–2855.
Pedersen, O. J., Lyregaard, P. E., and Poulsen, T. (**1977**). "The round robin test on impulsive noise," Rept. No. 22, The Acoustics Laboratory, Technical University of Denmark, pp. 1–180.
Poulsen, T. (**1981**). "Loudness of tone pulses in a free field," J. Acoust. Soc. Am. **69**(6), 1786–1790.
Rankovic, C., Viemeister, N., and Fantini, D. (**1988**). "The relation between loudness and intensity difference limens for tones in quiet and noise backgrounds," J. Acoust. Soc. Am. **84**(1), 150–155.
Scharf, B. (**1970**). "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. **I**.
Schlauch, R. S., DiGiocanni, J. J., and Ries, D. T. (**1998**). "Basilar membrane nonlinearity and loudness," J. Acoust. Soc. Am. **103**(4), 2010–2020.
Schlauch, R. S., DiGiocanni, J. J., and Ries, D. T. (**1997**). Abstract of the 20th MidWinter Research Meeting of ARO, p. 227.
Schneider, B., and Parker, S. (**1990**). "Does stimulus context affect loudness or only loudness judgements?," Percept. Psychophys. **48**(5), 409–418.
Small, A., Brandt, J., and Cox, P. (**1962**). "Loudness as a function of signal duration," J. Acoust. Soc. Am. **34**(4), 513–514.
Stephens, S. D. G. (**1974**). "Methodological factors influencing loudness of short duration sounds," J. Sound Vib. **37**, 235–246.
Stevens, J., and Hall, J. (**1966**). "Brightness and loudness as functions of stimulus duration," Percept. Psychophys. **1**, 319–327.
Stevens, S. S., and Greenbaum, H. B. (**1966**). "Regression effect in psychophysical judgement," Percept. Psychophys. **1**, 439–446.
Verhey, J., and Kollmeier, B. (**2002**). "Spectral loudness summation as a function of duration," J. Acoust. Soc. Am. **111**(3), 1349–1358.
Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics—Facts and Models* (Springer, Berlin).

Rhesus macaques spontaneously perceive formants in conspecific vocalizations

W. Tecumseh Fitch[a)]
School of Psychology, University of St. Andrews, St. Andrews, Fife, Scotland, KY169JP, United Kingdom

Jonathan B. Fritz[b)]
Laboratory of Neuropsychology, NIMH, NIH, Bethesda, Maryland 20742

(Received 7 March 2006; revised 2 July 2006; accepted 5 July 2006)

We provide a direct demonstration that nonhuman primates spontaneously perceive changes in formant frequencies in their own species-typical vocalizations, without training or reinforcement. Formants are vocal tract resonances leading to distinctive spectral prominences in the vocal signal, and provide the acoustic determinant of many key phonetic distinctions in human languages. We developed algorithms for manipulating formants in rhesus macaque calls. Using the resulting computer-manipulated calls in a habituation/dishabituation paradigm, with blind video scoring, we show that rhesus macaques spontaneously respond to a change in formant frequencies within the normal macaque vocal range. Lack of dishabituation to a "synthetic replica" signal demonstrates that dishabituation was not due to an artificial quality of synthetic calls, but to the formant shift itself. These results indicate that formant perception, a significant component of human voice and speech perception, is a perceptual ability shared with other primates. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2258499]

PACS number(s): 43.66.Gf, 43.80.Lb, 43.80.Ka, 43.71.An, 43.70.Bk [JAS] Pages: 2132–2141

I. INTRODUCTION

Formants (the spectral prominences generated by vocal tract resonances) provide a critical acoustic cue to many phonemic differences in human speech (Fant, 1960; Lieberman and Blumstein, 1988; Titze, 1994; Ladefoged, 2001). In addition to their central role in speech perception, recent studies indicate that human formants also act as a perceptual cue to body size (Fitch, 1994; Ives *et al.*, 2005; Smith *et al.*, 2005) and play a role in attractiveness judgments (Collins, 2000; Feinberg *et al.*, 2005). After many years of assuming that formants are a peculiarity of the human speech signal, it is becoming clear to bioacousticians that formants are present in animal calls (Fitch, 1997; Owren *et al.*, 1997; Rendall *et al.*, 1998; Riede and Fitch, 1999; Riede *et al.*, 2005). Thus there is an increasing interest in the roles that formants might play in animal communication, and in the degree to which formant perception in nonhuman animals represents a homologue to the mechanisms involved in human speech perception. Although it is clear that many animals (including dogs, horses, baboons, macaques, and various bird species) can be trained to discriminate between different speech signals, the critical questions for understanding animals' own communication systems are whether they perceive formants in their own, species-specific vocalizations. This requires, first, determining if the spectral prominences examined are, in fact, formants; second, generating an appropriate stimulus set where formants are shifted; and finally, determining if animal subjects perceive formant changes (preferably spontaneously, without any training) (Fitch and Kelley, 2000; Reby *et al.*, 2005). The specific types of information conveyed by formants (e.g., call type, size, identity, attractiveness) can then be further investigated.

Several approaches have been used to determine whether some particular spectral prominences are formants (derive from vocal tract filtering), rather than representing "pseudo-formants" generated by some other process, or even harmonics of the voice source. Nonlinear phenomena involving the voice source are the clearest alternative mechanism capable of generating pseudo-formants [e.g., in scream vocalizations (Fitch *et al.*, 2002)]. The most direct way to exclude the possibility of source-generated spectral prominences is to place the animal in a light-gas (e.g., heliox) environment and induce it to vocalize (Roberts, 1975; Nowicki, 1987; Amundin, 1991; Rand and Dudley, 1993). If a spectral prominence is caused by formant filtering, the increased speed of sound leads to an increase in vocal tract resonance frequency, and thus to the center frequency of the spectral prominence. Such experiments were critical in revealing that spectral prominences in vocalizations of several frog species are *not* influenced by filtering in the frog's vocal tract and therefore do not represent formants (Rand and Dudley, 1993). These pseudo-formants apparently result instead from frequency-modulation within the laryngeal source (e.g., Martin, 1971). In contrast, heliox testing in many bird species has revealed the predicted shift of frequencies, allowing researchers to conclude that formant filtering plays an important role in avian vocal production (Nowicki, 1987; Suthers and Hector, 1988; Nowicki *et al.*, 1989; Fletcher and Tarnopolsky, 1999). When feasible, heliox testing thus provides the

[a)]Author to whom correspondence should be addressed; electronic mail: wtsf@st-andrews.ac.uk
[b)]Currently at: Center for Acoustic and Auditory Research, Institute for Systems Research, ECE, University of Maryland, College Park, MD 20742.

"gold standard" for demonstrating formant filtering in animal vocalizations.

Unfortunately, for many large or free-living animal species, immersion in a heliox atmosphere is impractical or impossible, and many other animals will refuse to vocalize in confined conditions. In such cases other techniques are necessary to demonstrate formant filtering. One alternative approach relies on the acoustic link between vocal tract length and formant frequencies (Fant, 1960; Titze, 1994). If vocalizations from multiple individuals of different body sizes are available, and the different vocal tract lengths for each individual can be measured, a strong correlation between vocal tract length and the frequencies of spectral prominences provides strong evidence for formant filtering (Fitch, 1997; Riede and Fitch, 1999; Reby and McComb, 2003). Although direct anatomical or x-ray measurements of vocal tract length provide the strongest evidence, a correlation of spectral prominences with head length or overall body size provides weaker evidence for vocal tract filtering (Fitch, 2000a). If an animal makes prominent, visible changes in its vocal tract length while vocalizing, and spectral prominences move in synchrony, this is also evidence for formants (particularly if other acoustic variables such as fundamental frequency or higher harmonics do not change in synchrony) [e.g., Hauser *et al.*, 1993; Fitch and Reby, 2001; Harris *et al.* (2006)]. Even simple inspection of spectrograms, combined with basic acoustic considerations and anatomical measurements from museum skulls or preserved specimens, can provide an indication of whether a set of spectral prominences could represent formants. For example, two recent papers have claimed that spectral prominences in mouse vocalizations represent formants (Ehret and Riecke, 2002; Geissler and Ehret, 2002), but the relatively low frequencies of these spectral prominences would entail a vocal tract longer than a mouse's entire body if they were formants. These authors appear to have confused harmonics of the glottal source with formant frequencies.[1] In summary, there are several possible sources of confirmation that spectral peaks in a given species' vocalizations represent formants. Converging, consistent data from several sources will of course provide the most convincing demonstration.

Once it has been established that formants are present in a particular type of vocalization, perceptual experiments are necessary to test whether the species in question attends to these cues. While it is possible to perform such experiments with natural call exemplars with varying formant frequencies, this leaves open the possibility that changes in other, unmeasured variables were noticed instead. Thus, synthetic signals in which only formants are changed are preferable. To the extent that the source/filter theory of vocal production applies to many animal vocalizations (Fitch and Hauser, 1995; Fitch and Hauser, 2002), various well-understood signal processing techniques, developed by speech scientists, are available to modify formants without changing other aspects of the signal. For example, LPC-based analysis and resynthesis can be used to artificially separate the signal into source and filter components, if certain preconditions are fulfilled (Fitch, 1997; Owren and Bernacki, 1988). Then, the filter component can be modified in specific ways, and source and filter can be recombined to create a natural-sounding vocalization where only the spectral prominences have been shifted (Moorer, 1979; Moore, 1990; Fitch, 2002; Smith *et al.*, 2005). These and similar techniques can easily be implemented on desktop systems using software such as MATLAB or PRAAT. Such techniques have been successfully applied to whooping cranes (Fitch and Kelley, 2000) and red deer (Reby *et al.*, 2005), and are used in the present study with macaques. It is important to recognize, however, that not all vocalization types are appropriate for such techniques. If present, source-determined spectral peaks (in particular the fundamental frequency and low harmonics) must be lower than the lowest formant (as is the case in adult speech). Even in such ideal cases, however, it is difficult to distinguish formant frequency perception from the perception of differences in harmonic amplitudes [as seen in some birds Cynx *et al.*, 1990]. With very high fundamentals it also remains possible that source/filter interactions could occur, violating the independence assumption of the source/filter theory and of LPC (Fitch and Hauser, 1995). The ideal calls for resynthesis techniques and are therefore those in which no source-related peaks exist, e.g., calls with source components that consist of broadband noisy excitation or impulse trains. Given appropriate precautions, however, linear prediction or similar digital processing techniques allow researchers to generate a set of stimuli in which formants are manipulated, but all other acoustic variables are held constant.

Given an appropriate set of synthetized animal vocalizations, we can finally proceed to experimentally determine whether animals of a particular species perceive formant changes in their own species' calls. Presumably, given adequate prolonged training, any animal with adequate spectral sensitivity should be able to learn to distinguish between sounds with shifted formants (for example, many different vertebrate species have been trained to distinguish between synthetic human vowels differing only in formant frequencies). But if animals naturally make use of formants in their communication system, they should react to changes in formant frequency spontaneously, without requiring specific training or reinforcement. Thus, tests of spontaneous perception, such as habituation/discrimination techniques, provide the strongest evidence for a species' use of formants as a meaningful communicative parameter. Such techniques were introduced for perceptual experiments with human infants (Eimas *et al.*, 1971), they have been successfully used with many different animal species, including nonhuman primates (e.g., Seyfarth and Cheney, 1990; Rendall, 1996; Fischer, 1998; Hauser, 1998). However, results from such experiments employing resynthesized calls, where only formants change, are currently available for only two nonhuman species: whooping cranes (Fitch and Kelley, 2000) and red deer (Reby *et al.*, 2005).

In this study, we test the hypothesis that rhesus macaques (*Macaca mulatta*) spontaneously perceive formant frequencies in conspecific vocalizations. We also use a control condition to test the adequacy of our monkey call synthesis techniques, specifically to ensure that these techniques introduce no perceptually salient artificial quality to our synthetic calls. We chose the rhesus macaque, a common labo-

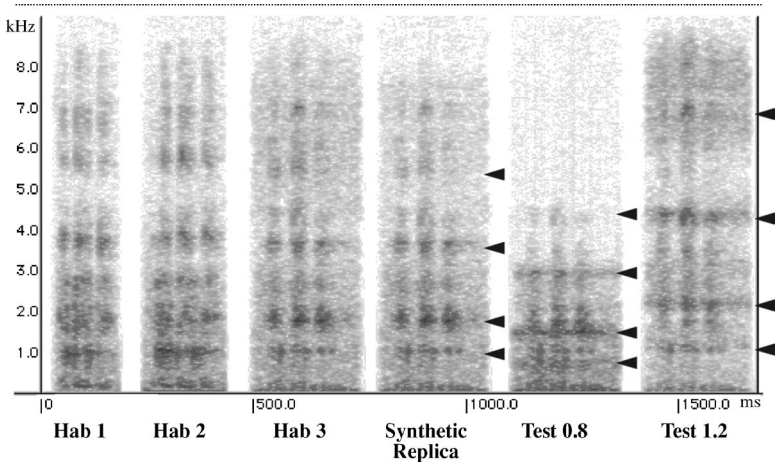


FIG. 1. Spectrogram illustrating the calls used in this experiment. The first three calls ("Hab") form the habituation set and are natural, recorded calls. The last three are synthetic calls. "Synthetic replica" has the formants unchanged from Hab 3, while in the two test calls, formants have been shifted down (0.8) or up (1.2) in frequency by 20%. Black arrows indicate the lowest four formant frequencies

ratory primate, as our test species because prior studies have demonstrated that formants are present in some calls of this species, and could potentially carry information about body size, call type, and/or individual identity (Hauser *et al.*, 1993; Fitch, 1997; Rendall *et al.*, 1998). Furthermore, with intensive training in the laboratory, macaques of a closely related species (*Macaca fuscata*) are able to recognize changes in formant frequencies in synthesized vowels with an accuracy rivaling humans' (Sommers *et al.*, 1992). However, no previous study has combined spontaneous perceptual testing with a synthetic stimulus set to conclusively demonstrate formant perception in this (or any other nonhuman primate) species.

Experimental Approach. We used a custom-designed monkey call synthesizer, based on the source/filter theory of vocal production (Fant, 1960; Titze, 1994; Fitch and Hauser, 1995; Fitch, 2002), to create natural-sounding macaque calls. Our central hypothesis was that rhesus monkeys spontaneously perceive changes in formant frequencies in conspecific vocalizations, and find these differences significant enough to warrant dishabituation. This was tested by habituating an animal to a set of natural stimuli from a single individual, and then playing a test call in which the formants had been digitally shifted while holding other acoustic variables constant. If a subject dishabituated to the modified "test" call, we concluded that it had perceived the test call as different from the preceding stimuli, and therefore perceived the formant changes. A nontonal noisy vocalization type, the aggressive "pant threat," was used to avoid the issues with harmonics cited earlier (Fig. 1).

Nonetheless, the cause of dishabituation in this case would remain ambiguous because a subject might simply perceive the formant-shifted call as "sounding synthetic" in general, rather than noticing the formant changes in particular. To exclude this possibility, we first tested a "synthetic replica" call created by applying the same software manipulations to one of the habituation calls, but without modifying formants (following Fitch and Kelley, 2000). To the human ear, synthetic replicas sound subtly cleaner (less noisy) than the original recordings, but otherwise identical. If the percept is similar for monkeys, we predicted that they would transfer habituation to this stimulus. If the subject did dishabituate to the synthetic replica, responding to resynthesis *per se*, the session was ended. A series of such results would indicate an inadequacy in our monkey call synthesis techniques. However, if the subject transferred habituation to the synthetic replica, we proceeded to test our core hypothesis by playing the synthetic, formant-shifted call, which differed from the synthetic replica presented previously only in that the formant frequencies were experimentally modified. All other acoustic aspects (e.g., duration, amplitude, pulse rate, other timbral cues, etc.) were identical to the call played just previously (Fig. 1).

II. MATERIALS AND METHODS

A. Animal subjects

The subjects in these experiments were 13 adult rhesus macaques (*Macaca mulatta*, age 4–9 yr, 7 females, 6 males). All experiments were conducted under an approved NIMH study proposal in accord with NIH Guidelines on the Care and Use of Primates. The monkeys were on a 12 h light/dark cycle (7 am–7 pm) and the experiments were conducted between 10 am and 5 pm. The monkeys were housed at NIH in individual cages in colony rooms, and had received no previous exposure to playback of vocal stimuli before these experiments were conducted. Each monkey was run individually.

B. Behavioral protocol

This behavioral protocol followed standard habituation/dishabituation paradigms (e.g., Seyfarth *et al.*, 1980; Hauser, 1998) in most respects. We concealed a loudspeaker in a testing room, and then introduced a monkey subject, sitting in a monkey chair facing directly away from the speaker. Using a video camera to monitor head position, we waited until the subject was looking directly away from the speaker ($180° \pm 20°$) before initiating each playback event. The criterion for response was a head turn in the direction of the speaker initiated within 3 s of sound playback. An experiment started with repeated playback of calls from the habituation set until the subject habituated (defined as a failure to respond to three successive playbacks). The average time between playbacks was 30 s (range 8–200 s; within the range of variation of pant-threat rates observed in free-living

populations (Fitch, unpublished data). After successful habituation, we played the synthetic replica. A dishabituation response to this "control" stimulus terminated the session. However, if subjects transferred habituation to the control, we then played the formant-shifted stimulus. If the animal dishabituated, we could safely conclude that the dishabituation was caused by the formant shifts, and not by other acoustic cues or artifacts of synthesis, and the experimental session was concluded: dishabituation to the "test" stimulus terminated the session. During the habituation phase, the same sound was sometimes played twice in a row, so dishabituation to the test stimulus could not have resulted simply from the monkey perceiving repetition. Thus, consistent dishabituation to the formant-shifted stimulus constitutes strong evidence that monkeys spontaneously perceive formants in conspecific calls. However, a null result (failure to dishabituate to either test stimulus) could be caused simply by general habituation: sensory fatigue, distraction, adaptation to the playback setting, or other confounds. To exclude this possibility, sessions in which the subject failed to dishabituate to either of the previous two stimuli were ended with a "post-test" stimulus, a monkey "shrill bark" alarm call, expected to reliably elicit a response, and thus reject this final control hypothesis (following Hauser, 1998).

C. Materials and testing procedure

The experiment was performed in an empty playback room (approximately 4×3 m, 2.5 m h) acoustically treated with Sonex 1 in. foam (9.4×2 ft panels, Sonex #10897, Illbruck, Minneapolis, MN). The experimenter, computer equipment, and playback speaker were hidden by a 2.3×1.9 m curtain made of heavy opaque cloth that bisected the room diagonally. This curtain was acoustically transparent (reducing audio levels by only 2 dB SPL and introducing no audible distortion). Monkeys were seated in custom-built plexiglas primate chairs (20×25 cm, 56 cm high) which allowed free head movement in the horizontal plane. Monkeys were seated in a fixed standard position in the room, facing away from the curtain and loudspeaker, with the loudspeaker 1.5 m directly behind them. Responses were filmed using a Panasonic Digital 5000 VHS video camera mounted on a tripod, 1.2 m away and directly facing the subject, monitored via a Sony Trinitron monitor during playbacks, and simultaneously recorded to VHS tapes (Fuji HQ-120, SP) with a Sony VHS Hi-Fi recording deck. Playbacks were performed using an Apple Powerbook computer and custom playback software, using the built-in sound output (44.1 kHz sampling rate, 16 bit quantization) attached to a Bose Roommate II self-powered speaker. Playback levels were determined with a Radio Shack Sound Level Meter, set for C-weighted, fast response measurement. Sound level measurements were made by mounting the SPL meter at the location occupied by the monkey's head during experiments. Broadband ambient ventilation noise in the playback room was 62–65 dB SPL, effectively masking computer-generated fan noise and keypresses. Playback levels were adjusted to 75–78 dB SPL and were very clearly audible above this background. Since playback time was initiated by the monkey facing calmly away from the speaker, a maximum delay of 4 min between playbacks was chosen. If exceeded, the experiment would have been terminated, but this never occurred.

D. Stimuli and signal processing

Monkey calls were synthesized using techniques developed by WTF (Fitch and Hauser, 1995; Fitch and Kelley, 2000; Fitch, 2002). The pant-threat vocalizations used in this study were recorded from a single adult female macaque, unknown to our monkey subjects, from a rhesus population on the island of Cayo Santiago near Puerto Rico (monkey 74B, recorded by Marc D. Hauser, Harvard University, using a Sennheiser microphone and Sony Walkman Professional cassette recorder). Call spectra were examined for high frequency energy; threat calls contained no appreciable energy above 8 kHz. The three highest-quality pant-threat calls were selected as habituation stimuli, low-pass filtered (8500 Hz) and downsampled to 18.5 kHz sampling rate for further digital processing. Final versions were upsampled and played back at 44.1 kHz. One habituation call was submitted to an 18-pole linear prediction analysis (512 sample window, no preemphasis, rectangular window), yielding a filter closely approximating the smoothed magnitude spectrum of the call. The calls was then inverse filtered using this filter, yielding an error signal which approximates the laryngeal source signal (this "source signal" consisted of three impulses with some attendant noise). Once the source signal and the filter were separated, various modifications of either are independently possible. In this experiment, we scaled the entire filter function, increasing or decreasing each resonance by 20% by finding its roots (corresponding to individual formants) and then multiplying each formant frequency by a fixed factor (1.2 or 0.8). Increasing (or decreasing) the formant frequencies is analogous to shortening (or lengthening) the vocal tract, respectively. The original model call had intermediate formant frequency values, so 20% up- or downshifting of formants corresponds to a VTL of approximately 8, or 9.5 cm (about 6 and 12 kg body weight, respectively), remaining well within the normal acoustic range for adult macaques (Fitch, 1997). This relatively large change was chosen to increase the chances that the acoustic difference would be not just perceptible, but also behaviorally meaningful to our subjects, and thus to induce dishabituation (Nelson and Marler, 1989). The modified filter function was then recombined with the original source (polynomialized back into a filter function and used to create a new synthetic signal by filtering the source signal). All signal processing was performed in MATLAB 5.1 (The Mathworks, Inc., Natick, MA) using the Signal Processing Toolbox and custom software.

III. DATA ANALYSIS

All trials were videotaped for additional offline analysis. All critical trials (last habituation, synthetic replica, test, and post-test) and a randomly selected set of habituation trials from each animal were digitized (Apple iMovie software) and scored by two observers (one blind to condition). Interobserver agreement was very high (97% agreement). To further quantify the strength of response we measured both la-

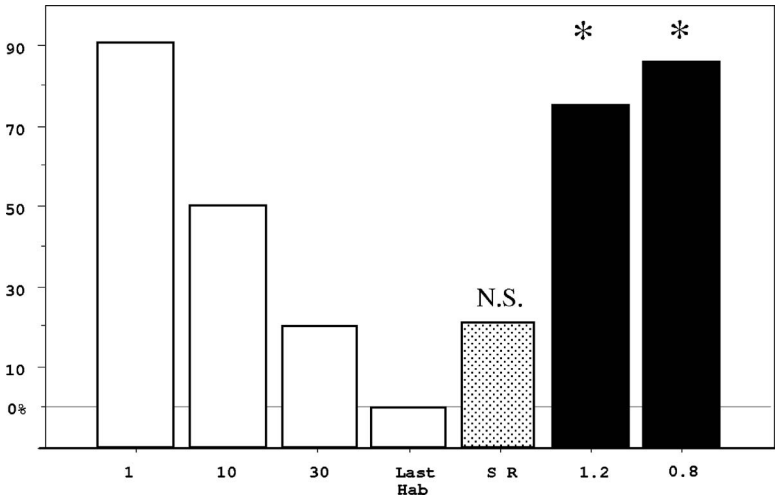


FIG. 2. Results of playbacks: Chance of response drops across habituation trials (1, 10, and 30 represent the first, tenth, and thirtieth trials), reaching 0 (by definition in our protocol) on the last habituation trial. Dishabituation to the synthetic replica (SR), where formants were not shifted, rarely occurred. Dishabituation nearly always occurred to the formant shifted stimuli (1.2 and 0.8), thus indicating perception of the formant shift imposed on these stimuli.

tency to look (in video frames, 30 frames/s) and magnitude of looks (in degrees from the initial head direction at playback to the maximum position during the look, max 180°). In two cases, monkeys' responses were ill-suited to an analysis of look latency and magnitude, once because a monkey began a head movement just before sound playback (negative latency), then followed it with a 135° turn toward the speaker in the opposite direction, and in another case because the monkey made a small 15° turn very quickly, followed by a large turn (90°) 6 s later (outside our arbitrary maximal cutoff window for scoring "looks"). In both cases, we then repeated the trial and received clear, short-latency looks that were used in the analysis.

Our original plan was to run each monkey twice, using both the up- or downshifted stimulus (randomly assigned). This would have enabled a statistical test to determine if up- or downshifting was more salient. Unfortunately, due to circumstances beyond our control, only 6 of the 13 monkeys were run a second time, yielding 19 experiments total. To ensure that repeated playbacks to a subset of monkeys did not influence our results we present analyses for both the first trials alone ($N=13$, which we term the "first" trials) and this total number ($N=19$). For statistical analysis we used one-tailed binomial tests to determine whether the proportion of looks (indicating dishabituation) relative to nonlooks (continued habituation) differed from chance. One-tailed tests are appropriate in this paradigm because the direction of the response is predicted in advance, and differs between conditions ("no" for synthetic replicas and "yes" for test stimuli). The significant p values we obtained with one-tailed statistics in general remain significant with a two-tailed test. The binomial test requires the specification of the base frequency of looking after habituation to three trials, which we took conservatively to be 50%. Statistics were performed in MATLAB 5.1 and STATVIEW 5.0 (SAS Institute, Cary, NC).

IV. RESULTS

We completed a total of 19 playback experiments with 13 different monkey subjects (six were tested twice as discussed above). The mean number of trials to habituation was 22.5 (min 4, max 45). The pattern of dishabituation is illustrated in Fig. 2.

Only four subjects dishabituated to the synthetic replica stimulus (50% binomial test 19(15), N.S.), indicating that the monkeys did not significantly dishabituate to this control stimulus. Only three monkeys dishabituated in the 13 "first" trials (50% binomial test 13(10), N.S.). Failure to consistently dishabituate to synthetic replica calls shows that the speech synthesis techniques used here are capable of generating realistic-sounding macaque calls. In the 15 remaining trials, 12 subjects who heard the formant-shifted test stimulus dishabituated (50% binomial test 15(3), $p=0.018$). For "first" trials, 8 of the remaining 10 subjects dishabituated to the test stimulus (50% binomial test 10(2), $p=0.055$). These results indicate that the monkeys heard and responded to the change in formant frequencies in these stimuli. Of the three subjects who did not dishabituate to the test stimulus, all responded to a post-test alarm call, indicating that no monkeys had habituated to the playback situation in general.

To further quantify the strength of response we measured both latency to look (in video frames, 30 frames/s) and magnitude of looks (in degrees from the initial head direction at playback to the maximum position during the look, max 180°). Monkeys looked faster to formant shifted stimuli than to synthetic replicas. Most looks (11/12) to test stimuli had latencies of less than 20 frames (667 ms) (mean 12 frames or 400 ms). In contrast, half of the looks to synthetic replicas (2/4) had latencies of more than 1 s (mean 26 frames or 867 ms). To further explore the reaction of subjects to synthetic calls, we compared looks to natural calls (last habituation, and post-test stimuli) with those to synthetic calls (the synthetic replica and test stimuli). There were no significant differences in the latency to look (Mann-Whitney $U=112$, $p=0.92$) or magnitude of looks (Mann-Whitney $U=109$, $p=0.84$) between synthetic and natural calls. We found no difference in monkeys' dishabituation to the upshifted versus downshifted test stimuli: of 12 dishabituations 6 were to upshifted and 6 to downshifted, and of the 3 failures to dishabituate 1 was to the upshifted and 2 were to the downshifted stimulus. Although monkeys looked more rapidly to the downshifted stimuli (mean 4.8 frames or 160 ms) than to the upshifted stimuli (mean 18.4 frames or 620 ms), this difference was not significant (unpaired t-test,

$t=1.45$, $p=0.174$). These findings thus reinforce the basic finding that monkeys ignored the synthetic replica and strongly responded to formant shifts in either direction.

V. DISCUSSION

The results of these experiments show that rhesus macaque perceive changes in formant frequencies in their own species-specific pant threat vocalizations. A significant proportion (12 out of 15) of our monkey subjects found formant shifts of 20% up or down to be salient enough to warrant dishabituation. Our use of a call type that lacks harmonics shows that this result cannot be attributed to perception of relative harmonic amplitude. The use of digital synthesis techniques allowed us to vary only formant frequencies, without varying other aspects of the signal. A failure to dishabituate to "synthetic replicas" which had been digitally processed *without* shifting formants shows that this result is not due to some unintended artificial quality imposed by the analysis/resynthesis technique. We conclude that rhesus macaques are sensitive to formants in vocalizations without training or reinforcement. It thus appears highly likely that formants play some role in the communication system of rhesus macaques.

These results are compatible with a number of previous results from rhesus macaques. Fitch (1997) showed that the spectral prominences in this species are formants, and showed that formant frequencies correlate with body size. Rendall *et al.* (1996) showed that free-ranging rhesus macaques distinguish identity of individuals by acoustic cues in their vocalizations, and acoustic analysis indicated that formant frequencies are potentially important cues for identity in this species (Rendall *et al.*, 1998). Injections of xylocaine into the perioral region in this species, which block the ability to produce the lip-rounding associated with "coo" calls, presumably affecting formants, led to differential reactions of conspecifics in this species (Hauser and Schön Ybarra, 1994). Combined, all of these studies converge on the conclusion that rhesus macaques perceive formant frequencies in their own vocalizations, though the information they extract from these cues remains uncertain (see the following).

Results from various other primate species provides additional converging evidence for formant perception in nonhuman primates. Several authors have termed spectral prominences in the calls of nonhuman primates "formants" with little further discussion or justification (Lieberman, 1968; Andrew, 1976; Richman, 1976). In baboons, species-specific "grunt" vocalizations have an acoustic structure quite similar to human vowels (Owren *et al.*, 1997) with spectral prominences hypothesized to represent formants, and significant correlations between body size and formants support this hypothesis (Rendall, 2005). A recent study shows that spectral prominences in guereza monkeys both correlate with body size, and change in close synchrony with lip movements, strongly suggesting that they represent formants (Harris *et al.*, 2006). Regarding perception, laboratory studies with the closely related Japanese macaque *Macaca fuscata*, although using synthetic vowel stimuli rather than conspecific calls, with training demonstrated an exquisite sensitivity to formants in this species, rivaling or exceeding that of humans (Sommers *et al.*, 1992). In a training paradigm baboons were shown to be quite sensitive to formant changes in grunts synthesized with a human Klatt synthesizer (and were similarly sensitive to formant changes in human vowels) (Hienz *et al.*, 2004). Finally, vervet monkeys possess spectral prominences that may represent vocal tract resonances (Owren and Bernacki, 1998), and perceptual tests show that vervets respond to these prominences in a classification task which involved training but did not demand that the subjects attend to that particular cue (Owren, 1990b, a). All of these data converge to suggest that formant perception may be a widespread capability in Old World monkeys.

Techniques similar to those used here have recently been utilized to demonstrate spontaneous formant perception in cranes and deer (Fitch and Kelley, 2000; Reby *et al.*, 2005). In both cases the species was tested because they possess unusual vocal adaptations hypothesized to modify formants. In whooping cranes *Grus americana*, the trachea is greatly elongated. Due to the anatomy of the avian vocal production system, tracheal elongation lowers formant frequencies, and has been hypothesized as a means of exaggerating size (Fitch, 1999). This hypothesis was tested by modifying formants in a nonharmonic call (the "contact call") using computer resynthesis, and demonstrating that listening cranes noticed this change (Fitch and Kelley, 2000). In red deer *Cervus elaphus*, adult stags have a permanently-descended larynx, and during territorial roars they lower the larynx even further to its anatomical limit. Again, this lowers formants, and was hypothesized to exaggerate projected body size (Fitch and Reby, 2001). This hypothesis was tested via playback of resynthesized calls, which demonstrated formant perception and use of formants as cues to size (Reby *et al.*, 2005). Thus, in at least two nonprimate species with formant-modifying vocal anatomy, formant perception is present.

These previous studies of animal formant perception leave open two evolutionary possibilities. First, formant perception and modification in primates, cranes and deer may represent convergent evolution. There are many examples of such convergence among vertebrate vocal communication, the most prominent being complex vocal imitation, which has evolved convergently several times (e.g., in humans, songbirds, and sea mammals), but is lacking in other primates (Janik and Slater, 1997; Fitch, 2000b; Marler and Slabbekoorn, 2004). Alternatively, formant perception in all of these species may be homologous, present in these widely separated species by virtue of inheritance from a common ancestor [the ancestral amniote, who lived some 300 million years ago (Smithson, 1989)]. If this latter hypothesis is correct, formant perception is predicted to be widespread among birds and mammals, even those which lack special formant-modifying anatomy. In particular, the crucial groups for testing the hypothesis that formant perception mechanisms are homologous in all these species are other nonhuman mammals, further bird species, and vocal reptiles such as the American alligator *Alligator mississipiensis*, which has clear formant-like bands in its vocalizations that correlate nicely

with body size (Fitch, unpublished data). It is also possible that amphibians may also perceive formants, though as discussed in Sec. I there is no evidence at present that the spectral prominences present in many anuran species represent formants (Rand and Dudley, 1993).

The ability to generate and control realistic animal acoustic signals (Fitch, 2002) opens new and exciting vistas in understanding the acoustic cues utilized in animal communication. The invention of speech synthesizers was a necessary prerequisite for the advances in speech perception starting in the 1970s, but off-the-shelf vocal synthesizers for animal calls still do not exist [although synthesizers designed for humans may be adequate in some cases (Hienz *et al.*, 2004)]. Fortunately, recent advances in our understanding of vertebrate vocal production have provided a major step forward in resolving this problem (Fitch and Hauser, 1995; Owren and Bernacki, 1988; Fitch, 2002). In particular, the realization that the source-filter theory of speech production also applies to animal vocalizations means that many of the algorithms developed by the speech community can now be applied, with the appropriate modifications, to nonhuman vocalizations. The ability to generate highly realistic animal vocalizations by computer allows us to choose and manipulate specific acoustic variables, leaving all other cues unchanged. With a careful choice of appropriate calls from a species' vocal repertoire, and new techniques for perceptual testing that do not involve training, we can now explore animal's perception of their own species-specific vocalizations at a level of detail previously impossible. Resynthesized calls can also be used in more traditional operant settings to allow accurate determination of difference limens and perceptual sensitivities (e.g., Owren, 1990b; Sinnott and Kreiter, 1991; Sommers *et al.*, 1992; Hienz *et al.*, 2004), or even in choice settings to explore perceptual preferences (McComb, 1991). Thus, the combination of digital synthesis of animal calls with playback experiments can provide a rich source of insight into the acoustic cues that play important roles in animal communication, in a wide variety of nonhuman vertebrate species.

What information might formants be providing? Differences in formant frequencies provide the primary cue to vowel identity in all human languages, and formant transitions also cue many important consonantal distinctions (Lieberman and Blumstein, 1988; Titze, 1994). Pitch information, though *present* in speech signals, is not necessary for their perception: even in so-called "tonal languages" like Chinese or Thai, formants are the key acoustic cue for most phonetic distinctions. Stimuli containing formant information alone are adequate to decode the phonetic content of speech (Remez *et al.*, 1981; Tartter, 1991), and the human auditory system automatically normalizes for vocal tract length and size information in speech from different speakers (Ives *et al.*, 2005; Smith *et al.*, 2005). Thus, of all the various acoustic cues that make up the complex speech signal, formants (or their synthetic analogues) are both necessary and sufficient for speech perception.

Previous studies suggest that formants may provide a similarly rich source(s) of information for animals (e.g., Sommers *et al.*, 1992; Fitch, 1997; Owren *et al.*, 1997; Rendall *et al.*, 1998; Riede and Fitch, 1999). Formants might provide several types of information to macaques. Formants are correlated with body size in macaques, baboons, and many mammals (Fitch, 1997; Riede and Fitch, 1999; Fitch, 2000a; Rendall, 2005), so formant perception could provide information about body size, as it does in humans (e.g., Fitch, 1994; Fitch and Giedd, 1999; Ives *et al.*, 2005; Smith *et al.*, 2005). In sexually dimorphic species such as rhesus macaques, size may also provide an indirect indication of sex or other secondary factors such as age, or degree of potential threat. Formants also may provide a reliable cue to individual identity: static differences in vocal tract anatomy (particularly in the nasal region) may remain invariant across calls and thus indicate identity (Rendall *et al.*, 1998). Formants may also provide one of the basic cues that differentiate different call types, similar to the way they distinguish different vowels in human speech (Lieberman, 1968). Hauser and colleagues (Hauser *et al.*, 1993; Hauser and Schön Ybarra, 1994) found that the vocal tract movements (specifically lip-protrusion associated with coo calls, and the retracted lips associated with screams), had well-defined acoustic effects on rhesus calls acoustics. Thus, macaques could potentially extract information about size, sex, identity, and call type from formants.

In this study we used a relatively gross manipulation—shifting all formant frequencies—that corresponds to a lengthening or shortening of overall vocal tract length. The positive response to these changes by our macaque listeners clearly opens the door to a more detailed exploration of formant perception in this and related species. In particular, while overall formant dispersion may be a cue to body size (and secondarily sex or age), more detailed aspects of the formant pattern, or changes in specific formants, may provide information about individual identity (Rendall *et al.*, 1998), call type (Hauser and Schön Ybarra, 1994), or other factors. The techniques developed in the current study should be seen as first steps, but clearly open the door to much more detailed study of these and other questions. It may be particularly interesting to learn whether the lowest three formants, which carry virtually all of the phonetic information in human speech, are preferentially attended to by macaques or other primates.

Neural basis of formant perception. If formants do indeed provide a multifaceted source of relevant information to nonhuman primate listeners, the corresponding neural mechanisms involved in decoding them might be quite complex, and thus may provide a richer neural substrate relevant to the evolution of speech perception than previously suspected. At least two broad regions of auditory cortex are likely to be involved in macaque formant perception, perhaps forming part of a more complex network for vocal perception. First, neurophysiological studies reveal neurons that show robust responses to bandpass-filtered noise stimuli (which are acoustically similar to formant frequencies) in the macaque lateral belt of auditory cortex, that may also play a role in analysis of conspecific vocalizations (Rauschecker *et al.*, 1995; Rauschecker and Tian, 2004), and recent fMRI analyses indicate potentially homologous activations in humans (von Kriegstein *et al.*, 2006). Second, multisensory

neurons in the superior temporal sulcus of the temporal lobe (STS) give robust responses to vocalizations (Ghazanfar, 2005), and neuroimaging data have provided additional evidence for activation of the STS in voice recognition and perception in both humans (Belin and Zatorre, 2003; Belin *et al.*, 2004; Uppenkamp *et al.*, 2006) and macaques (Poremba *et al.*, 2004). If formants play a crucial role in cueing vocal identity (Rendall *et al.*, 1998), such STS neurons should respond strongly and specifically to formant changes. In addition to lateral belt areas and STS, other cortical areas in the macaque and human may also play a role in vocal processing, including the rostral superior temporal gyrus (Poremba *et al.*, 2004) and lateral prefrontal cortex (Averbeck and Romanski, 2004; Gifford *et al.*, 2005; Romanski *et al.*, 2005). Whatever the neural substrates, the discovery that macaques perceive formants in their own vocalizations, and thus share a crucial component of human speech perception, opens the door to neuroscientific studies of formant perception that would be difficult or impossible in humans.

In conclusion, our experiments demonstrate that rhesus macaques are both capable of perceiving changes in formant frequencies in their own species-specific vocalizations, and that they spontaneously do so, without any training. Combined with previous data, this finding supports the hypothesis that formant perception is present in nonhuman primates, thus evolving prior to human speech, and indeed may be widespread among vertebrate species. Due to the importance of formants in speech, the mechanisms underlying macaque formant perception may represent an evolutionary precursor to the more sophisticated mechanisms underlying human speech perception, and this finding thus has important potential implications for our understanding of both primate communication systems and the evolution of spoken language. Wang has proposed that there may be an auditory cortical pathway specialized for processing vocal communication sounds in primates (Wang, 2000). Many scholars have suggested that human speech perception built upon pre-existing sensory mechanisms in our primate ancestors (Snowdon, 1982; Hauser and Fitch, 2003), while others suggest that at least some aspects evolved in humans *de novo* (Sinnott and Williamson, 1999; Pinker and Jackendoff, 2005). The ability to synthesize and manipulate specific aspects of primate calls using a variety of transformations [such as the source/filter model used here (Fitch, 2002), the Mellin transform (Smith *et al.*, 2005), or a parametric "virtual vocalization" model approach (DiMattina and Wang, 2005)] provides a new tool to help resolve these debates empirically, and opens the door to detailed exploration of the information-processing mechanisms underlying vocal perception in nonhuman primates.

ACKNOWLEDGMENTS

We thank Marc D. Hauser for advice on experimental design and for generously sharing his macaque recordings, and Ricardo Gil da Costa, Asif Ghazanfar, Mortimer Mishkin, Drew Rendall, and Richard C. Saunders and several anonymous reviewers for comments on the manuscript. We thank Ludise Malkova, Richard Saunders, and Mortimer Mishkin for their support in completing these experiments, conducted in the Laboratory of Neuropsychology, NIMH, NIH, and John Newman, Steven Suomi, Peggy O'Neill-Wagner, and Jennifer Stowe of the NIH Animal Care Center in Poolesville, MD for their help in piloting this paradigm. This work was funded by the NIH (NIDCD T32 DC00038 to W.T.F. and an NIH IRTA Postdoctoral Fellowship to J.B.F.) and by the NIMH IRP.

[1]A simple analysis demonstrates that the spectral peaks in the synthetic signals used in Ehret and Riecke's study could not represent the formants of a mouse, since a vocal tract capable of generating the frequency components that Ehret and Riecke term "formants" would have to be longer than a mouse pup's entire body. The lowest predicted formant frequencies for a mouse pup (assuming a 1 cm vocal tract length) would be about 9 kHz, but the synthetic mouse calls had frequency components at 3.8, 7.6, and 11.4 kHz, a frequency spacing of 3.8 kHz, which (if the components were formants) would correspond to a vocal tract length of 4.6 cm. This is longer than the entire body length of a neonatal mouse (about 3 cm), whose calls the synthetic stimuli are supposed to mimic, and are indeed considerably beyond the values predicted for an adult mouse with a vocal tract length of (liberally) 2 cm (4.3, 13.1, 21.9 kHz). It would thus be physically impossible for a neonatal mouse, or even an adult mouse, to produce formants with the stated frequencies. The individual frequency components manipulated in Ehret and Riecke's synthetic stimuli, therefore represent harmonics: aspects of the laryngeal source and not the vocal tract filter (Roberts, 1975; Weisz *et al.*, 2001; Liu *et al.*, 2003). The distinction between formants and harmonics is central in speech science, and redefining such terms in a different species "to stress {their} perceptual significance" (Geissler and Ehret, 2002) both deviates from long-accepted usage and is highly misleading if such data are to be related to speech in humans, or to vocalizations in other species.

Amundin, M. (**1991**). "Helium effects on the click frequency spectrum of the Harbor porpoise, *Phocoena phocoena*," J. Acoust. Soc. Am. **90**, 53–59.

Andrew, R. J. (**1976**). "Use of formants in the grunts of baboons and other nonhuman primates," Ann. N.Y. Acad. Sci. **280**, 673–693.

Averbeck, B. B., and Romanski, L. M. (**2004**). "Principal and independent components of macaque vocalizations: Constructing stimuli to probe high-level sensory processing," J. Neurophysiol. **91**, 2897–2909.

Belin, P., Fecteau, S., and Bedard, C. (**2004**). "Thinking the voice: neural correlates of voice perception," Trends in Cognitive Science **8**, 129–135.

Belin, P., and Zatorre, R. J. (**2003**). "Adaptation to speaker's voice in right anterior temporal lobe," NeuroReport **14**, 2105–2109.

Collins, S. A. (**2000**). "Men's voices and women's choices," Anim. Behav. **60**, 773–780.

Cynx, J., Williams, H., and Nottebohm, F. (**1990**). "Timbre discrimination in Zebra finch (*Taeniopygia guttata*) song syllables," J. Comp. Psychol. **104**, 303–308.

DiMattina, C., and Wang, X. (**2005**). "Virtual vocalization stimuli for investigating neural representations of species-specific vocalizations," J. Neurophysiol. **95**, 1244–1262.

Ehret, G., and Riecke, S. (**2002**). "Mice and humans perceive multiharmonic communications sound in the same way," Proc. Natl. Acad. Sci. U.S.A. **99**, 479–482.

Eimas, P. D., Siqueland, P., Jusczyk, P., and Vigorito, J. (**1971**). "Speech perception in infants," Science **171**, 303–306.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., and Perrett, D. I. (**2005**). "Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices," Anim. Behav. **69**, 561–568.

Fischer, J. (**1998**). "Barbary macaques categorize shrill barks into two call types," Anim. Behav. **55**, 799–807.

Fitch, W. T. (**1994**). *Vocal Tract Length Perception and the Evolution of Language* (UMI Dissertation Services, Ann Arbor, MI).

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. (**1999**). "Acoustic exaggeration of size in birds by tracheal elongation: Comparative and theoretical analyses," Journal of Zoology (London) **248**, 31–49.

Fitch, W. T. (**2000a**). "Skull dimensions in relation to body size in nonhuman mammals: The causal bases for acoustic allometry," Zoology **103**, 40–58.
Fitch, W. T. (**2000b**). "The evolution of speech: A comparative review," Trends in Cognitive Science **4**, 258–267.
Fitch, W. T. (**2002**). "Primate vocal production and its implications for auditory research," in *Primate Audition: Ethology and Neurobiology*, edited by A. A. Ghazanfar (CRC Press, Boca Raton, FL), pp. 87–108.
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.
Fitch, W. T., and Hauser, M. D. (**1995**). "Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on 'honest' advertisement," Am. J. Primatol **37**, 191–219.
Fitch, W. T., and Hauser, M. D. (**2002**). "Unpacking 'Honesty:' Vertebrate vocal production and the evolution of acoustic signals," in *Acoustic Communication*, edited by A. M. Simmons, R. F. Fay, and A. N. Popper (Springer, New York), pp. 65–137.
Fitch, W. T., and Kelley, J. P. (**2000**). "Perception of vocal tract resonances by whooping cranes, *Grus americana*," Ethology **106**, 559–574.
Fitch, W. T., Neubauer, J., and Herzel, H. (**2002**). "Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production," Anim. Behav. **63**, 407–418.
Fitch, W. T., and Reby, D. (**2001**). "The descended larynx is not uniquely human," Proc. R. Soc. London, Ser. B **268**, 1669–1675.
Fletcher, N. H., and Tarnopolsky, A. (**1999**). "Acoustics of the avian vocal tract," J. Acoust. Soc. Am. **105**, 35–49.
Geissler, D. B., and Ehret, G. (**2002**). "Time-critical integration of formants for perception of communication calls in mice," Proc. Natl. Acad. Sci. U.S.A. **99**, 9021–9025.
Ghazanfar, A. A. (**2005**). "Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex," J. Neurosci. **25**, 5004–5012.
Gifford, G. W., III, MacLean, K. A., Hauser, M. D., and Cohen, Y. E. (**2005**). "The neurophysiology of functionally meaningful categories: Macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations," J. Cogn Neurosci. **17**, 1471–1482.
Harris, T. R., Fitch, W. T., Goldstein, L. M., and Fashing, P. J. (**2006**). "Black and white colobus monkey (*Colobus guereza*) roars as a source of both honest and exaggerated information about body mass," Ethology **112**, 911–920.
Hauser, M. D. (**1998**). "Functional referents and acoustic similarity: Field playback experiments with rhesus monkeys," Anim. Behav. **55**, 1647–1658.
Hauser, M. D., Evans, C. S., and Marler, P. (**1993**). "The role of articulation in the production of rhesus monkey (*Macaca mulatta*) vocalizations," Anim. Behav. **45**, 423–433.
Hauser, M. D., and Fitch, W. T. (**2003**). "What are the uniquely human components of the language faculty?," in *Language Evolution*, edited by M. Christiansen and S. Kirby (Oxford University Press, Oxford), pp. 158–181.
Hauser, M. D., and Schön Ybarra, M. (**1994**). "The role of lip configuration in monkey vocalizations: Experiments using xylocaine as a nerve block," Brain Lang **46**, 232–244.
Hienz, R. D., Jones, A. M., and Weerts, E. M. (**2004**). "The discrimination of baboon grunt calls and human vowel sounds by baboons," J. Acoust. Soc. Am. **116**, 1692–1697.
Ives, D., Smith, D. R., and Patterson, R. D. (**2005**). "Discrimination of speaker size from syllable phrases," J. Acoust. Soc. Am. **118**, 3816–3822.
Janik, V. M., and Slater, P. B. (**1997**). "Vocal learning in mammals," Advances in the study of behavior **26**, 59–99.
Ladefoged, P. (**2001**). *Vowels and Consonants: An Introduction to the Sounds of Languages* (Blackwell, Oxford).
Lieberman, P. (**1968**). "Primate vocalization and human linguistic ability," J. Acoust. Soc. Am. **44**, 1574–1584.
Lieberman, P., and Blumstein, S. E. (**1988**). *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge University Press, Cambridge, UK).
Liu, R. C., Miller, K. D., Merzenich, M. M., and Schreiner, C. E. (**2003**). "Acoustic variability and distinguishability among mouse ultrasound vocalizations," J. Acoust. Soc. Am. **114**, 3412–3422.
Marler, P., and Slabbekoorn, H. (**2004**). *Nature's Music: The Science of Birdsong* (Academic, New York).
Martin, W. F. (**1971**). "Mechanics of sound production in toads of the genus *Bufo*: Passive elements," J. Exp. Zool. **176**, 273–294.
McComb, K. E. (**1991**). "Female choice for high roaring rates in red deer, *Cervus elaphus*," Anim. Behav. **41**, 79–88.
Moore, F. R. (**1990**). *Elements of Computer Music* (Prentice Hall, Engelwood Cliffs, NJ).
Moorer, J. A. (**1979**). "The use of linear prediction of speech in computer music applications," J. Audio Eng. Soc. **27**, 134–140.
Nelson, D. A., and Marler, P. (**1989**). "Categorical perception of a natural stimulus continuum: Birdsong," Science **244**, 976–978.
Nowicki, S. (**1987**). "Vocal tract resonances in oscine bird sound production: Evidence from birdsongs in a helium atmosphere," Nature (London) **325**, 53–55.
Nowicki, S., Mitani, J. C., Nelson, D. A., and Marler, P. (**1989**). "The communicative significance of tonality in birdsong: Responses to songs produced in helium," Bioacoustics 2, 35–46.
Owren, M. J. (**1990a**). "Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans. I. Natural calls," J. Comp. Psychol. **104**, 20–28.
Owren, M. J. (**1990b**). "Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans. II. Synthetic calls," J. Comp. Psychol. **104**, 29–40.
Owren, M. J., and Bernacki, R. (**1988**). "The acoustic features of vervet monkey (*Cercopithecus aethiops*) alarm calls," J. Acoust. Soc. Am. **83**, 1927–1935.
Owren, M. J., and Bernacki, R. H. (**1998**). "Applying linear predictive coding (LPC) to frequency-spectrum analysis of animal acoustic signals," in *Animal Acoustic Communication: Sound Analysis and Research Methods*, edited by S. L. Hopp, M. J. Owren, and C. S. Evans (Springer, New York), pp. 130–162.
Owren, M. J., Seyfarth, R. M., and Cheney, D. L. (**1997**). "The acoustic features of vowel-like *grunt* calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and functions," J. Acoust. Soc. Am. **101**, 2951–2963.
Pinker, S., and Jackendoff, R. (**2005**). "The faculty of language: what's special about it?," Cognition **95**, 201–236.
Poremba, A., Malloy, M., Saunders, R. C., Carson, R. E., Herscovitch, P., and Mishkin, M. (**2004**). "Species-specific calls evoke asymmetric activity in the monkey's temporal poles," Nature (London) **427**, 448–451.
Rand, A. S., and Dudley, R. (**1993**). "Frogs in helium: The anuran vocal sac is not a cavity resonator," Physiol. Zool. **66**, 793–806.
Rauschecker, J., and Tian, B. (**2004**). "Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey," J. Neurophysiol. **91**, 2578–2589.
Rauschecker, J. P., Tian, B., and Hauser, M. (**1995**). "Processing of complex sounds in the macaque nonprimary auditory cortex," Science **268**, 111–114.
Reby, D., and McComb, K. (**2003**). "Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags," Anim. Behav. **65**, 519–530.
Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., and Clutton-Brock, T. (**2005**). "Red deer stags use formants as assessment cues during intrasexual agonistic interactions," Proc. R. Soc. London, Ser. B **272**, 941–947.
Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947–950.
Rendall, C. A. (**1996**). "Social communication and vocal recognition in free-ranging rhesus monkeys (*Macaca mulatta*)," University of California, Davis.
Rendall, D. (**2005**). "Pitch (Fo) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry," J. Acoust. Soc. Am. **117**, 944–955.
Rendall, D., Owren, M. J., and Rodman, P. S. (**1998**). "The role of vocal tract filtering in identity cueing in rhesus monkey (Macaca mulatta) vocalizations," J. Acoust. Soc. Am. **103**, 602–614.
Rendall, D., Rodman, P. S., and Emond, R. E. (**1996**). "Vocal recognition of individuals and kin in free-ranging rhesus monkeys," Anim. Behav. **51**, 1007–1015.
Richman, B. (**1976**). "Some vocal distinctive features used by gelada monkeys," J. Acoust. Soc. Am. **60**, 718–724.
Riede, T., Bronson, E., Hatzikirou, H., and Zuberbühler, K. (**2005**). "Vocal production mechanisms in a non-human primate: Morphological data and a model," J. Hum. Evol. **48**, 85–96.
Riede, T., and Fitch, W. T. (**1999**). "Vocal tract length and acoustics of vocalization in the domestic dog *Canis familiaris*," J. Exp. Biol. **202**,

2859–2867.
Roberts, L. H. (**1975**). "The rodent ultrasound production mechanism," Ultrasonics **13**, 83–88.
Romanski, L. M., Averbeck, B. B., and Diltz, M. (**2005**). "Neural representation of vocalizations in the primate ventrolateral prefrontal cortex," J. Neurophysiol. **93**, 734–747.
Seyfarth, R. M., and Cheney, D. L. (**1990**). "The assessment by vervet monkeys of their own and another species' alarm calls," Anim. Behav. **40**, 754–764.
Seyfarth, R. M., Cheney, D. L., and Marler, P. (**1980**). "Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication," Science **210**, 801–803.
Sinnott, J. M., and Kreiter, N. A. (**1991**). "Differential sensitivity to vowel continua in Old World monkeys (Macaca) and humans," J. Acoust. Soc. Am. **89**, 2421–2429.
Sinnott, J. M., and Williamson, T. L. (**1999**). "Can macaques perceive place of articulation from formant transition information?," J. Acoust. Soc. Am. **106**, 929–937.
Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H., and Irino, T. (**2005**). "The processing and perception of size information in speech sounds," J. Acoust. Soc. Am. **117**, 305–318.
Smithson, T. R. (**1989**). "The earliest known reptile," Nature (London) **342**, 676–678.
Snowdon, C. T. (**1982**). "Linguistic and psycholinguistic approaches to primate communication," in *Primate Communication*, edited by C. T. Snowdon, C. H. Brown, and M. R. Petersen (Cambridge University Press, New York), pp. 171–211.
Sommers, M. S., Moody, D. B., Prosen, C. A., and Stebbins, W. C. (**1992**). "Formant frequency discrimination by Japanese macaques (*Macaca fuscata*)," J. Acoust. Soc. Am. **91**, 3499–3510.
Suthers, R. A., and Hector, D. H. (**1988**). "Individual variation in vocal tract resonance may assist oilbirds in recognizing echoes of their own sonar clicks," in *Animal Sonar: Processes and Performances*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 87–91.
Tartter, V. C. (**1991**). "Identifiability of vowels and speakers from whispered syllables," Percept. Psychophys. **49**, 365–372.
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice Hall, Englewood Cliffs, NJ).
Uppenkamp, S., Johnsrude, I. S., Norris, D., Marslen-Wilson, W., and Patterson, R. D. (**2006**). "Locating the initial stages of speech-sound processing in human temporal cortex," Neuroimage **31**, 1284–1296.
von Kriegstein, K., Warren, J., Ives, D., Patterson, R. D., and Griffiths, T. D. (**2006**). "Processing the acoustic effect of size in speech sounds," Neuroimage (to be published).
Wang, X. (**2000**). "On cortical coding of vocal communication sounds in primates," Proc. Natl. Acad. Sci. U.S.A. **97**, 11843–11849.
Weisz, D. J., Yang, B. Y., Fung, K., and Amirali, A. (**2001**). "The mechanism of ultrasonic vocalization in the rat," Abstr. Soc. Neurosci. **27**, 88.19.

Enhancing and unmasking the harmonics of a complex tone

William M. Hartmann[a] and Matthew J. Goupell
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

(Received 23 June 2005; revised 22 June 2006; accepted 23 June 2006)

Alternately eliminating and reintroducing a particular harmonic of a complex tone can cause that harmonic to stand out as a pure tone—separately audible from the rest of the complex-tone background. In the psychoacoustical literature the effect is known as "enhancement." Pitch matching experiments presented in this article show that although harmonics above the 10th are not spectrally resolved, harmonics up to at least the 20th can be enhanced. Therefore, resolution is not required for enhancement. Further, during those experimental intervals in which a harmonic is eliminated, excitation pattern models suggest that listeners should be able to hear out a neighboring harmonic—separately audible from the background. The latter effect has been called "unmasking." In the present article we provide the first experimental evidence for unmasking. Harmonics of 200 Hz, with harmonic numbers between about 5 and 16, are readily unmasked. Their pitches are usually matched by sine tones with frequencies that are not exactly those of the unmasked harmonics but are shifted in a direction away from the frequency of the pulsed harmonic. Phase relationships among the harmonics that produce temporally compact cochlear excitation lead to reduced enhancement but greater unmasking.
[DOI: 10.1121/1.2228476]

PACS number(s): 43.66.Hg, 43.66.Ba, 43.66.Dc [AJO] Pages: 2142–2157

I. INTRODUCTION

Demonstration number one on the compact disc *Auditory Demonstrations* (Houtsma, Rossing, and Wagenaars, 1987) is called "Cancelled Harmonics." In this demonstration a particular harmonic of a complex tone is alternately turned off (cancelled) and turned on, i.e., pulsed. As a result, that particular harmonic is heard as an independent entity standing out from the complex-tone background. Rather surprisingly, this harmonic can often be heard for an indefinitely long time after finally being turned back on. The effect is not new. Helmholtz (1877) quotes Seebeck's remark that the duration of the percept depends on the "… liveliness of our recollection of the tones heard separately." In the more modern literature the electronic version of the technique was attributed to Schouten by Cardozo (1967), and the pulsed harmonic is said to be "enhanced" (Viemeister, 1980; Viemeister and Bacon, 1982; Summerfield *et al.*, 1987). By this technique, many harmonics of a complex tone can be made individually audible. By contrast, normal listening to the complex tones of speech or music does not reveal individual harmonics. Instead, the harmonics are collectively absorbed into the global property of tone color or timbre.

The Cancelled Harmonics demonstration found on the *Auditory Demonstrations* disc presents a complex tone with a fundamental frequency of 200 Hz and 20 harmonics of equal amplitude. In order of increasing harmonic number, the first ten harmonics are turned off and on in 7-interval sequences. Thus the maximum harmonic frequency exposed in this way is $10 \times 200 = 2000$ Hz.

The fact that enhancement can expose harmonics up to the tenth can be contrasted with the harmonic identification task of Plomp and Mimpen (1968), which showed that harmonics could be resolved up to a limit between the fifth and seventh harmonics. *A priori*, it seems possible that the cancelled harmonic demonstration is just a less demanding version of the same thing—capable of exposing harmonics as high as the tenth because they are, at least to some extent, resolvable. The tenth harmonic was recently identified as the limit in the resolution and pitch experiments of Bernstein and Oxenham (2003). That view would suggest that the enhanced harmonic effect might not work much beyond the tenth, though pitch matching experiments by Gibson (1971) found an effect for harmonics as high as the 11th or 12th. By contrast, Yes-No detection experiments by Goupell *et al.* (2003) and Goupell and Hartmann (2004) showed that some trace of a pulsed harmonic could be detected for harmonic numbers as high as 69.

Because of the huge difference in experimental results from different experimental methods, we decided to try the pitch matching method. The pitch matching task is demanding because there is no prior constraint on the matching frequency. If a listener can successfully match the pitch of a pulsed harmonic then that is good evidence that the harmonic is audible.

II. EXPERIMENT 1—SEVEN-INTERVAL SEQUENCE—SINE PHASE

Experiment 1 followed the format of the compact disc's Cancelled Harmonics demonstration, making a selected harmonic separately audible by pulsing it off and on in a seven-interval sequence, as shown by the spectrogram in Fig. 1. Listeners were required to match the pitch of the harmonic

[a]Corresponding author. 1266 BPS Building, Michigan State University, East Lansing, Michigan 48824. Electronic mail: hartmann@pa.msu.edu

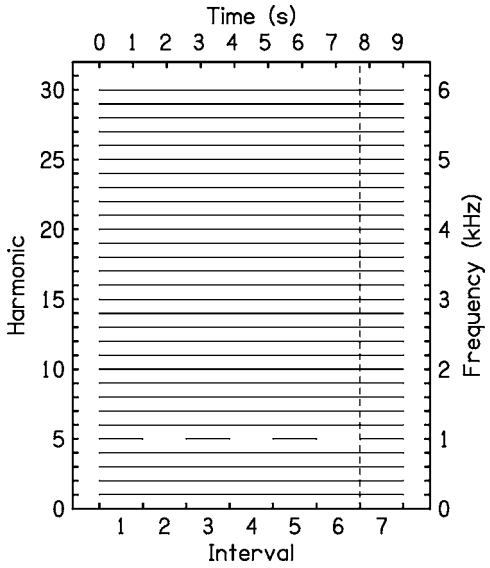


FIG. 1. Spectrogram showing the the seven intervals of the complex stimulus tone when the fifth harmonic was pulsed on and off. The tone was actually continuous for 9.171 s. Only the fifth harmonic was pulsed. The dashed line shows where a six-interval sequence ends.

that was exposed in the sequence. Because the sequence ended with the pulsed harmonic turned on, the sequence naturally cued the enhanced harmonic.

A. Stimulus

As with the compact disc demonstration, the tone, with fundamental frequency f_0, was continuous throughout the sequence. Only the selected harmonic was pulsed. The tone consisted of the first 30 harmonics with equal amplitudes. Fundamental frequency f_0 was nominally 200 Hz, but on different trials it was randomized over a ±5% range with a rectangular distribution as a guard against interaction among successive trials. The harmonics were added in *sine phase*.

The complex-tone sequence was computed in a Tucker-Davis array processor (AP2), converted by a 16-bit DAC (DD1), and given 10-ms raised-cosine edges (SW2). The signal was then lowpass filtered with a corner frequency of 20 kHz and a rolloff of −115 dB/octave. Each of the seven intervals of the sequence had a duration of 1310.72 ms, equivalent to two buffers converted at a sample rate of 50 kHz. Therefore, the duration of the seven-interval sequence was about 9.2 s. The selected harmonic was on or off for the entire duration of an interval, as shown in Fig. 1.

The stimulus was presented diotically at a level of 60 dB SPL (45 dB SPL for each component) through Etymotic ER2 insert earphones. The insert earphones provided an acceptably flat frequency response for the wide range of frequencies of interest in this experiment. Listeners were tested individually in a double-walled sound attenuating room.

B. Procedure

Four male listeners participated in the experiment: B, M, W, and Z. All listeners were between the ages of 21 and 25, except for W, who was 65. Listeners M and W were the authors. An experimental run consisted of 10 trials. On a low-frequency run one harmonic was selected on each trial, chosen randomly from the range 1–10. In a high-frequency run, the range was 11–20. Final data were based on ten runs, i.e., five matches for each harmonic, 1–20.

To begin a trial, the listener pressed the green button on the response box. The computer responded by selecting a harmonic to be pulsed off and on and then playing the seven-interval sequence. Then after a 300-ms silent gap, the computer began a series of matching sine tones (WG2) pulsed on and off (on for 1 s, off for 0.3 s) with 10-ms raised cosine edges (SW2). The series of matching tones continued indefinitely. The listener could vary the frequency of the matching tone by using up/down push buttons to establish a range (two octaves) and a ten-turn potentiometer for fine control within that range. The listener could adjust the level of the matching tone—from inaudible to loud—using a second ten-turn potentiometer, and could mute the matching tone with the blue push button. To hear the seven-interval sequence again, with the same f_0 and same pulsed harmonic number, the listener pressed the red button. Then after a 300-ms gap, the seven-interval sequence was presented once again, followed by a series of matching tones as before. There was no limit to the number of seven-interval sequences or matching tones that could be heard for any trial.

By this procedure the listener converged on a satisfactory matching frequency. When the listener was satisfied with his match, he pressed the green button again to end the trial. The computer then recorded the matching frequency for the trial. Alternatively, the listener could press the white button to abort the trial if he decided that it was impossible to make a match. Such aborted trials were indicated in the data as "no-match." When the green button was pressed again, the next trial began, with a different pulsed harmonic number and a somewhat different f_0. A run of ten trials required as little as six minutes and as much as 30 for completion, depending on the listener.

C. Results

The results of Experiment 1 are shown in Fig. 2 for listeners B, M, and W (listener Z did not participate in this experiment). The figures show the matching error, namely the difference between the matching frequency and the pulsed harmonic frequency as a percentage of the pulsed harmonic frequency. The matching error is shown as a function of the pulsed harmonic number, n, ranging from 1 to 20.

Figure 2 shows each individual match using five different symbols for the different runs. The solid line at zero error shows perfect matching. The dashed lines show the error corresponding to a match at harmonic $n+1$ or $n-1$. Numbers at the top of each plot show the number of failed responses. These are mostly no-match responses (30/33), but there are three matches (listener B) off the chart, i.e., matching attempts outside the range of the graphs. The string of "5"s for harmonic 16 and higher for listener M are all no-matches because M did not hear a harmonic standing out. The lowest numbered matching failure occurred on one match to pulsed harmonic 11 (listener B). Apart from that single failure at harmonic 11, all five matches for all three listeners for every

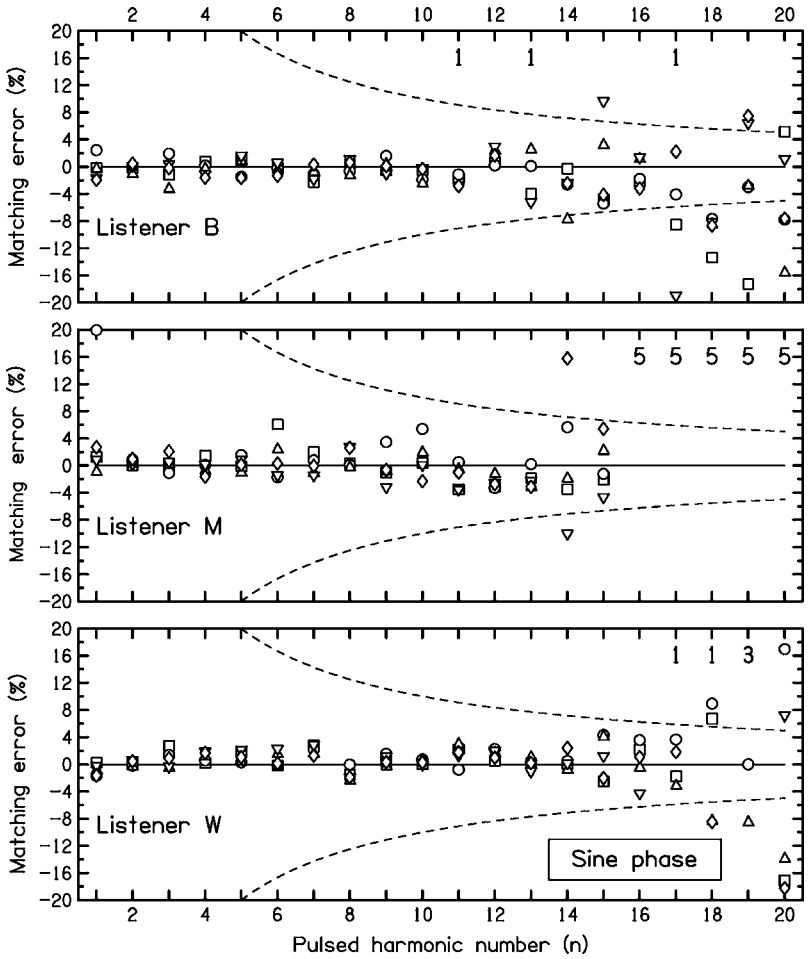


FIG. 2. Experiment 1 pitch matching error in percent showing the discrepancy between the frequency of the matching tone and the frequency of the pulsed (enhanced) harmonic for the seven-interval sequence when all harmonics were presented in sine phase. Each of three listeners, B, M, and W, made five matches. All matches are shown. Matches with common symbols were made in the same experimental run. Dashed lines indicate the frequencies of harmonics $n+1$ and $n-1$, given by the formula $\pm 100/n$. Numbers at the top of the plots indicate unsuccessful matches, either off the chart or no-matches.

pulsed harmonic from 1 to 12, were closer to the pulsed-harmonic frequency than to the frequency of any other harmonic.

D. Discussion

Experiment 1 served to validate the pitch-matching method. Harmonics up to the tenth, and somewhat beyond, could always be successfully matched, in rough agreement with the results of Bernstein and Oxenham (2003). A match closer to the pulsed harmonic than to any other harmonic is shown by a symbol that is closer to the solid line in Fig. 2 than to either dashed line. However, this criterion for the success of matches raises the spectre of spectral pitch shifts, where the pitch of a harmonic in the context of a complex tone is different from the pitch of a sine tone with that frequency (Terhardt, 1971). Except for matches by listener M for harmonics 11, 12, and 13, that run systematically flat, and matches by W for harmonic 7, that run sharp, there is no good evidence for such shifts. Similarly, the pitch-matching and forced-choice experiments of Peters *et al.* (1983) and the zero-mistuning results of Hartmann and Doty (1996) showed no shifts.

Although Experiment 1 did not approach very high harmonic numbers, the results shown in Fig. 2 exhibit large scatter for pulsed harmonics above 16, suggesting that the enhancement effect is limited to harmonics less than about 16, at least for sine phases as used in this experiment. Also, the large number of no-matches above harmonic 16 would seem to be in conflict with the conclusion of Goupell and Hartmann (2004) that pulsed harmonics up to 69 can be detected. However, the latter work was done with harmonics having random phases. The procedure there allowed listeners to take advantage of other phase relationships among harmonics. By contrast, Experiment 1 was restricted to sine phases. To look for the role of relative phases in the enhancement experiment, we performed Experiment 2, which randomized phases. We also expected that comparing the results of Experiments 1 and 2 would lead to relevant insight into the resolution of the harmonics.

III. EXPERIMENT 2—SEVEN-INTERVAL SEQUENCE—RANDOM PHASE

There were four listeners in Experiment 2, B, M, W, and Z—the same as in Experiment 1, plus listener Z.

A. Stimulus

Experiment 2 was identical to Experiment 1 except that every time the listener pressed the red button he was presented with a new randomization of the phases. As phases were rerandomized, the value of f_0 was not changed, though, as in Experiment 1, f_0 was randomized across different trials. For high pulsed-harmonic numbers listeners often took advantage of the randomization by pressing the red button several times to obtain a "good" set of phases for matching.

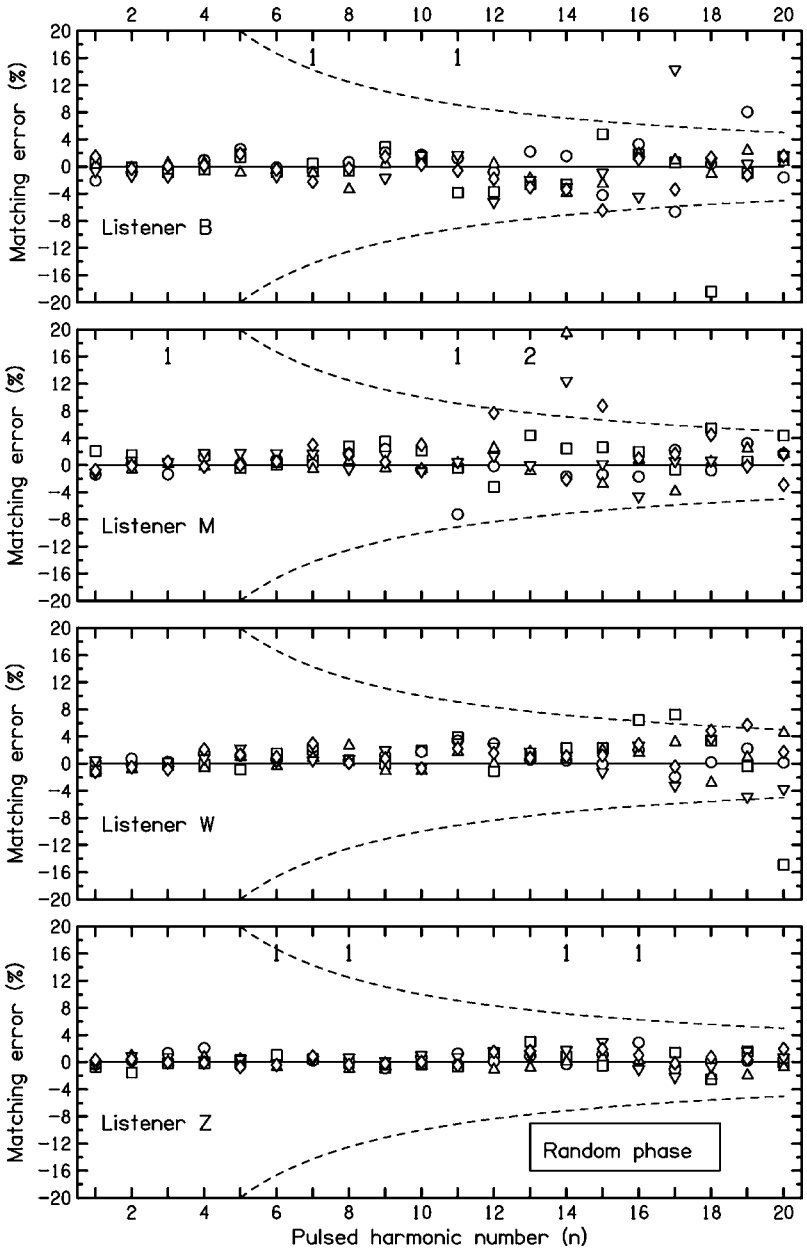


FIG. 3. Experiment 2 pitch matching error, as in Fig. 2. In Experiment 2 listeners could hear as many different random phase relationships as they wanted before making a pitch match. A fourth listener, Z, was added to the group of listeners.

B. Results

The results of Experiment 2 are shown in Fig. 3. Here matching performance was more successful than for Experiment 1 shown in Fig. 2. In contrast to Experiment 1 (sine phases) with 11% of failed matches, Experiment 2 (random phases) led to a failure rate of only 3%, and none of these were no-matches. Listeners B and Z successfully matched the 20th harmonic (ten times out of ten). Matches that are not failures are said to be "on the chart," and they can be compared for listeners B, M, and W, who were in both Experiments 1 and 2. In Experiment 1, 10% of on-the-chart matches fell outside the dashed lines. By contrast, in Experiment 2, only 3% of them fell outside those lines. Listener M, who could not hear pulsed harmonics above 15 for the sine phase, successfully matched out to harmonic 20 when given the chance to choose a phase relationship.

According to informal reports from listeners, randomizing the phases led to important perceptual differences in the high-frequency runs, where harmonics 11–20 were pulsed. The listeners observed that some phase relationships led to enhanced harmonics with a clear tonal character—easy to match. Other phase relationships led to enhanced harmonics (or enhanced spectral regions) that were buzzy in character and hard to match, and still other phases led to no enhanced sensation at all.

Relative phases are expected to be important for high frequencies, where harmonics are poorly resolved. One measure of the minimum harmonic number for which phases play a role can be found in comparing the data for B, M, and W in Fig. 3 (random phases) with their data in Fig. 2 (sine phase). To make the comparison quantitative, we counted the number of matches that fell outside a central band of ±4%. We found no systematic differences between random phase and sine phase for pulsed harmonics less than 14. But for all harmonics from 14 to 20 the number outside the central band was twice as large for sine phase as for random phase. This

measure, with its ±4% limits, was chosen because it gives a quantitative value to an effect that is clear to the eye in comparing Figs. 2 and 3. This measure points to harmonic 14 as the onset of *dramatic* phase dependence. As another measure of onset, the authors (listeners M and W) listened to ten different phase relationships in an experiment like Experiment 2, and rated the clarity of enhanced harmonics on a scale of 0 (buzzy) to 5 (tonal). Pulsed harmonics below 8 received scores of 5. Pulsed harmonics 8 and above received variable scores. For example, the scores for harmonic 8 ranged from 2 to 5. This measure points to harmonic 8 as the onset of phase dependence.

C. Discussion

A comparison of the data in Figs. 2 and 3 provides good evidence that randomizing the phases allows listeners to find phase relationships that give better tonal sensations than the sine phase, at least for pulsed-harmonic numbers greater than 13. Records were kept of the number of seven-interval sequences heard by listeners. For Experiment 1, listeners heard an average of 3.3 sequences per trial (i.e., per match). For Experiment 2, where the listeners could find a favorable phase condition, the number of sequences per trial increased to 5.6 for the three listeners common to Experiment 1, and 6.2 when the data from listener Z were added. The number of randomizations, 6.2, is not large. Repeated randomizations were discouraged by the fact that each request for a new randomization produced another stimulus sequence, adding 9.2 s to the trial. Because the number of rerandomizations was small, whatever phase relationship is responsible for strong enhancement cannot be very special in the way that sine phases are special. However, it was clear that even among the few phase relationships actually explored on any trial some were more favorable than others.

The results of Experiment 2, especially the outstanding performance of listener Z, indicate that with nonspecial phases, harmonics up to the 20th can be heard out. Although no matching experiments were done above the 20th harmonic, nothing in the results of Experiment 2 contradicts the randomized-phase results from the Yes-No experiment that showed an audible effect beyond the 60th (Goupell and Hartmann, 2004). However, pulsed harmonics as high as 60 lead to a buzz sensation and not a pure-tone sensation. For that reason, and because pitch perception becomes poor above 5000 Hz, pitch matching much above the 20th harmonic is likely to be inaccurate.

The results of Experiment 2 can be compared with those of Bernstein and Oxenham (2003), who used signal levels similar to ours and also randomized phases. Their experiments found little effective audibility above the tenth harmonic, which they interpreted as indicating that harmonics above the tenth are not resolved. Our experiments tend to agree about resolution, but place the limit between 8 and 14. In our interpretation, that region indicates the upper limit for resolution because that is where important monaural phase effects begin.

The limit between harmonics 8 and 14 can be compared with the onset of phase dependence in fundamental frequency discrimination for 200 Hz complex tones, as measured by Houtsma and Smurzynski (1990), who found a negligible phase dependence when the lowest harmonic was 10, considerable phase dependence when the lowest harmonic was 13, and dramatic phase dependence when the lowest harmonic was greater than 15. Similarly, Shackleton and Carlyon (1994) found a transition to phase dependence in pitch perception when they tested the band including harmonics 11 through 15. Moore *et al.* (2006) found phase effects in fundamental frequency discrimination for harmonics as low as 8, consistent with the onset of phase effects reported here.

The experiments of Bernstein and Oxenham (2003) indicated that the frequencies of pulsed harmonics could be discriminated, to within 3.5–5%, up to about the 10th harmonic, whereas our experiments found successful matching within similar percentage limits up to the 20th, given random phase relationships. The difference between the results is probably related to the experimental techniques. The Bernstein-Oxenham complex tones were brief single shots, one second in duration, but our trials permitted listeners to listen as many times as desired to the 9.2 s sequence. For reasons unrelated to the experiment at hand, Bernstein and Oxenham added a noise background, but our experiments were done in quiet. Their listeners received no feedback, but our listeners were trained over the months of experimenting and received feedback on early runs. On early runs, the ratio of the matching frequency to the pulsed harmonic frequency was displayed on the response box after a match had been made.

Although the differences in results obtained by Bernstein-Oxenham and in our experiments can be attributed to differences in methods, the results lead to different conclusions about the enhancement effect itself. Bernstein and Oxenham regarded an enhancement experiment as a measurement of peripheral resolvability. They found an upper limit at the tenth harmonic and observed that this result is consistent with experiments on the highest numbered harmonics that are useful in producing a strong virtual pitch. Bernstein and Oxenham discounted more central effects whereby an enhanced harmonic behaves as though it were physically more intense, e.g., by the adaptation of suppression as invoked by Viemeister and Bacon (1982). By contrast, the work presented here also suggests that the limit of peripheral resolvability occurs near the 10th harmonic, but the enhancement effect persists at least up to the 20th. Consequently, our experiments suggest a role for a process beyond peripheral resolution that effectively boosts the level of an enhanced harmonic, at least relative to other harmonics. Evidence for such a relative boost, both in firing rate and in synchrony, was found in auditory nerve recordings by Palmer *et al.* (1995).

IV. EXPERIMENT 3—SEVEN-INTERVAL SEQUENCE—SCHROEDER PHASE

The phase effect seen for harmonic numbers 11–20 in Experiments 1 and 2 was intriguing. Clearly some random-phase relationships proved to be more effective than sine-phase in enhancing a harmonic. However, Experiment 2 did

not keep track of the successful and unsuccessful phases. Therefore, to try to discover the nature of the phase effects, we performed Experiment 3 in which the relative harmonic phases were controlled.

Experiment 3 used three different phase relationships, sine phase, and two forms of Schroeder phase, *m*+ and *m*−, as defined in Eq. (1). The sine phase leads to a waveform with peaks in tuned channels that contain several harmonics. However, it is expected that dispersion in the inner ear results in cochlear excitation that is less peaked than the waveform itself at the high-frequency place corresponding to those channels. The *m*+ phase compensates for the cochlear dispersion and increases the temporally peaked character at the high-frequency place. The *m*− phase relationship has properties similar to the *m*+ phase signal in that both tend to be small-peak factor waveforms (Schroeder, 1970), but the curvature of the phase-versus-frequency function is opposite to that of *m*+. The temporally compact character of the *m*+ excitation has been observed in masking experiments that reveal dramatically smaller masking produced by *m*+ compared to *m*−. (Smith *et al.* 1986; Kohlrausch and Sander, 1995; Oxenham and Dau, 2001). A compact masker allows a signal to be detected, or glimpsed, during the "quiet" portions of the masker.

A. Method

Experiment 3 was identical to Experiment 1 in that a fixed phase relationship was used for each experimental run. Listeners did five runs per phase condition. Because phase effects were mainly seen in the high-frequency runs of Experiments 1 and 2, only harmonics 11 through 20 were pulsed on and off in Experiment 3.

The sine-phase runs of Experiment 3 were in every way identical to the high-frequency runs of Experiment 1. The Schroeder-phase runs added a phase shift for harmonic n given by

$$\phi_n = \pm \pi n(n-1)/N, \tag{1}$$

where N equals 30, the maximum harmonic number in the spectrum. The plus sign corresponds to *m*+ and the minus sign to *m*−. The sine phase has no added phase shift ($\phi_n = 0$) and can be designated "*m*0." Listeners, B, M, W, and Z, participated.

B. Results

The results of Experiment 3 are shown in Fig. 4, where matching data are plotted for all three phase relationships. Open symbols indicate the average of five matches. Filled symbols indicate the average of four or three matches, which occurred when there were one or two no-matches or matches off the chart. If fewer than three matches were on the chart, then no symbol is plotted. Several measures were used to compare results for the different phase relationships.

(1) *No-matches*: The number of no-matches is an indication of the difficulty of hearing the enhanced harmonic. All four listeners indicate the same ordering, with *m*+ leading to the greatest number of no-matches and *m*− leading to the least. Summed over the listeners, the numbers of no-matches were as follows: For *m*+, 42; for *m*0, 26; and for *m*−, only 1.

(2) *Standard deviations*: The standard deviations (SD) are shown in Fig. 4 by the error bars. A fair assessment of average SD can be obtained from the data for those listeners and those pulsed harmonic numbers for which there are five matches (out of a possible five) for each of the three phase relationships. Over the four listeners, there were 19 such cases (out of a possible 40). Averaged over the four listeners, the SD were as follows: For *m*+, 3.3% (±0.9); for *m*0, 2.7% (±2.4); for *m*−, 1.7% (±0.8). The large error for *m*0 resulted from an anomalously large SD for listener Z. For all listeners, the SD for *m*+ was larger than the SD for *m*−.

(3) *More standard deviations*: More SD data can be included by comparing phase conditions in pairs, where all five matches were made for both phases of a pair, and using a sign test. The SD for *m*+ was greater than the SD for *m*0 on 15 out of 20 such comparisons ($p \leqslant 0.05$). The SD for *m*+ was greater than the SD for *m*− on 17 out of 21 comparisons ($p \leqslant 0.01$). The SD for *m*0 was greater than the SD for *m*− on 21 out of 29 comparisons ($p \leqslant 0.05$).

(4) *Absolute deviations*: An absolute deviation is the distance, expressed in percent, between the mean matching frequency and the frequency of the pulsed harmonic, as shown in Fig. 4. Casual inspection shows that the squares (stimulus *m*+) are usually farther from the solid line at 0% than are the other symbols. Averaged over the listeners for the 19 cases with all matches on the chart, the absolute values of the deviations were as follows: For *m*+, 4.1% (±2.3); for *m*0, 2.1% (±0.7); for *m*−, 1.5% (±1.1). Thus, a detailed calculation agrees with casual observation.

C. Discussion

The four measures of pitch matching success presented in Sec. IV B all agree that the *m*− phase condition leads to the most successful performance, and the *m*+ condition leads to the least successful. Consequently, performance improves monotonically as the phase curvature goes from positive to zero to negative. The positive curvature condition, *m*+, is expected to lead to temporally compact excitation at high-frequency places on the basilar membrane (Smith *et al.*, 1986). The compact time dependence allows a signal to be glimpsed during the fractions of the cycle when the excitation is small. The zero-curvature condition, *m*0, with all sine phases, leads to a compact waveform, as seen on an oscilloscope, but is expected to be less compact at the high-frequency place. The *m*− phase condition can be expected to produce an even more uniform distribution of excitation throughout the stimulus cycle. In that sense, the *m*− condition is similar to typical random-phase conditions (Kohlrausch and Sander, 1995). Thus, the results obtained in Experiment 3 agree with what was learned in comparing Experiment 1 (sine phase) with Experiment 2 (random phase). The results support the suspicion from Experiment 2 that what is important about random phase is that it is not special like sine phase. Cochlear dispersion notwithstanding,

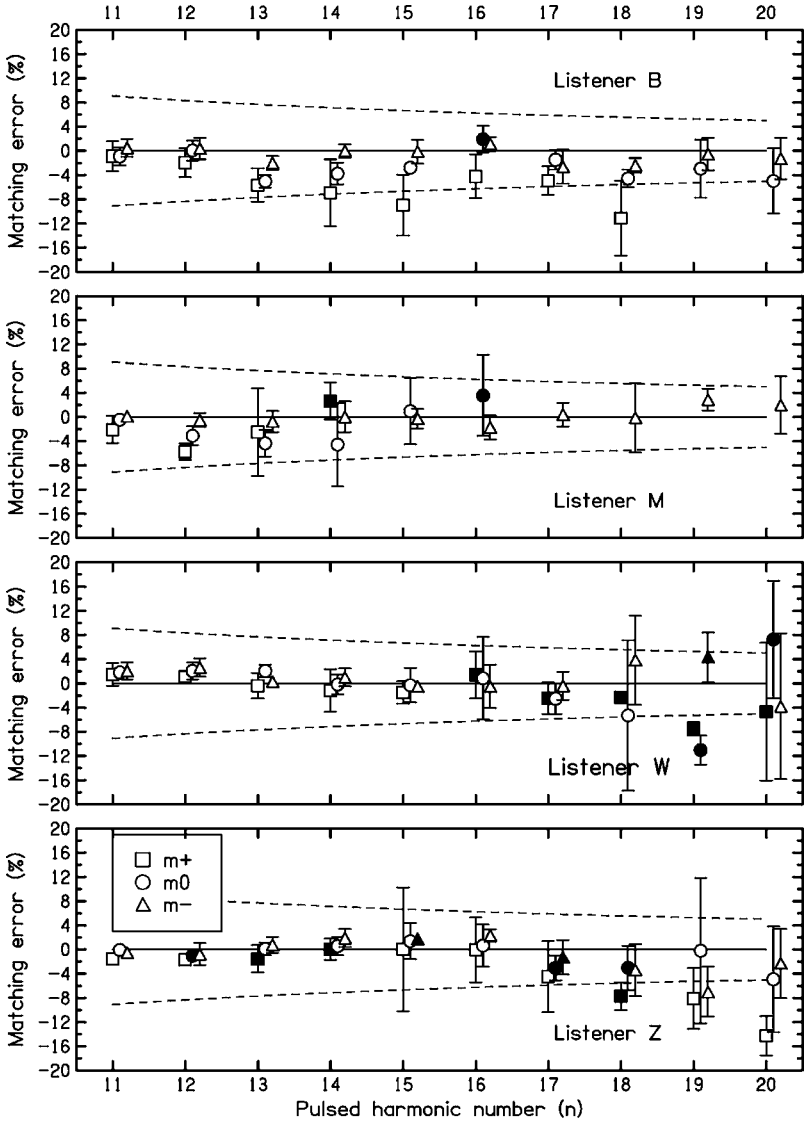


FIG. 4. Experiment 3 pitch matching error in percent, showing the mean discrepancy between the matching tone and the frequency of the pulsed harmonic for the seven-interval sequence with three phase relationships: squares for *m*+ (Schroeder plus), circles for *m*0 (sine phase), and triangles for *m*− (Schroeder minus). Open symbols indicate that all five matches are on the chart. Filled symbols indicate three or four matches. Error bars are two standard deviations in overall length.

the sine phase seems to retain some peaked character at the 2200–4000 Hz places that are probed by pulsing harmonics 11 through 20. The conclusion of this experiment is that enhancement is promoted if the excitation is not peaked in time but is relatively homogeneously distributed over the period of the stimulus. In short, what is good for release from masking (glimpsing) is bad for enhancement, and vice versa.

Because phase effects are important for higher harmonic numbers, a summary of Experiments 1, 2, and 3 for $11 \leqslant n \leqslant 20$ appears in Table I, which shows the number of matches closer to the frequency of n than to the frequency of any other harmonic. Ideally, the entries for the two sine-phase experiments should be the same, and they nearly are. The summary suggests that, overall, the benefits of randomizing the phases in Experiment 2 were all captured by fixing the phase at *m*− in Experiment 3.

V. EXPERIMENT 4—SIX-INTERVAL SEQUENCE—SINE PHASE

Experiments 1, 2, and 3 demonstrated enhancement, an effect that occurs when a harmonic amplitude changes from zero to full on. It is reasonable to expect that the enhancement effect would be similarly present if the harmonic amplitude varied from some small value to full on, or if some other dramatic increase in the amplitude were to occur. A particularly interesting increase in effective amplitude—not *a priori* evident—is suggested by excitation pattern models such as that proposed by Terhardt *et al.* (1982a, b). This excitation pattern model is nonlinear in that the effective strength of each harmonic is reduced by the strengths of

TABLE I. Summary of Experiments 1, 2, and 3—enhancement for unresolved harmonics. For four listeners (B, M, W, and Z) entries in the table show the number of matches that were closer in frequency to the pulsed harmonic (n) than to any other harmonic. The denominator in the Total row indicates the maximum possible total. The table includes only results for $11 \leqslant n \leqslant 20$.

Listener	Expt 1 sine	Expt 2 random	Expt 3 *m*+	Expt 3 sine	Expt 3 *m*−
B	4	7	2	4	9
M	3	4	1	2	6
W	4	6	5	5	6
Z	NA	10	3	4	6
Total	11/30	27/40	11/40	15/40	27/40

neighboring harmonics. This mutual masking, or suppression, is most effective in the upward direction, i.e., the strength of a harmonic is reduced primarily by the harmonic immediately below it. The strength of a harmonic is less affected by the harmonic immediately above it.

A consequence of this nonlinear excitation model is that when a harmonic is turned off, as in a pulsed harmonic experiment, its masking or suppression effect on neighboring harmonics disappears. Therefore, the model predicts an effective boost in the strength of the neighboring harmonics, particularly the harmonic immediately above. Thus turning off harmonic n is predicted to lead to a boost in the effective amplitude of harmonic $n+1$. The boost might be adequate to "pop out" harmonic $n+1$, just as harmonic n pops out when that harmonic is turned on. Exposing harmonic $n+1$ by turning off harmonic n has been called "unmasking" (Hartmann, 1997). The effect has been observed informally, but it has never been experimentally verified. The purpose of Experiment 4 was to try to find clear evidence for the predicted unmasking effect.

Experiment 4 used a six-interval sequence intended to facilitate the matching of an unmasked harmonic. There were four listeners in Experiment 4, the same who participated in Experiments 2 and 3.

A. Stimulus

The six-interval sequence of Experiment 4 was identical to the seven-interval sequence of Experiment 1 except that the final interval was omitted, terminating the sequence at the dashed line shown in Fig. 1. Therefore, the sequence duration was 7.8 s. On the sixth and final interval, the pulsed harmonic was off, leaving the listener with the impression of the unmasked harmonic—if any. The listener's task was to match what he heard standing out on the sixth interval of the sequence.

B. Results

The results of Experiment 4 are shown in Fig. 5, which shows the difference between the matching frequency and the frequency of the pulsed harmonic (harmonic n). The vertical scale limits are ±50% of the pulsed harmonic frequency, in contrast to Figs. 2–4 with limits of ±20%. The solid lines show the frequencies of harmonics $n+1$ and $n-1$. They are solid because they indicate the matches that are expected from the theory of unmasking. Unlike the seven-interval experiments, matches are not expected at 0% difference, i.e., at harmonic n. The dashed lines in Fig. 5 show harmonics n, $n+2$, and $n-2$.

As before, numbers at the top of the graphs indicate no-matches and matches off the chart. The great majority of these responses, including all of them for listeners M and W, were no-matches.

1. Evidence for unmasking

For pulsed harmonics 4, 5, and 6, listeners B, M, and W matched perfectly, in the sense that their matches were always closer to harmonics $n+1$ or $n-1$ than to any other harmonic. Listener Z matched perfectly for pulsed harmonics 5–10, and listener W matched perfectly for pulsed harmonics 1–7.

Pulsing harmonic 1 would be expected to unmask the second harmonic and lead to a matching difference of 100%, off the chart in Fig. 5. Only listeners B and W attempted matches for this harmonic. B's matches were scattered, with a mean matching difference of 123% (±31). W's matches were nearly exact, with a mean matching difference of 99% (±1). Matches at $n-1$ for pulsed harmonic $n=2$ (listener B) have an ambiguous interpretation in that they are at the fundamental frequency corresponding to the pitch of the complex tone as a whole.

The most important result of Experiment 4 is that the hypothesized unmasking effect clearly exists. The experimental matches cluster in the vicinity of the unmasked harmonics as predicted. The odds against such clustering by accident are overwhelming.

2. Relative variability

Unmasking can be compared with enhancement. The variability of the matches to unmasked harmonics in Experiment 4 can be compared to the variability for enhanced harmonics in Experiment 1 for the three listeners common to both experiments. The comparison was made by forming the ratio of the standard deviations in the two experiments for those pulsed harmonics where all five matches were on the chart for both experiments. For Experiment 4, the absolute value of the deviation from f_n was taken to compensate for the bimodal character of the matches. For listener B there were 10 ratios for n between 1 and 18. For M there were 11 ratios for n between 4 and 15, and for W there were 12 ratios for n ranging from 1 to 12. Average ratios for B, M, and W were, respectively, 1.6, 2.2, and 2.4. (Because this statistic is a ratio, it is the same whether calculated as the ratio of frequency standard deviations or as the ratio of frequency standard deviations relative to the pulsed harmonic frequency.) The conclusion of the calculation is that the variability for matches to unmasked harmonics is about twice the variability for matches to enhanced harmonics.

C. Discussion

The data in Fig. 5 clearly demonstrate the unmasking of harmonics. Thus, the question that inspired this experiment is answered in the affirmative: The unmasking effect suggested by the nonlinear excitation pattern model exists. The pitch matches cluster well near the expected frequencies, corresponding to $n+1$ and $n-1$. The standard deviations among the matches to unmasked harmonics in the six-interval experiment are only about twice as large as the standard deviations among the matches to enhanced harmonics in the seven-interval experiment. The standard deviation can be taken as a measure of the relative salience of unmasking and enhancement. Thus, unmasking is found to be about half as salient as enhancement.

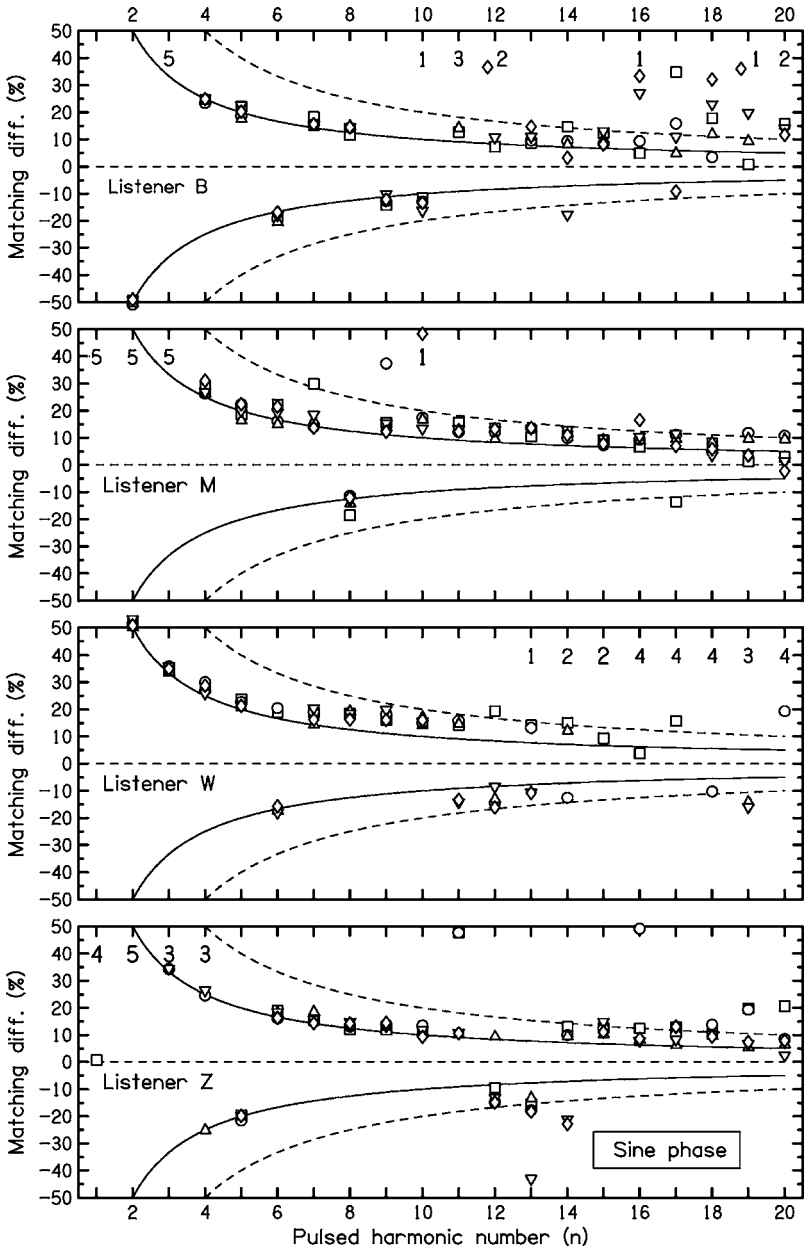


FIG. 5. Experiment 4 results, showing the mean difference between the frequencies of the matching tone and the pulsed harmonic for the six-interval sequence when all the harmonics were presented in sine phase. The solid lines show the frequency differences for expected matches at harmonics $n+1$ and $n-1$. Dashed lines show the frequency differences for $n+2$ and $n-2$ as well as n itself. Numbers at the tops of plots show unsuccessful matches, mostly no-matches but also a few matches off the chart.

1. Upward tendency

According to the unmasking hypothesis, there is mutual masking among the excitation patterns for the harmonics of a complex tone. Each harmonic is partially masked by its neighbors. When harmonic n is removed, the masking of a neighboring harmonic is reduced causing that neighboring harmonic to stand out. Because masking is expected to spread upward in frequency, one would expect that the harmonic made salient by unmasking would be the harmonic that is immediately above the harmonic that is pulsed, i.e., removing harmonic n should unmask harmonic $n+1$. That view is consistent with 82% of the matches for harmonics 3–20 in Fig. 5. Because upward spread of masking is a stronger effect for more intense maskers (Egan and Hake, 1950), the percentage would likely have been even larger if the experiment had been done at a higher intensity level. By contrast, some pulsed harmonics (n) systematically unmasked harmonic $n-1$, and the value of n for which this occurred was different for each listener: $n=2$, 6, 9, and 10 for B; $n=8$ for M; $n=12$ for W; and $n=5$, 12, and 13 for Z. The apparent idiosyncratic unmasking of individual harmonics is reminiscent of other individual differences—otoacoustic emissions and pitch shifts with intensity (Verschuure and van Meeteren, 1975). One is tempted to speculate that the effects are due to differences in the auditory periphery, perhaps in the cochlea. Monaural unmasking experiments that compared the two ears of a given listener would help check that conjecture, as would experiments with odd and even harmonics presented to opposite ears, *per* Bernstein and Oxenham (2003).

Whether the origin is cochlear or more central, it seems unlikely that unmasking results from an effective increase in mere firing rate without regard for timing. The clear pitch of some unmasked harmonics suggests a role for phase locking. Removing a harmonic expands the range of tonotopic axis that can be made synchronous with a neighboring harmonic.

A similar point was made by Moore and Ohgushi (1993) to explain the greater salience of harmonics that occur at spectral boundaries.

2. Pitch shifts

Because of the nature of the unmasking effect, one might expect it to be particularly susceptible to pitch shifts. Figure 5 shows that the great preponderance of shifts are upward—the unmasked harmonics are matched by sine-tone frequencies that are above the frequency of the unmasked harmonic. The exception to that rule occurs when the unmasked harmonic is $n-1$. For the nine cases of systematic $n-1$ matches noted above, the mean match was lower than f_{n-1} for eight.

Because of the positive pitch shifts, matches differ systematically from f_{n+1}. We believe that this is the reason that matches often fall between f_{n+1} and f_{n+2}, and often fall closer to the latter for pulsed harmonics 9–18. Above pulsed harmonic 18 the variation among the matches becomes large, even for the most reliable listeners M and Z.

3. Relationship to the Duifhuis pitch effect

The unmasking effect bears a superficial similarity to the Duifhuis pitch effect (DPE) in that turning off a harmonic causes a perceived tone to emerge (Duifhuis, 1970, 1971; Lin and Hartmann, 1997). However, the two effects are really very different. First, the DPE makes strong demands on the density of the harmonics. It has been studied for fundamental frequencies of 100 Hz and below, which leads to dense spectra at higher frequencies. Our informal attempts to hear the DPE with a fundamental of 200 Hz failed. Similarly, the DPE occurs only if the pulsed harmonic number is rather high. Depending on the listener, Duifhuis (1970) found a minimum harmonic number between 16 and 20. Lin and Hartmann found a minimum between 12 and 20. In contrast to these observations with the DPE, unmasking can be observed with a 200-Hz fundamental and with lower numbered pulsed harmonics. For example, all the listeners in Experiment 4 made 100-percent-successful matches when harmonic 5 was pulsed.

Second, the Duifhuis pitch corresponds to the frequency of the harmonic that is removed, harmonic n, but the unmasked pitch occurs at $n+1$ or $n-1$. Figure 5 makes it clear that the pitch is not at harmonic n.

A third feature of the DPE is that it requires a phase relationship that leads to a small-amplitude portion in the waveform. Duifhuis (1972) discovered that the pitch effect disappeared when phases were randomized. Lin and Hartmann agreed, though the DPE did occur for a sawtooth waveform (zero curvature). The disappearance of the DPE when phases are randomized presented another opportunity to draw a distinction between the DPE and the unmasked pitch. Experiment 5 studied unmasking with random phases.

VI. EXPERIMENT 5—SIX-INTERVAL SEQUENCE—RANDOM PHASES

Based on our experience with the enhancement effect, it seemed possible that exposing listeners to random phases would lead to a stronger unmasking effect and an even tighter distribution of matches than seen in Experiment 4. Alternatively, as suggested above, randomizing the phases might eliminate the effect, as it does for the DPE. There were three listeners in Experiment 5, B, M, and W—the same who participated in Experiment 1.

A. Stimulus

Experiment 5 was identical to Experiment 4, except that the phases of the harmonics were rerandomized, as in Experiment 2, every time the listener pressed the red button to request another six-interval sequence. Thus, Experiment 5 was different from Experiment 4 in the same way that Experiment 2 was different from Experiment 1.

B. Results

The results of Experiment 5 are shown in Fig. 6, which is identical in form to Fig. 5. Not shown are matches by B and W when the first harmonic was pulsed, leading to mean deviations of 118% (±25) for B and 98% (±1) for W. The expected matching deviation is 100%, and the results for B and W are similar to those in Experiment 4.

It is interesting to compare matches for sine phase and random phase for the three listeners who were common to Experiments 4 and 5. A comparison of Figs. 5 and 6 shows that listener B produced nine clusters of five matches for sine phase but only two such clusters for random phase. For listener M there are ten clusters for sine phase and none for random phase. For listener W there are nine clusters (including matches to the pulsed fundamental) for both sine and random phase conditions. All listeners reported that different phase relationships unmasked different harmonics, either above the pulsed harmonic or below it. The fact that W tended to match $n+1$ for random phases, as for sine phase, reflects this listener's choice to wait for a phase condition for which a harmonic above the pulsed harmonic was unmasked. Further, listeners reported that for some phase relationships no harmonic at all was unmasked when n was pulsed for n greater than 10.

C. Discussion

The results of Experiment 5 show that the unmasking effect occurs when the phases are random. Therefore, the results support the contention from the previous main section that the unmasking effect is different from the DPE. However, although the unmasking effect can be seen when the phases are randomized, a comparison of Experiments 4 and 5 shows that changing from the sine phase to the random phase caused matches to become less consistent. Some of the random phase conditions preferentially unmasked the harmonic above the pulsed harmonic while other phase conditions unmasked the harmonic below. Perhaps because of this confusion, each match required more complex tone sequences for the random phase than for the sine phase, on average, 6.0 and 3.5, respectively. Figure 6 shows that the matches to $n+1$ and the matches to $n-1$ were often in separate tight clusters, but the division of a small total number of matches (five) into two clusters made the random phase data less

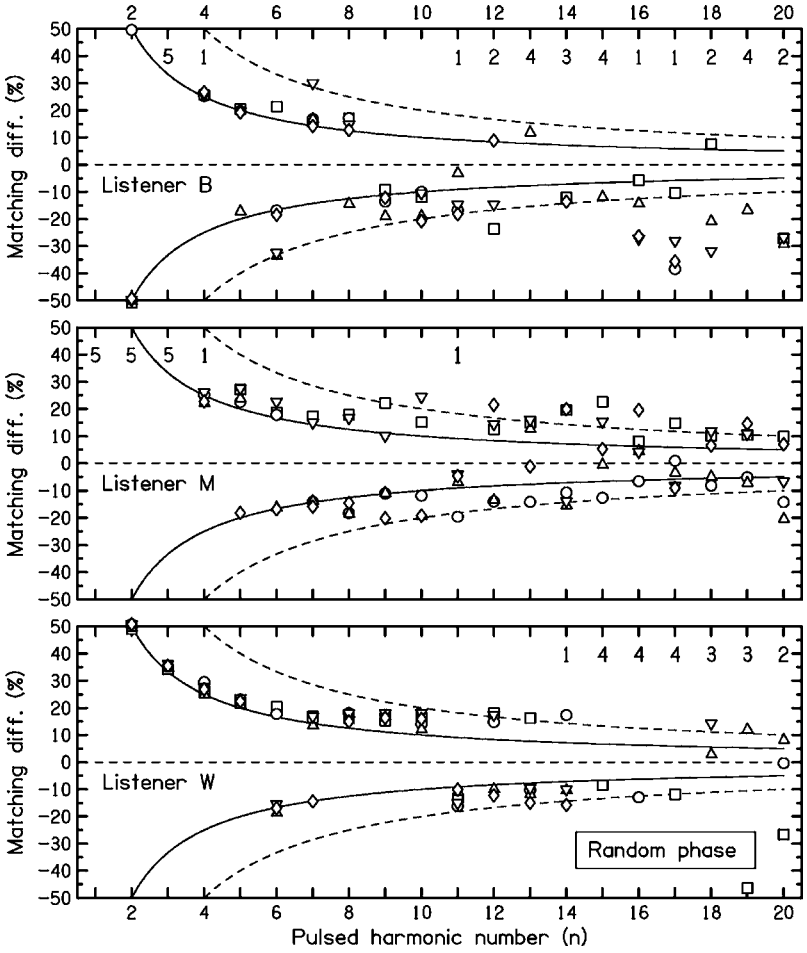


FIG. 6. Experiment 5 results, as in Fig. 6, except that in Experiment 5 listeners could hear as many different random phase relationships as they wanted before making a pitch match.

useful than the sine-phase data.

VII. EXPERIMENT 6—SIX-INTERVAL SEQUENCE—SCHROEDER PHASE

By comparing pitch matches for different phase-versus-frequency curvatures, Experiment 3 arrived at some insight concerning enhancement. Experiment 6 was performed with the hope that employing different curvatures would also lead to some insight concerning unmasking. Especially, it was hoped to discover why some phase relationships in Experiment 5 apparently caused listeners to match harmonic $n+1$ but others caused them to match harmonic $n-1$.

A. Method

Experiment 6 was like Experiment 4 in that the relative phases among harmonics were fixed for each run. Whereas Experiment 4 used only sine phase, Experiment 6 used the sine phase ($m0$) as well as Schroeder phase relationships $m+$ and $m-$, as in Experiment 3. Because phase relationships were expected to be important only for high harmonic numbers, only harmonics 11 through 20 were pulsed in Experiment 6. However, listener W had difficulty hearing unmasking for pulsed harmonics greater than 14 (see Figs. 5 and 6), and, consequently, W was given pulsed harmonics 5 through 15.

B. Results

The pitch matching results of Experiment 6 are plotted in Fig. 7 for the three phase relationships, as in Fig. 4. Because Fig. 7 plots mean and standard deviation, and because listeners matched either up or down (expected to be $n+1$ or $n-1$), this figure led to the unique problem of plotting means in two groups. In practice, it was normally not difficult to decide whether to include any given match in the positive deviation group or the negative group because deviations near zero were rather rare. Consequently, the decisions were made automatically, based only on the sign of the deviation.

For each listener, Fig. 7 shows patterns of matching that resemble those in Figs. 5 and 6, except that listener B had apparently learned to make more consistent matches at high values of n. The patterns were similar for all phase relationships. There was no evidence linking phase curvature with a tendency to match up or to match down. The only consistent trend visible in the figure is that the matching frequencies for listeners B and M for pulsed harmonics 13 through 18 increase monotonically as the curvature decreases from positive to zero to negative. Matches for listener Z did not exhibit that dependence.

A systematic trend did appear in the number of no–matches. As the phase curvature decreased from positive to zero to negative, the number of no-matches summed over

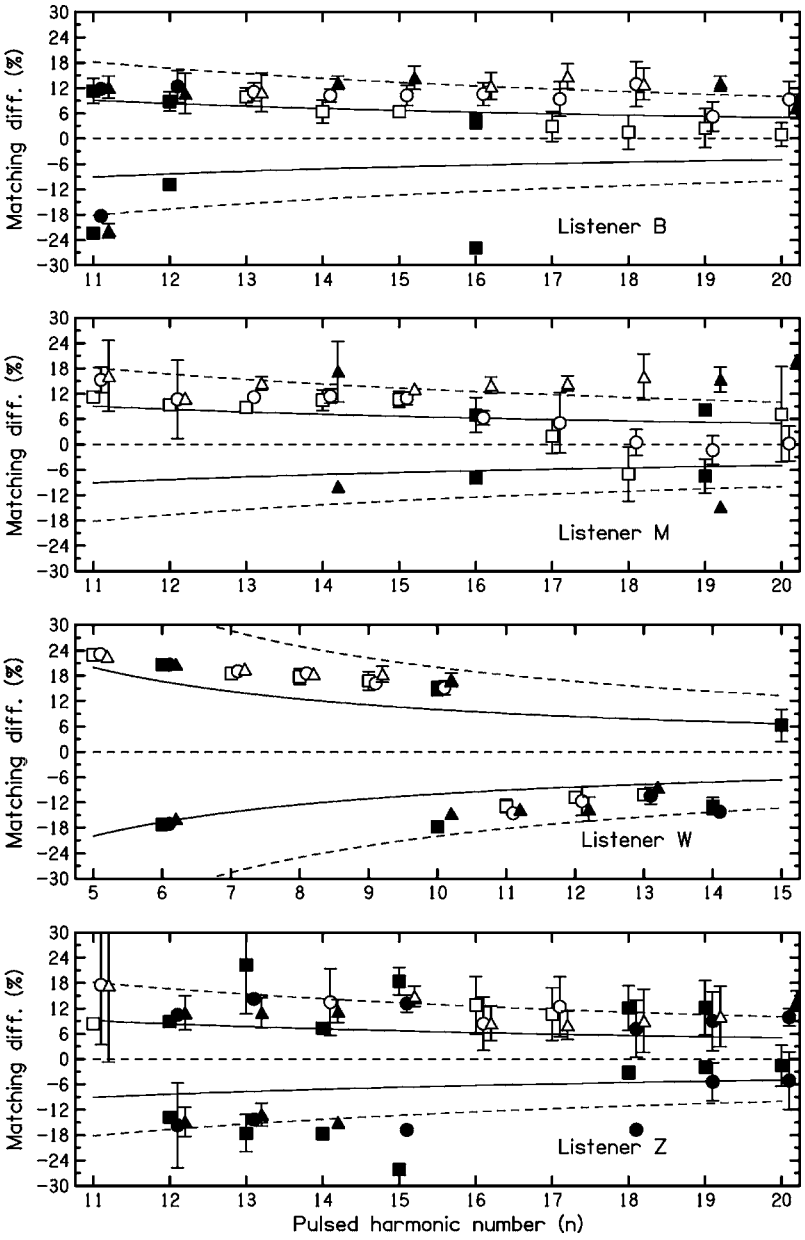


FIG. 7. Experiment 6 results, showing the mean difference between the frequencies of the matching tone and of the pulsed harmonic for the six-interval sequence with three phase relationships: squares for $m+$ (Schroeder plus), circles for $m0$ (sine phase), and triangles for $m-$ (Schroeder minus). Open symbols indicate that all five matches are averaged. Filled symbols indicate 1–4 matches. Error bars are two standard deviations in overall length.

listeners increased from 2 to 15 to 26. For every listener, the number of no-matches for $m+$ was less than the number for $m-$.

C. Discussion

Experiment 6 discovered little, if any, systematic dependence of the pitch of an unmasked harmonic on phase curvature. In particular, it did not make it possible to manipulate a listener's tendency to match $n-1$ instead of $n+1$, or vice versa. Some other form of phase variation, beyond our simple curvature dimension, must be responsible for the effects seen with random phase.

Experiment 6 did discover a suggestive curvature effect in unmasked harmonic detection. The no-match data indicated that detecting an unmasked harmonic is easier for increasing positive curvature. In this sense, the data from the unmasking experiment resemble the data from the detection experiments of Kohlrausch and Sander (1995) and Oxenham and Dau (2001). Apparently, the unmasking effect is aided by a small-amplitude portion during a stimulus cycle, as is the case for the DPE.

VIII. A PLACE MODEL FOR UNMASKING

As noted in the introduction to unmasking, the unmasking effect is suggested by excitation pattern models, including that of Terhardt *et al.* (1982a, b). This model crystallized concepts of excitation patterns, partial masking, and spectral pitch shifts that had been developed in Munich over several decades. The goal of the algorithm is to compute a virtual pitch of a complex tone, based on the strengths and spectral pitches of individual harmonics. It is the strengths and spectral pitch aspects that are applicable in the present article—especially the strengths. To our knowledge, this model is unique in assigning strengths and pitches to individual har-

monics of a tone. The strengths of the harmonics, originally called "SPL excess" by Terhardt *et al.* are here called "spectral strengths." They are measured in dB.

A. Spectral strength

A calculation of the spectral strength begins with the level of the harmonic in dB, compared to a standard absolute threshold for a sine tone of that frequency. Next, the effective level is reduced by accounting for partial masking by neighboring harmonics. The masking pattern caused by a harmonic is assumed to be triangular on a coordinate system of decibels versus the tonotopic scale in bark. Neighboring harmonics *above* the harmonic of interest produce a partial masking that is attenuated at a rate of 27 dB/bark. Neighboring harmonics *below* the harmonic of interest produce additional partial masking with a variable slope—shallower (more effective masking) for lower frequencies and for higher levels of the neighbor.

As noted above, the algorithm applies to the unmasking effect because, as a harmonic is removed from the spectrum, its masking effect on its neighbors is reduced. That reduction in partial masking leads to an effective boost in the strength of a neighbor, especially the neighbor immediately above the removed harmonic. The sudden boost in strength is imagined to cause the neighbor to stand out perceptually so that its pitch can be matched. Because the algorithm ignores phase information, it is evident from the outset that it will not be able to reproduce all the features seen experimentally in the unmasking effect. However, the algorithm applies so naturally that it cannot be ignored, and it does make some interesting predictions.

Figure 8 shows the predictions of the model. The horizontal axis, "Pulsed harmonic number" ranges from $n=1$ to $n=20$, as in our experiments. Above and below that axis are plots that show the effect of pulsing the nth harmonic on the strengths of harmonics $n+1$ and $n-1$, respectively. An open bar shows the strength before harmonic n is removed. The adjacent filled bar shows the strength after n is removed. The difference in height between filled and open bars is a measure of the sudden boost in strength that unmasks a harmonic.

When the spectral strength is negative in Fig. 8, the harmonic is predicted to be inaudible according to the model. This prediction is consistent with the general tendency for matches to fail for pulsed harmonics greater than 15, as observed for listeners B and W for sine phase, but not for M and Z. Thus, the prediction does not hold in general. The spectral strength of high-frequency components is primarily limited by masking from lower harmonics. Increasing the level of the components leads to a shallower masking slope with the result that the spectral strength of high-frequency components becomes more negative as the stimulus level increases.

Because masking from below is more effective than masking from above, removing harmonic n is predicted to have a much greater effect on harmonic $n+1$ than on harmonic $n-1$. This effect can be seen in Fig. 8, where the difference between filled and open bars is greater in the upper graph than in the lower graph. As the pulsed harmonic number increases beyond 10, this disparity is reduced and the algorithm tends to predict that harmonic $n-1$ would be unmasked almost as effectively as harmonic $n+1$. These predictions partly agree with experiment in that there is a clear tendency for listeners to hear out harmonic $n+1$ and not $n-1$. But, as noted in Sec. V C, idiosyncratic and reproducible tendencies for listeners to hear out $n-1$ do occur for n less than 10, and this is not predicted by the algorithm.

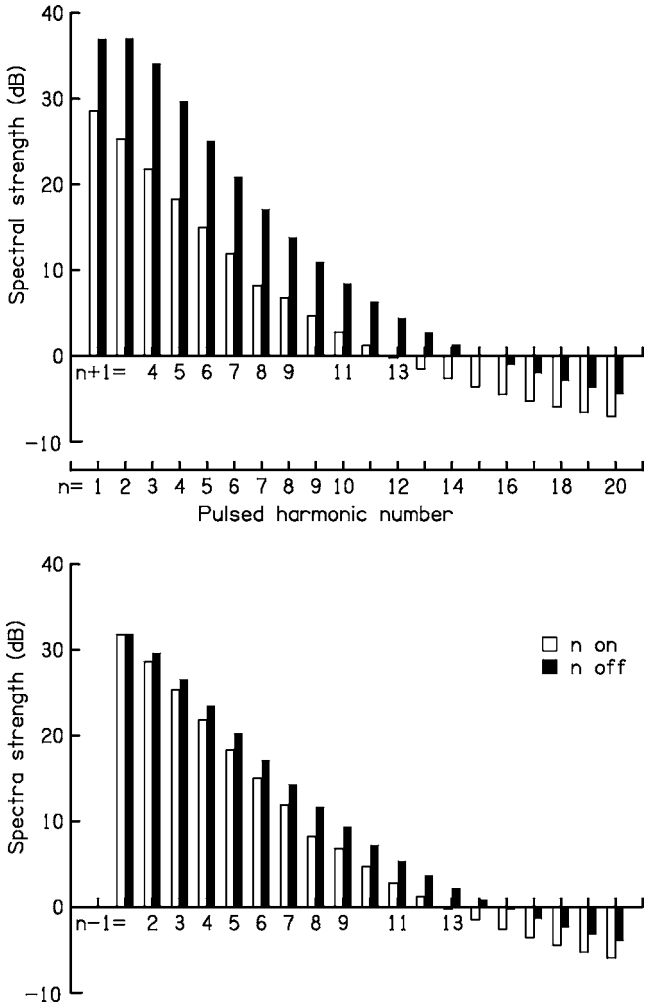


FIG. 8. Spectral strength of unmasked harmonics compared to their normal masked strength as calculated from the algorithm by Terhardt *et al.* The strengths of harmonics $n+1$ and $n-1$ with harmonic n OFF are shown by solid bars. The strengths with harmonic n ON, as normal, are shown by open bars. Negative strengths correspond to harmonics that should not be audible according to the algorithm. A harmonic with a negative strength closer to zero is stronger than one with a negative strength farther from zero.

Figure 8 shows the calculated unmasking only for harmonics $n+1$ and $n-1$. Calculations within the algorithm show that there is some unmasking for harmonics $n+2$ and $n-2$, but the effects are quite small. The predicted unmasking of $n+2$ is always appreciably less than the unmasking of $n-1$. Therefore, the algorithm predicts that the most likely alternative to a match at $n+1$ is not $n+2$ but is $n-1$ instead.

B. Pitch shifts

The algorithm of Terhardt *et al.* computes pitch shifts for each harmonic of a complex tone. The appropriate compari-

son between the algorithm and our experiment is to compute the difference between the pitch of harmonic $n+1$ (or $n-1$), with harmonic n absent, and the pitch of a sine tone having the level and frequency of the experimental matching tone. As used here a "pitch shift" does not refer to a comparison of the pitch of an unmasked harmonic with the pitch of that harmonic in the context of all harmonics. Instead it refers to the pitch of the unmasked harmonic compared to the pitch of a sine tone in quiet. The differences calculated from the algorithm ought to agree with the shifts of the matches from the solid line in Figs. 5–7.

The algorithm includes pitch shifts of the matching tone per Stevens's rule (Fletcher, 1934) such that pitch p is related to frequency f (Hz) by

$$p = f[1 + (f - 2000)(L - 60) \times 2 \times 10^{-7}],$$

where L is the level in dB SPL. In our experiment, the levels of the matching tone, which listeners adjusted to taste, varied from about 40–70 dB SPL. Over this range the Stevens's rule shifts are less than one percent. Compared to the variability in our data, an uncertainty of this magnitude is not important, and pitch shifts are here computed using matching tone frequencies and not intensity-adjusted matching tone pitches.

Pitch shifts were computed for six-interval runs using only data from those pulsed harmonics for which the listener put all five matches in a cluster. Because Experiment 5 with random phases so often led to matches to both $n+1$ and $n-1$ there were few such clusters. Matches were more consistent in Experiment 4 with the sine phase, leading to 41 clusters for n in the range 2–18 inclusive: 9 for B, 13 for M, 8 for W, and 11 for Z. Of these 41 clusters, there were only 5 at $n-1$. These data, together with the predictions of the algorithm appear in Fig. 9.

The solid lines in Fig. 9 show the predictions of the algorithm. The heavy dashed line shows a shift of harmonic $n+1$ that is so large that the frequency equals that of harmonic $n+2$, i.e., $100[(n+2)/(n+1)-1]$, which sets a terminus for the applicability of the algorithm, somewhat below harmonic 17. Data points show pitch shifts from Experiment 4 for the four listeners, as indicated.

Figure 9 shows that the algorithm successfully predicts the order of pitch shifts for unmasked harmonics $n+1$. It underestimates the shifts for listener W and usually overestimates them for the other listeners. By contrast, the algorithm fails to predict the negative pitch shifts that occur for matches to $n-1$.

The pattern of experimental pitch shifts shown in Fig. 9 is a familiar one indicating contrast enhancement in a matching experiment. Given a target that is above a reference, a listener tends to match high. Given a target that is below a reference, a listener tends to match low. Contrast is enhanced in that a listener's response to a displaced stimulus exaggerates the displacement. Similar results are seen in the pitch shifts caused by a leading tone (Hartmann and Blumenstock, 1976; Rakowski and Hirsh, 1980). These shifts were modeled in a contrast enhancement theory by Kanistanaux and Hartmann in 1979. Similar results occur for a mistuned harmonic (Hartmann and Doty, 1996; Lin and Hartmann, 1998),

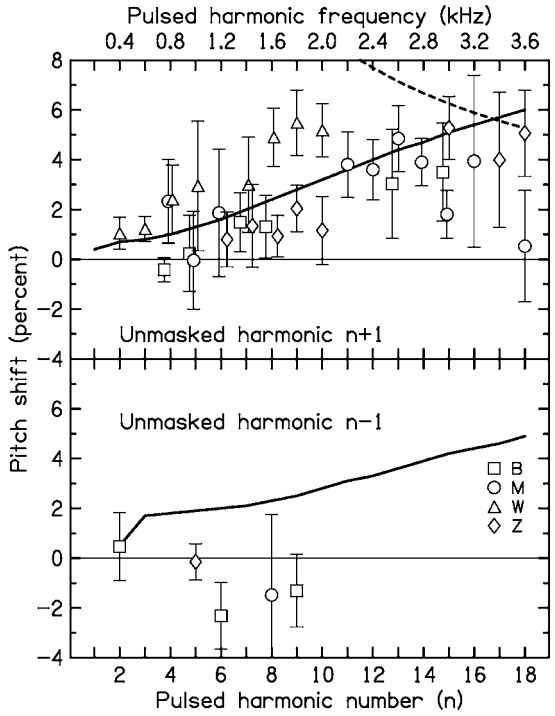


FIG. 9. Pitch shifts calculated from the algorithm of Terhardt *et al.* are shown by solid lines for unmasked harmonics above $(n+1)$ and below $(n-1)$ the pulsed harmonic. The dashed line shows the frequency of harmonic $n+2$. Experimental pitch shifts from Experiment 4 are shown by symbols for four listeners, B, M, W, and Z. Error bars are two standard deviations in overall length.

and a contrast enhancement model was constructed to try to account for these results by de Cheveigné (1997). Roberts and Brunstrom (2003) turned this particular shift into an effective tool for measuring the perceptual fusion of harmonics in a complex tone. As observed in other pitch shift experiments, the unmasking experiment finds that positive pitch shifts in response to a target above a reference tend to be larger and more reliable than negative shifts in response to a target below a reference.

IX. CONCLUSIONS

Alternately omitting and restoring a harmonic in a periodic complex tone creates two salient effects, enhancement and unmasking. In the enhancement effect, the pulsed harmonic itself stands out from the rest of the periodic tone. Enhancement was studied in Experiments 1 through 3 of this article for a 200-Hz complex tone. It was found that the relative phases among harmonics affect the matching data for pulsed harmonics greater than the 14th, though listeners can detect phase effects for pulsed harmonics as low as the 8th. The onset of sensitivity to relative phase indicates a failure of spectral resolution in the auditory periphery. But although harmonics above the 10th (Bernstein and Oxenham, 2003), or 8th–11th (Moore *et al.*, 2006), or 8th–14th (this work) are not well resolved, successful matches can be made to enhanced harmonics up to the 20th for relative phases that are not unfavorable. As shown in Experiment 3, unfavorable phases are those that lead to temporally compact (peaked) excitation. Favorable phases are those that distribute the ex-

citation more uniformly throughout the period of the complex tone. The conclusion of these experiments is that the audibility of an enhanced harmonic is not limited by spectral resolution. Therefore, theories of the enhancement effect that depend on the adaptation or inhibition of tonotopic regions are likely to have difficulty explaining the enhancement of harmonics as high as the 20th. Experiments 1 through 3 also found that the pitches of enhanced harmonics were not shifted significantly.

In the unmasking effect, the pulsed harmonic causes a neighboring harmonic to be audible. Specifically, when harmonics are presented in sine phase and when harmonic n is turned off, harmonic $n+1$ tends to pop out from the complex tone background. However, about 10%–20% of the time it is harmonic $n-1$ that pops out. When harmonics are presented in a random phase, harmonic $n-1$ pops out almost as frequently as harmonic $n+1$, depending on the phase relationship that actually occurs. Phase relationships that produced temporally compact cochlear excitation were most effective in unmasking harmonics.

For favorable phases, unmasking can be reliably heard by some listeners for pulsed harmonics as high as the 18th. Unlike enhanced harmonics, unmasked harmonics are subject to clear pitch shifts. The pitch shifts tend to exaggerate the spacing between the pulsed harmonic and the unmasked harmonic. In other words, matches to $n-1$ tend to be flat and matches to $n+1$ tend to be sharp compared to harmonic frequencies, indicating contrast enhancement.

The preference for unmasking harmonic $n+1$ over harmonic $n-1$ finds a natural explanation in terms of partial masking, as exemplified by the algorithm of Terhardt *et al.* (1982a, b). Because of the upward spread of masking, the removal of harmonic n has the greatest effect on the excitation pattern for a harmonic higher than n. However, the algorithm admits no role for harmonic phases, and so it fails to predict the phase effects observed in Experiment 5, where phases were randomized. Further, the algorithm predicts no unmasking for pulsed harmonics above the 14th, contrary to experiment.

The algorithm also predicts pitch shifts. It is successful for the positive pitch shifts observed for matches to harmonics $n+1$, but it fails for the negative pitch shifts observed for matches to harmonics $n-1$. The existence of negative shifts and the alternative explanation in terms of contrast enhancement raise questions about the role of partial masking in spectral pitch as implemented in the algorithm. However, invoking both partial masking and contrast enhancement allows one to have one's cake and eat it too. The shifts caused by partial masking are always positive; the shifts from contrast enhancement are exaggerations of the difference between target and reference. If the contrast enhancement shifts are the larger, then positive pitch shifts in response to a high target and negative shifts in response to a low target are both predicted. Further, the positive shifts are predicted to be larger and more reliable than the negative shifts, as observed experimentally.

Both the enhancement effect and the unmasking effect offer promising opportunities to study tonotopic and temporal aspects of neural excitation patterns, admittedly at a yet-to-be-determined level of the system. Both effects involve selective attention, though the relative importance of attention *per se* is not clear. In any event, the *unmasking* effect, where the perceived harmonic is different from the pulsed harmonic, reveals a mechanism whereby a harmonic is exposed as the excitation pattern changes. Thus the effect is aptly named "unmasking," though specific roles of pattern masking and suppression are unknown. The unmasking effect provides a way to study excitation patterns experimentally without using powerful masking noises that may change the operating point of the system under study. Unmasking reveals both spectral resolution and phase effects. By studying unmasking with a pitch matching task in the present article, we were able to make use of the extraordinary precision of human pitch perception to reveal details of the effect.

ACKNOWLEDGMENTS

We are grateful to Dr. N. F. Viemeister, Dr. B. A. Wright, Dr. A. J. Oxenham, and J. G. Bernstein, for discussions concerning the enhancement effect. Dr. B. C. J. Moore and several anonymous reviewers made useful suggestions on an earlier version of the manuscript. This work was supported by the NIDCD of the NIH under Grant No. DC 00181.

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.

Cardozo, B. L. (**1967**). "Ohm's law and masking," J. Acoust. Soc. Am. **38**, 1193 (abstract).

de Cheveigné, A. (**1997**). "Harmonic fusion and pitch shifts of mistuned partials," J. Acoust. Soc. Am. **102**, 1083–1087.

Duifhuis, H. (**1970**). "Audibility of high harmonics in a periodic pulse," J. Acoust. Soc. Am. **48**, 888–893.

Duifhuis, H. (**1971**). "Audibility of high harmonics in a periodic pulse II, Time effect," J. Acoust. Soc. Am. **49**, 1155–1162.

Duifhuis, H. (**1972**). "Perceptual analysis of sound," Thesis, University of Eindhoven, The Netherlands (unpublished).

Egan, J. P., and Hake, H. W. (**1950**). "On the masking pattern of a simple auditory stimulus," J. Acoust. Soc. Am. **22**, 622–630.

Fletcher, H. (**1934**). "Loudness, pitch and the timbre of musical tones and their relation to the intensity, frequency and overtone structure," J. Acoust. Soc. Am. **6**, 59–69.

Gibson, H. L. (**1971**). "The ear as an analyzer of musical tones," J. Acoust. Soc. Am. **49**, 127 (abstract).

Goupell, M. J., and Hartmann, W. M. (**2004**). "Spectral analysis for the ear," *Proceedings of the 18th International Congress on Acoustics*, We2E2.

Goupell, M. J., Zhang, P. X., and Hartmann, W. M. (**2003**). "Cancelled harmonics—How high does the effect go?," J. Acoust. Soc. Am. **113**, 2292 (abstract).

Hartmann, W. M. (**1997**). *Signals, Sound, and Sensation* (Springer-Verlag, New York), p. 128.

Hartmann, W. M., and Blumenstock, B. J. (**1976**). "Time dependence of pitch perception: Pitch step experiment," J. Acoust. Soc. Am. **60**, S40 (abstract).

Hartmann, W. M., and Doty, S. L. (**1996**). "On the pitches of the components of a complex tone," J. Acoust. Soc. Am. **99**, 567–578.

Helmholtz, H. L. F. (**1877**). *On the Sensations of Tone*, 4th ed. (Dover, New York), p. 61 (translated by A. J. Ellis).

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.

Houtsma, A. J. M., Rossing, T. D., and Wagenaars, W. M. (**1987**). *Auditory Demonstrations*, Acoustical Society of America (Eindhoven, The Netherlands).

Kanistanaux, D. C., and Hartmann, W. M. (**1979**). "On the shift in pitch of a sine tone caused by a preceding tone," J. Acoust. Soc. Am. **65**, S37 (abstract). See also "The effect of a prior tone on the pitch of a short tone," *Proc. Res. Symposium on the Psychology and Acoustics of Music*, edited

by W. V. May, University of Kansas. Department of Music Report, **1979**.
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.
Lin, J.-Y., and Hartmann, W. M. (**1997**). "On the Duifhuis pitch effect," J. Acoust. Soc. Am. **101**, 1034–1043.
Lin, J.-Y., and Hartmann, W. M. (**1998**). "The pitch of a mistuned harmonics: Evidence for a template model," J. Acoust. Soc. Am. **103**, 2608–2617.
Moore, B. C. J., and Ohgushi, K. (**1993**). "Audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **93**, 452–461.
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.
Oxenham, A. J., and Dau, T., (**2001**). "Reconciling frequency selectivity and phase effects in masking," J. Acoust. Soc. Am. **110**, 1525–1538.
Palmer, A. R., Summerfield, Q., and Fantini, D. A. (**1995**). "Responses of auditory-nerve fibers to stimuli producing psychophysical enhancement," J. Acoust. Soc. Am. **97**, 1786–1799.
Peters, R. W., Moore, B. C. J., and Glasberg, B. R. (**1983**). "Pitch of components of a complex tone," J. Acoust. Soc. Am. **73**, 924–929.
Plomp, R., and Mimpen, A. M. (**1968**). "The ear as a frequency analyzer II," J. Acoust. Soc. Am. **43**, 764–767.
Rakowski, A., and Hirsh, I. J. (**1980**). "Post-stimulatory pitch shifts for pure tones," J. Acoust. Soc. Am. **68**, 467–474.
Roberts, B., and Brunstrom, J. M. (**2003**). "Spectral pattern, harmonic relations, and the perceptual grouping of low-numbered harmonics," J. Acoust. Soc. Am. **114**, 2118–2134.
Schroeder, M. R. (**1970**). "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation," IEEE Trans. Inf. Theory **IT-16**, pp. 85–89.
Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.
Smith, B. K., Sieben, U. K., Kohlrausch, A., and Schroeder, M. R. (**1986**). "Phase effects in masking related to dispersion in the inner ear," J. Acoust. Soc. Am. **80**, 1631–1637.
Summerfield, Q., Sidwell, A., and Nelson, T. (**1987**). "Auditory enhancement of changes in spectral amplitude," J. Acoust. Soc. Am. **81**, 700–708.
Terhardt, E. (**1971**). "Pitch shifts of harmonics, an explanation of the octave enlargement phenomenon," *Proc. 7th ICA*, Budapest, Vol. **3**, pp. 621–624.
Terhardt, E., Stoll, G., and Seewann, M. (**1982a**). "Pitch of complex signals according to virtual-pitch theory: Test, examples, and predictions," J. Acoust. Soc. Am. **71**, 671–678.
Terhardt, E., Stoll, G., and Seewann, M. (**1982b**). "Algorithm for extraction of pitch and pitch salience from complex tonal signals," J. Acoust. Soc. Am. **71**, 679–688.
Verschuure, J., and van Meeteren, A. A. (**1975**). "The effect of intensity on pitch," Acustica **32**, 33–44.
Viemeister, N. F. (**1980**). "Adaptation of masking," in *Psychological, Physiological, and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft Univ. Press, Delft), pp. 190–198.
Viemeister, N. F., and Bacon, S. P. (**1982**). "Forward masking by enhanced components in harmonic complexes," J. Acoust. Soc. Am. **71**, 1502–1507.

Perception of acoustic scale and size in musical instrument sounds

Ralph van Dinther[a] and Roy D. Patterson
Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3EG UK

(Received 29 March 2005; revised 28 July 2006; accepted 28 July 2006)

There is size information in natural sounds. For example, as humans grow in height, their vocal tracts increase in length, producing a predictable decrease in the formant frequencies of speech sounds. Recent studies have shown that listeners can make fine discriminations about which of two speakers has the longer vocal tract, supporting the view that the auditory system discriminates changes on the acoustic-scale dimension. Listeners can also recognize vowels scaled well beyond the range of vocal tracts normally experienced, indicating that perception is robust to changes in acoustic scale. This paper reports two perceptual experiments designed to extend research on acoustic scale and size perception to the domain of musical sounds: The first study shows that listeners can discriminate the scale of musical instrument sounds reliably, although not quite as well as for voices. The second experiment shows that listeners can recognize the *family* of an instrument sound which has been modified in pitch and scale beyond the range of normal experience. We conclude that processing of acoustic scale in music perception is very similar to processing of acoustic scale in speech perception. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2338295]

PACS number(s): 43.66.Lj, 43.66.Ba [AK] Pages: 2158–2176

I. INTRODUCTION

When a child and an adult say the same word, it is only the message that is the same. The child has a shorter vocal tract and lighter vocal cords, and as a result, the wave form carrying the message is quite different for the child and the adult. The form of the size information is illustrated in Fig. 1, which shows four versions of the vowel /a/ as in "hall." From the auditory perspective, a vowel is a "pulse-resonance" sound, that is, a stream of pulses, each with a resonance showing how the vocal tract responded to that pulse.[1] The "message" of the vowel is contained in the shape of the resonance (i.e., the relative height and spacing of the ripples following each pulse) which is the same in every cycle of all four waves of Fig. 1. The left column shows two versions of /a/ spoken by one adult using a high pitch (a) and a low pitch (b); the glottal pulse rate determines the pitch of the voice. The resonances are the same since it is the same person speaking the same vowel. The right column shows a child (c) and the same adult (d) speaking the /a/ sound on the same pitch. The pulse rate and the shape of the resonance are the same, but the *scale* of the resonance within the glottal cycle is dilated in the lower panel. The adult has a longer vocal tract which reduces the formant frequencies and the formant bandwidths. The reduction in formant bandwidth means that the resonances ring longer. Thus, vowel sounds contain two forms of information about the size of the source—the pulse rate (PR) and the resonance scale (RS); the shape of the resonance carries the message.

A high-quality vocoder has been developed that can manipulate the PR and RS of natural speech (Kawahara *et al.*, 1999; Kawahara and Irino, 2004); it is referred to as STRAIGHT and it has been used to manipulate the PR and RS of speech sounds in a number of perceptual experiments. For example, Smith *et al.* (2005) and Ives *et al.* (2005) used STRAIGHT to scale the resonances in vowels and syllables, respectively, and they showed that listeners can reliably discriminate the changes in RS that are associated with changes in vocal-tract length (VTL). Listeners heard the difference between scaled vowels as a difference in speaker size, and the just-noticeable difference (JND) in perceived size was less than 10% for a wide range of combinations of PR and RS. Smith and Patterson (2005) also used STRAIGHT to scale the resonances in vowels and showed that listeners' estimates of speaker size are highly correlated with RS. They manipulated the PR as well as the RS and found that PR and RS interact in speaker size estimates.

Assmann *et al.* (2002) and Smith *et al.* (2005) showed that vowels manipulated by STRAIGHT are readily recognized by listeners. Assmann *et al.* showed that a neural net model could be used to explain the recognition performance provided the vowels were not scaled too far beyond the normal speech range. Smith *et al.* showed that good recognition performance is achieved even for vowels scaled well beyond the normal range. They argued that it was more likely that the auditory system had general mechanisms for normalizing both PR and RS, as suggested by Irino and Patterson (2002), rather than a neural net mechanism for learning all of the different acoustic forms of each vowel type, as suggested by Assmann *et al.* (2002). We will return to the issue of mechanisms in the discussion (Sec. V).

[a]Electronic mail: rv230@cam.ac.uk. Portions of this work were reported at meetings of the British Society of Audiology, London, 2004 and the 149th meeting: Acoustical Society of America, Vancouver, 2005.

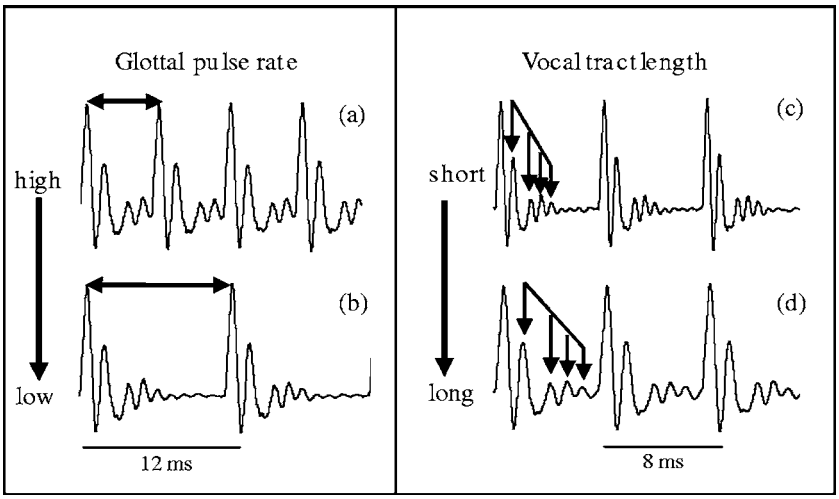


FIG. 1. The internal structure of pulse-resonance sounds illustrating the pulse rate and the resonance scale of vowel sounds.

The purpose of the current study was to extend the research on the perception of scaled sounds to another class of everyday sounds where it is clear that source size plays a role in what we perceive. The sounds are the musical notes produced by the sustained-tone instruments of the orchestra. They come in families (e.g., brass instruments) which have similar shape and construction, and which differ mainly in size (e.g., trumpet, trombone, euphonium and tuba). We hear the members of a family as sounding the same in the sense of having the same timbre (the message), but at the same time we are able to tell whether the sound we are listening to is from a larger or smaller member of the family.

Section II of this paper briefly reviews the form of size information in the notes of sustained-tone instruments, and the role of RS in the perception of musical sounds. The section shows that, although the mechanisms whereby musical instruments produce their notes are very different from the way humans produce vowels, the notes are, nevertheless, pulse-resonance sounds, and the information about instrument size is largely summarized in the PR and RS values. Section III describes an experiment on the discrimination of RS in sounds produced by four musical instruments taken from four families; strings, woodwinds, brass and voice. The just-noticeable difference (JND) for the RS dimension is found to be a little larger for brass, string and woodwind instruments than it is for the voice (as an instrument). Section IV describes a recognition experiment for 16 musical sounds (four instruments in each of four families). The experiment shows how much the RS of an instrument can be modified without rendering the individual instruments unrecognizable. In both experiments, the scaling is implemented with STRAIGHT. Then in Sec. V, we return to the question of how the auditory system might process RS information in music and speech sounds.

II. RESONANCE SCALE IN MUSICAL INSTRUMENTS

The wave forms of a trumpet and a trombone are shown in Fig. 2, which shows that they both have a pulse-resonance form.[1] They also have the same PR, and so they are playing the same note ($B_3^\flat$, 233 Hz). The RS of the trombone is more dilated than that of the trumpet, and it is this that makes the trombone sound larger than the trumpet when they play the same note. Dilation in the time domain corresponds to contraction of the envelope in the frequency domain. In this section we briefly describe the mechanisms that control the PR and RS in instruments capable of producing sustained sounds. (Note that the human voice is a sustained-tone instrument, in this sense.) The purpose of the analysis is to illustrate the ubiquitous nature of size information in everyday sounds. First, we review the "source-filter" model which is traditionally used to explain the sounds produced by sustained-tone instruments. Then, we show that scaling the spatial dimensions of an instrument proportionately produces wave forms whose structure is invariant, and which differ only in terms of the value of a scaling constant. This illustrates that scale is a property of sound, just like time and frequency (Cohen, 1993).[2]

A. Sustained-tone instruments

Musical instruments that produce sustained tones can be modeled as *linear* resonant systems (e.g., air columns, cavi-

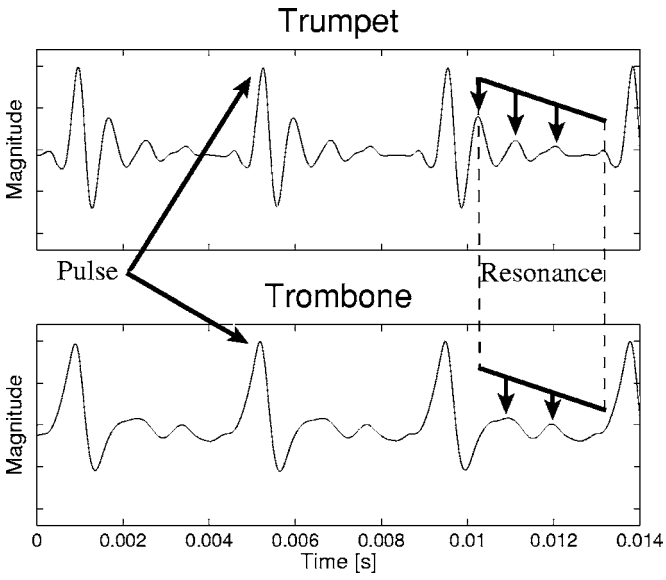


FIG. 2. Wave forms of notes produced by a trumpet (top panel) and a trombone (bottom panel). The pulses and resonances are indicated by arrows.

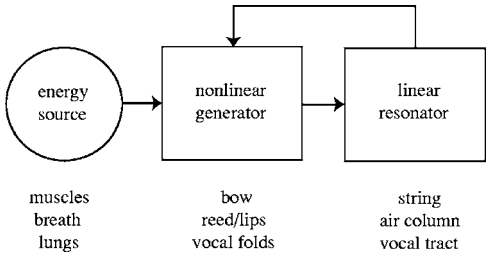


FIG. 3. Block diagram of a sustained-tone instrument.

ties, strings) excited by a *nonlinear* generator (e.g., vocal folds, lips, reeds, bows). The nonlinear generator produces acoustic pulses, and when the generator is coupled to the resonator, as indicated by the feedback loop in Fig. 3, the system produces a temporally regular stream of acoustic pulses, similar to a click train. Thus, the nonlinearity of the generator virtually ensures that the waves of sustained-tone instruments are pulse-resonance sounds (Benade, 1976; Fletcher, 1978; Mclntyre *et al.*, 1983). A similar analysis applies to the voiced sounds of speech (Chiba and Kajiyama, 1941; Fant, 1960). The Fourier spectrum of a click train is a set of phase-locked harmonics of the click rate, and Fletcher (1978) has confirmed that the notes of sustained-tone instruments have overtones that are strictly harmonic up to fairly high harmonic numbers—locked to the fundamental both in frequency *and phase*.

1. The "source" in sustained-tone instruments

The excitation source in stringed instruments is a combination of bow and string. Over the first 50 ms or so, the string forces the vibration produced by bowing into a standing wave with a quasi-sawtooth shape. Fourier analysis of this wave form shows a spectrum containing all harmonics of the fundamental, with amplitudes decreasing at approximately 6 dB/octave (Fletcher, 1999). All of the harmonics are in phase, as indicated by the sharp rise at the start of each cycle of the sawtooth wave.

The excitation of woodwind, brass and vocal instruments can be modeled by standard fluid mechanics, in terms of "valves" that control the momentary closing of a stream of air. For woodwind instruments, the valve is the reed, for brass instruments, it is the lips, and for the voice, it is the vocal folds. The reeds of the saxophone and clarinet are designated "inward-striking" valves (Helmholtz, 1877). The lips exciting brass instruments and the vocal folds exciting the vocal tract are designated "outward striking" or "sideways-striking" valves (Fletcher and Rossing, 1998). The pulsive nature of the excitation generated by reed, lip and vocal-fold vibrations, and the temporal regularity of the pulse stream, mean that the dominant components of the spectrum are strictly harmonic *and* they are phase locked (Fletcher and Rossing, 1998). Fletcher (1978) provides a mathematical basis for understanding mode locking in musical instruments.

2. Spectral filtering in sustained-tone instruments

The spectral envelope of the source wave of a sustained-tone instrument is modified by the resonant properties of the instrument's components. For stringed instruments, the prominent resonances are associated with the plates of the body (wood resonances), the body cavities (air resonances), and the bridge (Benade, 1976). For brass and woodwind instruments, the prominent resonances are associated with the shape of the mouthpiece, which acts like a Helmholtz resonator, and the shape of the bell which determines the efficiency with which the harmonics radiate into the air (Benade, 1976; Benade and Lutgen, 1988). Woodwind instruments have a tube resonance like brass instruments, however, the spectrum is complicated due to the "open-hole cutoff frequency." The dominant resonances of speech sounds are determined by the shape of the vocal tract (Chiba and Kajiyama, 1941; Fant, 1960). The important point in this brief review, however, is that these body resonances do not affect the basic pulsive nature of the sounds produced by sustained-tone instruments.

In summary, the harmonic structure of the notes produced by sustained-tone instruments, and the fact that the components are phase locked, indicates that a simple model with a nonlinear pulse generator and a coupled linear resonator works quite well for these instruments, including the voice. From the point of view of the instrument maker and the physicist, the review emphasizes that there are many ways to produce a regular stream of pulses and many ways to filter the excitation when producing music and speech. From the auditory perspective, these are pulse-resonance sounds which are characterized by a pulse rate, a resonance scale and a message which is the shape of the resonance.

B. Relation between resonance scale and the size of an instrument

If all three spatial dimensions of an instrument are increased by a factor, λ, keeping all materials of the instrument the same, the natural resonances *decrease* in frequency by a factor of $1/\lambda$. The shape of the spectral envelope is preserved under this translation; the envelope simply expands or contracts by $1/\lambda$ (on a log-frequency scale, the spectrum shifts as a unit without expansion or contraction). This scaling relationship is called "the general law of similarity of acoustic systems" (Fletcher and Rossing, 1998). It is easy to confirm the law for simple vibrators such as the Helmholtz resonator or a flat plate. The natural frequency of a Helmholtz resonator is

$$f = \frac{c}{2\pi}\sqrt{\frac{A}{VL}}. \tag{1}$$

If the spatial dimensions are scaled up by a factor λ, then

$$f' = \frac{c}{2\pi}\sqrt{\frac{\lambda^2 A}{(\lambda^3 V)(\lambda L)}} = \frac{c}{2\pi\lambda}\sqrt{\frac{A}{VL}} = \frac{1}{\lambda}f. \tag{2}$$

The resonance frequencies of a plate with dimensions L_x and L_y and thickness h are

$$f_{nm} = kh\left[\left(\frac{m}{L_x}\right)^2 + \left(\frac{n}{L_y}\right)^2\right], \tag{3}$$

where k is a constant and (n,m) are numbers of nodal lines in the y and x direction of the plate. Scaling of the spatial dimensions by a factor λ results in

$$f'_{nm} = k\lambda h\left[\left(\frac{m}{\lambda L_x}\right)^2 + \left(\frac{n}{\lambda L_y}\right)^2\right] = \frac{1}{\lambda} f_{nm}. \quad (4)$$

Scaling the spatial dimensions of an instrument to produce another member of the family in a different register can result in an instrument which is too large, or too small, to play. To solve this problem, instrument makers often adjust the scale of the instrument by (a) changing the spatial dimensions less than would be required to achieve the register change, and at the same time, (b) changing some other property of the instrument that affects scale (such as the thickness, mass or stiffness of one or more of the components) to achieve the desired RS. In this way, they preserve the formant relationships without disproportionate scaling of the spatial dimensions. For example, Hutchins (1967, 1980) constructed a family of eight instruments covering the entire range of orchestral registers, based on the properties of the violin. If the dimensions of the contra bass were six times greater then those of the violin, the formant ratios of the contra bass body would be the same as those of the violin. However, a contra bass six times as large as a violin would be 3.6 m tall, which would be completely impractical. So, in the construction of the new family of violins, the body size and string lengths were scaled to fit human proportions, and the RSs and PRs required for the lower registers were obtained by adjusting the thickness of the body plates, the mass of the strings and the tension of the strings (Benade, 1976). Similarly, the dimensions of the *f* holes were adjusted to attain the required air resonance frequencies. So for the string family, the law of similarity is actually a law of similarity of shape; the spatial scale factors are smaller than would be required by a strict law of similarity; they have to be augmented by mass and thickness scaling to produce the formant ratios characteristic of the string family in an instrument with a large RS.

The law of similarity also applies to brass instruments, and with similar constraints. Luce and Clark (1967) analyzed 900 acoustic spectra from a variety of brass instruments and showed that the spectral envelopes of the trumpet, trombone, open French horn and tuba were essentially scaled versions of one another, and Fletcher and Rossing (1998) report that the size of the cup scales roughly with the size of the instrument. However, the instrument makers adjust the shape of the bell beyond what would be indicated by strict spatial scaling to produce a series of harmonic resonances and to improve tone quality. So the notes of brass instruments would be expected to differ mainly in PR and RS as dictated by the law of similarity, with differences in bell shape having a smaller effect.

In summary, scaling the spatial dimensions of an instrument would shift the frequencies of the resonances in a way that would preserve formant frequency ratios and produce a family of instruments with the same timbre in a range of registers. Thus, when we change the RS of an instrument sound with STRAIGHT, the listener is very likely to perceive the sound as a larger or smaller instrument of the same family. For practical reasons, instrument makers achieve the desired RS for the extreme members of a family with a combination of spatial dimension scaling and scaling of other properties like mass and thickness. Thus, if listeners were asked to estimate the spatial size of instruments from sounds scaled by STRAIGHT, we might expect, given their experience with natural instruments, that they would produce estimates that are less extreme than the resonance scaling would produce if it were entirely achieved by increasing the spatial dimensions of the instrument. This means that the experiments in this paper are strictly speaking about the perception of acoustic scale in musical instruments. However, listeners do not have a distinct concept of scale separate from size, and they associate changes in acoustic scale with changes in spatial size, and so the experiments are about source size in the sense that people experience it. We will draw attention to the distinction between acoustic scale and size at points where it is important.

III. RESONANCE SCALE DISCRIMINATION IN MUSICAL INSTRUMENTS

The purpose of this experiment was to determine the just-noticeable difference (JND) for a change in the resonance scale of an instrument over a large range of PR and RS. The experiment is limited to relative judgments about RS, and so the distinction between acoustic scale and source size does not arise; there is a one-to-one mapping between acoustic scale and source size in this experiment.

A. Method

1. Stimuli and experimental design

The musical notes for the experiments were taken from an extensive, high-fidelity database of musical sounds from 50 instruments recorded by Real World Computing (RWC) (Goto *et al.*, 2003). This database provided individual sustained notes for four families of instruments (strings, woodwind, brass and voice) and for several members within each family. We chose these specific instrument families for two reasons: (1) They produce sustained notes, and so there is little to distinguish the instruments in their temporal envelopes. (2) The sounds have a pulse-resonance structure and there is a high-quality vocoder that can manipulate the PR and RS in such sounds. The vocoder is referred to as STRAIGHT (Kawahara *et al.*, 1999; Kawahara and Irino, 2004) and its operation is described below. In the database, individual notes were played at semitone intervals over the entire range of the instrument. For the stringed instruments, the total range of notes was recorded for each string. The notes were also recorded at three sound levels (forte, mezzo, piano); the current experiments used the mezzo level. The recordings were digitized into "wav" files with a sampling rate of 44 100 Hz and 16 bit amplitude resolution.

The first experiment focused on the baritone member of each instrument family: for the string family, it is the cello; for the woodwind family, the tenor saxophone; for the brass family, the French horn, and for the human voice, the baritone. Each note was extracted with its initial onset and a total duration of 350 ms. The onset of the recorded instrument was included to preserve the dynamic timbre cues of the

TABLE I. The five conditions from which the JNDs were measured.

	Pulse rate		1/Resonance scale
Condition	F_0 [Hz]	Keys	Factor
1	98	G_2	1
2	49	G_1	$2^{-2/3}$
3	196	G_3	$2^{2/3}$
4	49	G_1	$2^{2/3}$
5	196	G_3	$2^{-2/3}$

instrument. A cosine-squared amplitude function was applied at the end of the wave form (50 ms offset) to avoid offset clicks.

The notes were scaled using the vocoder, STRAIGHT, described by Kawahara *et al.* (1999); Kawahara and Irino (2004). It is actually a sophisticated speech processing package designed to dissect and analyze an utterance at the level of individual glottal cycles. It segregates the glottal-pulse rate and spectral envelope information (vocal-tract shape information *and* vocal-tract length information), and stores them separately, so that the utterance can be resynthesized later with arbitrary shifts in glottal-pulse rate and vocal-tract length. Utterances recorded from a man can be transformed to sound like a woman or a child. The advantage of STRAIGHT is that the spectral envelope of the speech that carries the vocal-tract information is smoothed as it is extracted, to remove the harmonic structure associated with the original glottal-pulse rate, *and* the harmonic structure associated with the frame rate of the Fourier analysis window. For speech, the resynthesized utterances are of extremely high quality, even when the speech is resynthesized with PRs and vocal-tract lengths beyond the normal range of human speech. Assmann and Katz (2005) compared the recognition performance for vowels vocoded by STRAIGHT with performance for natural vowels and vowels from a cascade formant synthesizer. They found that performance with the vowels vocoded with STRAIGHT was just as good with natural vowels, whereas performance was 9%–12%, lower with the vowels from the cascade formant synthesizer. Liu and Kewley-Port (2004) have also reviewed the vocoding provided by STRAIGHT and commented very favorably on the quality of its resynthesized speech.

STRAIGHT also appears to be a good "mucoder" (i.e., a device for encoding, manipulating and resynthesizing musical sounds) for the notes of sustained-tone instruments where the excitation is pulsive. There are audio file examples available on our website to demonstrate the naturalness of the mucoded notes.[3] STRAIGHT was used to modify the PR and RS of the notes required for the discrimination experiment, in which the JND was measured for five combinations of PR and RS as indicated in Table I. The experiment was performed with short melodies instead of single notes to preclude listeners performing the task on the basis of a shift in a single spectral peak. The notes shown in this table indicate the octave and key of the tonal melodies presented to the listeners. Figure 4 shows the PR-RS plane and the points where the JND was measured; the arrows show that the JND was measured in the RS dimension. The stimuli were presented over headphones at a level of approximately 60 dB SPL to listeners seated in a sound attenuated booth.

2. Auditory images of the stimuli

The effects of scaling the stimuli with STRAIGHT are illustrated in Figs. 5–7, using "auditory images" of the stimuli (Patterson *et al.*, 1992, 1995) that illustrate the form of the PR and RS information in the sound. The spectral and temporal profiles of the images provide summaries of the PR information from the RS information. Figure 5(a) shows the auditory image produced by the baritone voice with a PR of 98 Hz (G_2) and the original VTL, that is, a RS value of 1. Figure 5(b) shows the auditory image for the corresponding French horn note. The auditory image is constructed from the sound in four stages: First, a loudness contour is applied to the input signal to simulate the transfer function from the sound field to the oval window of the cochlea (Glasberg and Moore, 2002). Then a spectral analysis is performed with a dynamic, compressive, gammachirp auditory filterbank (Irino and Patterson, 2006) to simulate the filtering properties of the basilar partition. Then each of the filtered waves is converted into a neural activity pattern (NAP) that simulates the aggregate firing of all of the primary auditory nerve fibres associated with that region of the basilar membrane (Patterson, 1994a). Finally, a form of "strobed temporal in-

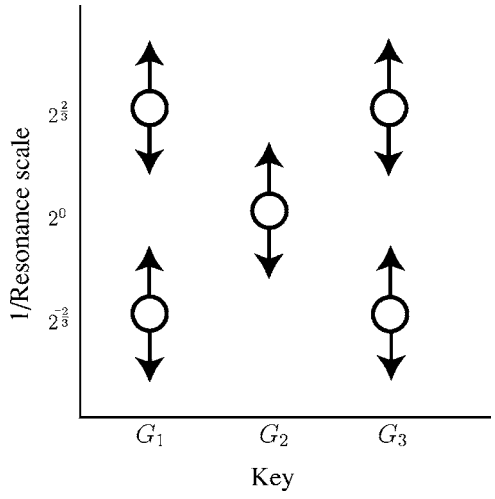


FIG. 4. The "standard" conditions for the psychometric functions on the PR-RS plane. The abscissa is pulse rate in musical notation; the ordinate is the factor by which the resonance scale was modified. The arrows show the direction in which the JNDs were measured. Demonstrations of the five standard sounds are presented on our website[3] for the four instruments in the experiment (cello, sax, French horn and baritone voice).

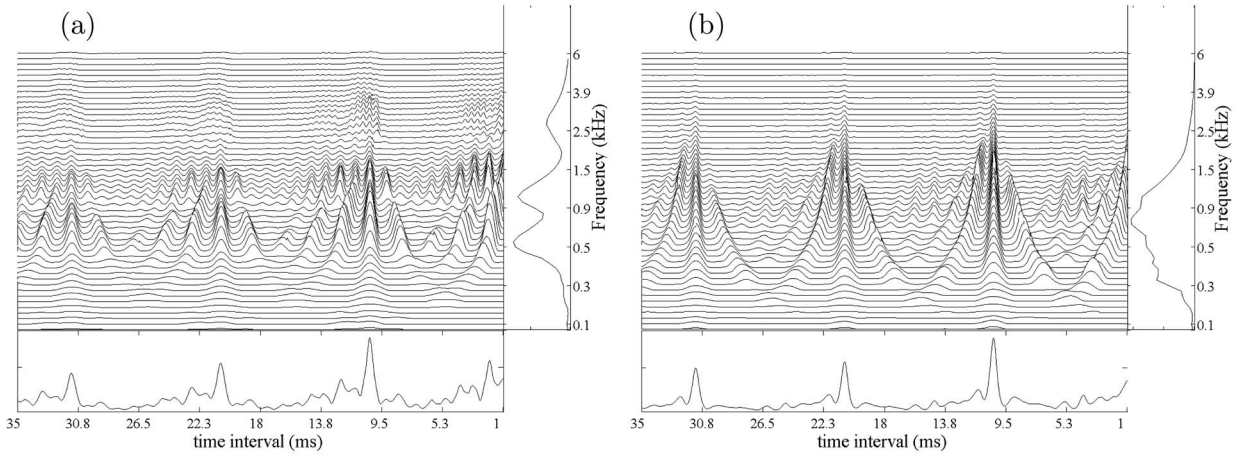


FIG. 5. Auditory images of the sustained portion of the original note for the baritone voice (left panel) and French horn (right panel).

tegration" is used to calculate the time intervals between peaks in the NAP and construct a time interval histogram for each of the filter channels (Patterson, 1994b). The array of time-interval histograms (one for each channel of the filterbank) is the auditory image; see Patterson *et al.* (1995, Fig. 2) for a discussion of the outputs of the different stages of processing. The auditory image is similar to an autocorrelogram (Meddis and Hewitt, 1991) but strobed temporal integration involves far less computation and it preserves the temporal asymmetry of pulse resonance sounds which autocorrelation does not (Patterson and Irino, 1998).

The auditory image is the central "waterfall" plot in Figs. 5(a) and 5(b); the vertical ridge in the region of 10 ms, and the resonances attached to it, provide an aligned representation of the impulse response of the instrument as it appears at the output of the auditory filterbank. The profile to the right of each auditory image is the average activity across time interval; it simulates the tonotopic distribution of activ-

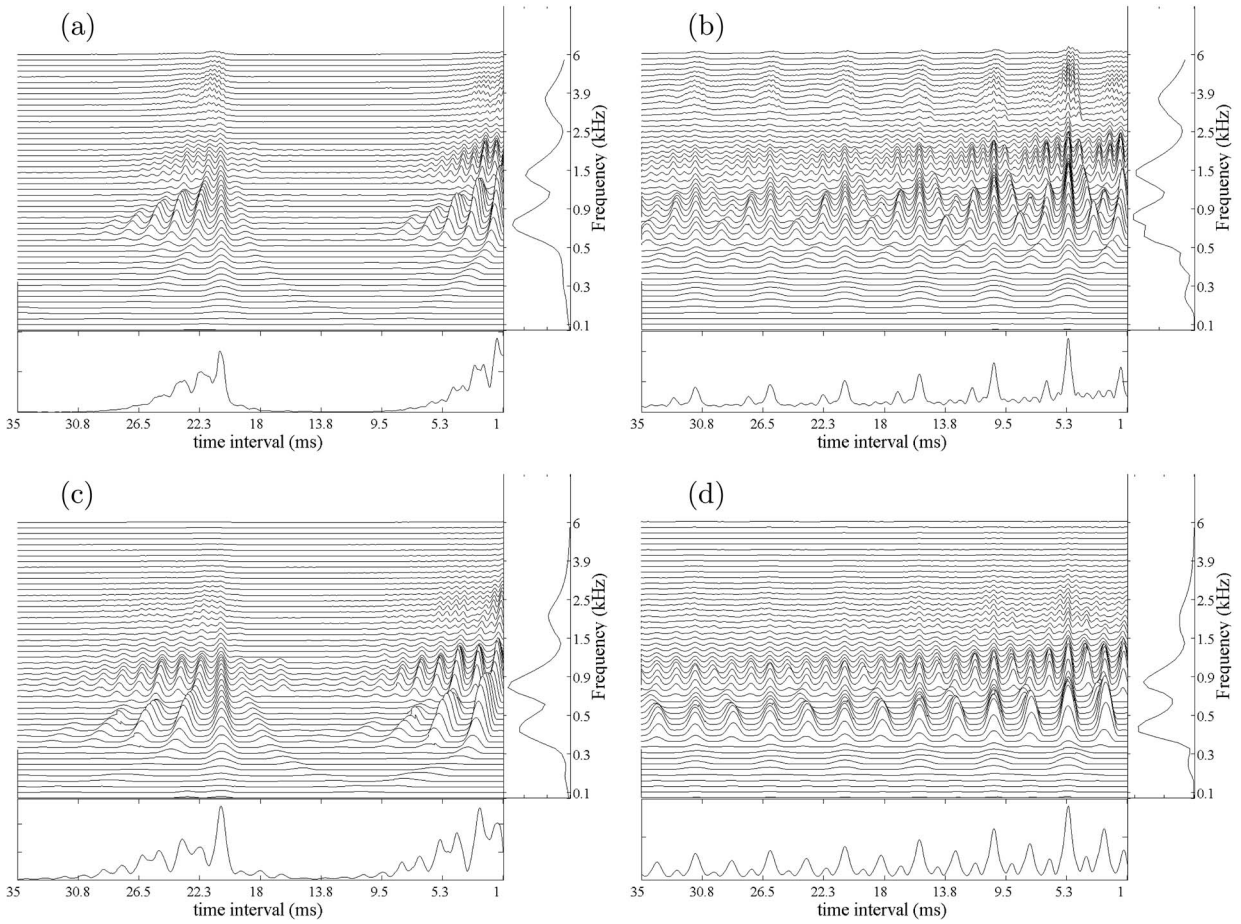


FIG. 6. Auditory images showing the effect of STRAIGHT on the baritone voice. The four panels show how the auditory image changes when the pulse rate and resonance scale are changed to the combinations presented by the outer four points in Fig. 4.

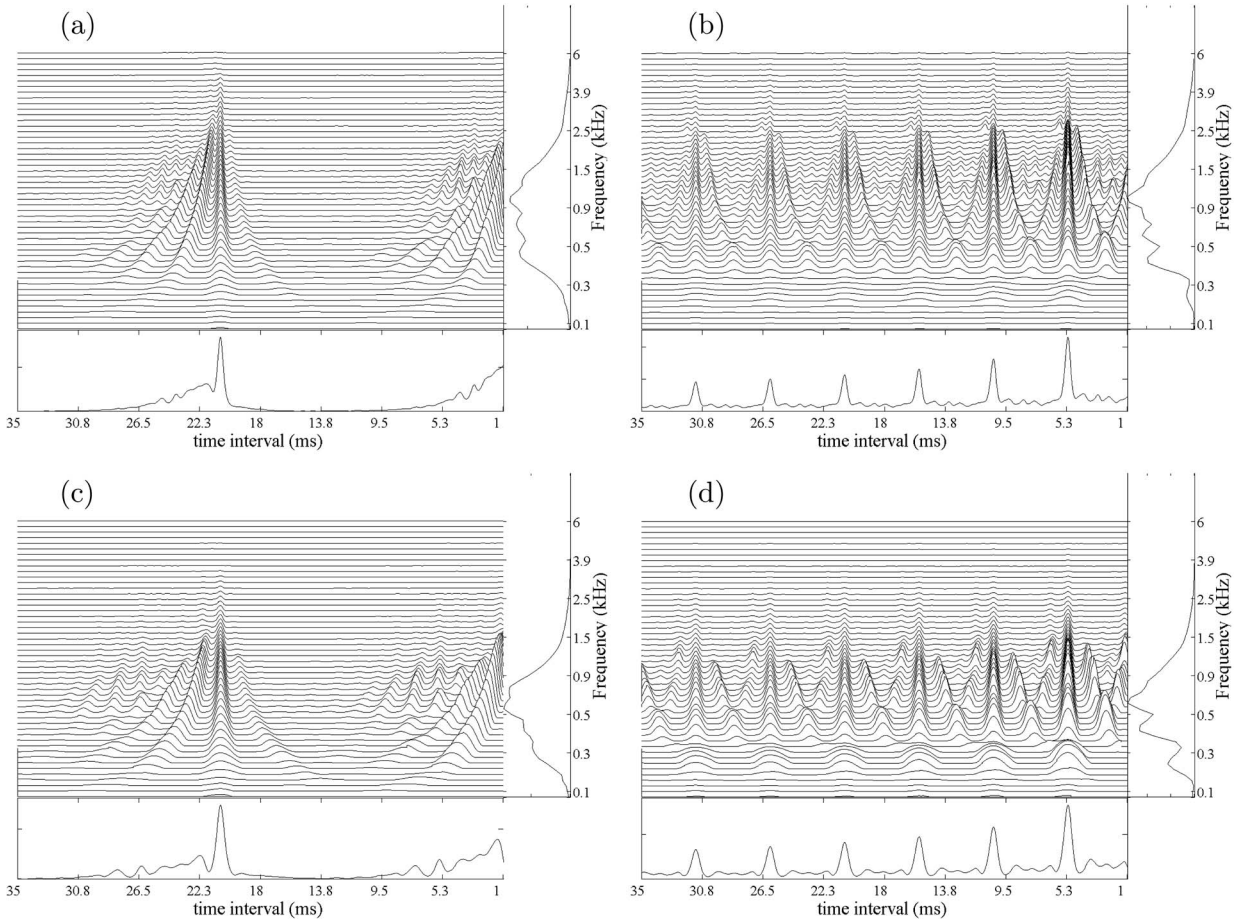


FIG. 7. Auditory images showing the effect of STRAIGHT on the French horn. The four panels show how the auditory image changes when the pulse rate and resonance scale are changed to the combinations presented by the outer four points in Fig. 4.

ity in the cochlea or the auditory nerve, and it is similar to an excitation pattern. The unit on the axis is frequency in kHz and it is plotted on a quasi-logarithmic "ERB" scale (Moore and Glasberg, 1983). The peaks in the spectral profile of the voice show the formants of the vowel. The profile below each auditory image shows the activity averaged across channel, and it is like a summary autocorrelogram (Yost *et al.*, 1996) with temporal asymmetry; the largest peak in the time-interval profile (in the region beyond about 1.25 ms) shows the period of the sound (G_2; 10 ms), much as the first peak in the summary autocorrelogram shows the pitch of a sound (Yost *et al.*, 1996). Comparison of the time-interval profiles for the two auditory images shows that they have the same PR, and thus the same temporal pitch (G_2). Comparison of the spectral profiles shows that the voice is characterized by three distinct peaks, or formants, whereas the horn is characterized by one broad region of activity.

The effect of STRAIGHT on the baritone voice is illustrated by the four panels in Fig. 6; they show how the auditory image changes when the PR and RS are altered to produce the values represented by the outer four stimulus conditions in Fig. 4. Comparison of the auditory image of the original baritone note in Fig. 5(a) with the images in the left-hand column of Fig. 6 shows that the PR has been reduced by an octave; the main vertical ridge in the image, and the largest peak in the time-interval profile (beyond 1.25 ms) have shifted from 10 to 20 ms. The panels in the right-hand column, show the images when the PR has been increased by an octave; the main ridge and the main peak now occur at 5 rather than 10 ms. Comparison of the time-interval profile for the original French horn note in Fig. 5(b), with the time-interval profiles in the left-hand and right-hand columns of Fig. 7, shows the same effect on PR; that is, the rate is reduced by an octave for both panels in the left-hand column and increased by an octave in both panels of the right-hand column. Together the figures illustrate that the pitch of pulse resonance sounds is represented by the position of the main vertical ridge of activity in the auditory image itself, and by the main peak in the time-interval profile (beyond about 1.25 ms).

Comparison of the auditory image of the original baritone note [Fig. 5(a)] with the images in the upper row of Fig. 6 shows the effect when STRAIGHT is used to reduce RS; the pattern of activity in the image, and the spectral profile, move up in frequency. For example, the second formant has shifted from about 0.9 to 1.2 kHz, although the vowel remains the same. In the lower row, the RS has been increased, with the result that the pattern of activity in the image, and the spectral profile, move down in frequency. Comparison of the original French horn note in Fig. 5(b) with the scaled versions in Fig. 7, shows the same effect on RS for the French horn; that is, the pattern moves up as a unit when RS

decreases, and down as a unit when RS increases. Moreover, a detailed examination shows that the patterns move the same amount for the two instruments. Together the figures illustrate that the RS information provided by the body resonances is represented by the vertical position of the pattern in the auditory image.

The auditory images and spectral profiles of the baritone voice notes in Fig. 6 and the French horn notes in Fig. 7 suggest that RS is a property of timbre. That is, the notes in each column of each figure have the same pitch, so if the notes were equated for loudness, then the remaining perceptual differences would be timbre differences, according to the usual definition. There are two components to the timbre in the current example, instrument family which distinguishes the voice notes from the horn notes, and instrument size which distinguishes the note in the upper row from the note in the bottom row, in each case. The two components of the timbre seem largely independent which supports the hypothesis that RS is a property of timbre. In this case, we might expect to find that listeners use RS to distinguish instruments, and since RS reflects the size of body resonances, we might expect listeners to hear RS differences as differences in instrument size. The question then arises as to how large a difference in RS is required to reliably discriminate two instruments, and this is the motivation for the discrimination experiment.

The mathematics of acoustic scale lends support to the hypothesis that RS is a property of auditory perception; however, the mathematics indicates that RS is a property of sound itself, rather than a component of timbre. We will return to this topic in the Discussion. The purpose of the current experiment is to demonstrate that RS provides a basis for discriminating the relative size of two instruments on the basis of their sounds. It is not crucial to the design of the experiments or the interpretation of the results, whether RS is a property of timbre or an independent property of sound itself.

3. Procedure and listeners

A two-interval forced-choice procedure was used to measure the JND for RS. Each trial consisted of two intervals with random tonal melodies played by one instrument. Short diatonic melodies were presented to convey the impression of tonal music and to preclude discrimination based on a simple spectral strategy, like tracking a single spectral peak. Each melody consisted of four different notes chosen randomly without replacement from the following five notes: G_i, A_i, B_i, C_{i+1} and D_{i+1}, where $i \in \{1,2,3\}$ depended on the condition presented in Table I. One of the stimulus intervals contained a melody with notes having a "standard" RS value, while the other interval contained a melody with notes having a slightly larger or slightly smaller RS value. The order of the intervals was randomized. The listener's task was to listen to the melodies and indicate which interval contained the smaller instrument. Since a change in RS represents a proportionate change in the spatial dimensions of the instrument, it is reasonable to assume that the perceptual cue is closely related to the natural perception of a change in size, particularly since the scale differences within a trial were relatively small. No feedback was given after the response.

Psychometric functions were generated about the standard RS value using six modified RS values, three below the standard and three above the standard, ranging between factors of $2^{-1/2}$ to $2^{1/2}$ for the cello, tenor sax and French horn, and between factors of $2^{-9/24}$ to $2^{9/24}$ for the voice. The ranges were chosen following pilot listening to determine the approximate range of the psychometric function for each instrument. A run from one of the five conditions in Table I consisted of 240 trials (four instruments $\times$ six points on the psychometric function $\times$ ten trials); the order of the trials was randomized. Each psychometric function was measured four times, so each of the six points on the function was contrasted with the standard 40 times. The set of points describes a two-sided psychometric function showing how much the RS of the instrument has to be decreased or increased from that of the standard for a specific level of discrimination.

Four listeners, aged between 20 and 35, participated in the experiment. There was one female and three males, all with normal hearing confirmed by an audiogram, and none of them reported any history of hearing impairment.

To familiarize the listeners with the task, a set of 50 trials was presented before each run. The RS differences in these trials were large to make discrimination easy. During the training, feedback was given indicating whether the response was correct or incorrect; a trial was judged correct if the listener chose the sound with the smaller RS. The listeners had some difficulty with the cello, so several retraining trials were presented after every 90 trials of a run, to remind listeners of the perceptual cue.

B. Results and discussion

A sigmoid function was fitted to the discrimination data for each instrument, in each of the five experimental conditions set out in Table I to characterize the psychometric function for that condition. The full set of psychometric functions is shown in Fig. 8. The layout of the five panels in this figure corresponds to that shown in Fig. 4. The data from the four listeners all exhibited the same overall form and so the data were averaged over listeners. The solid, dashed, dotted and dashed-dotted lines represent the psychometric functions obtained for the cello, saxophone, French horn and baritone voice, respectively. The JND values for the four instruments are presented in the individual panels of Fig. 8. The JND was taken to be the percentage increase in RS required to support 76% correct performance on the psychometric function, or since the psychometric function is symmetric, the percentage decrease in RS required to support 24% correct performance. The JNDs for the baritone voice are the smallest; they are about 3% in the upper three panels, and rise to 10% in the lower, right-hand panel. The JNDs for the French horn are more uniform around 7%, while those for the saxophone are around 12%.

The JNDs in the current experiment are largest for the cello. The JNDs are about 10% when the pitch is low and the instrument is small (or the pitch is high and the instrument is

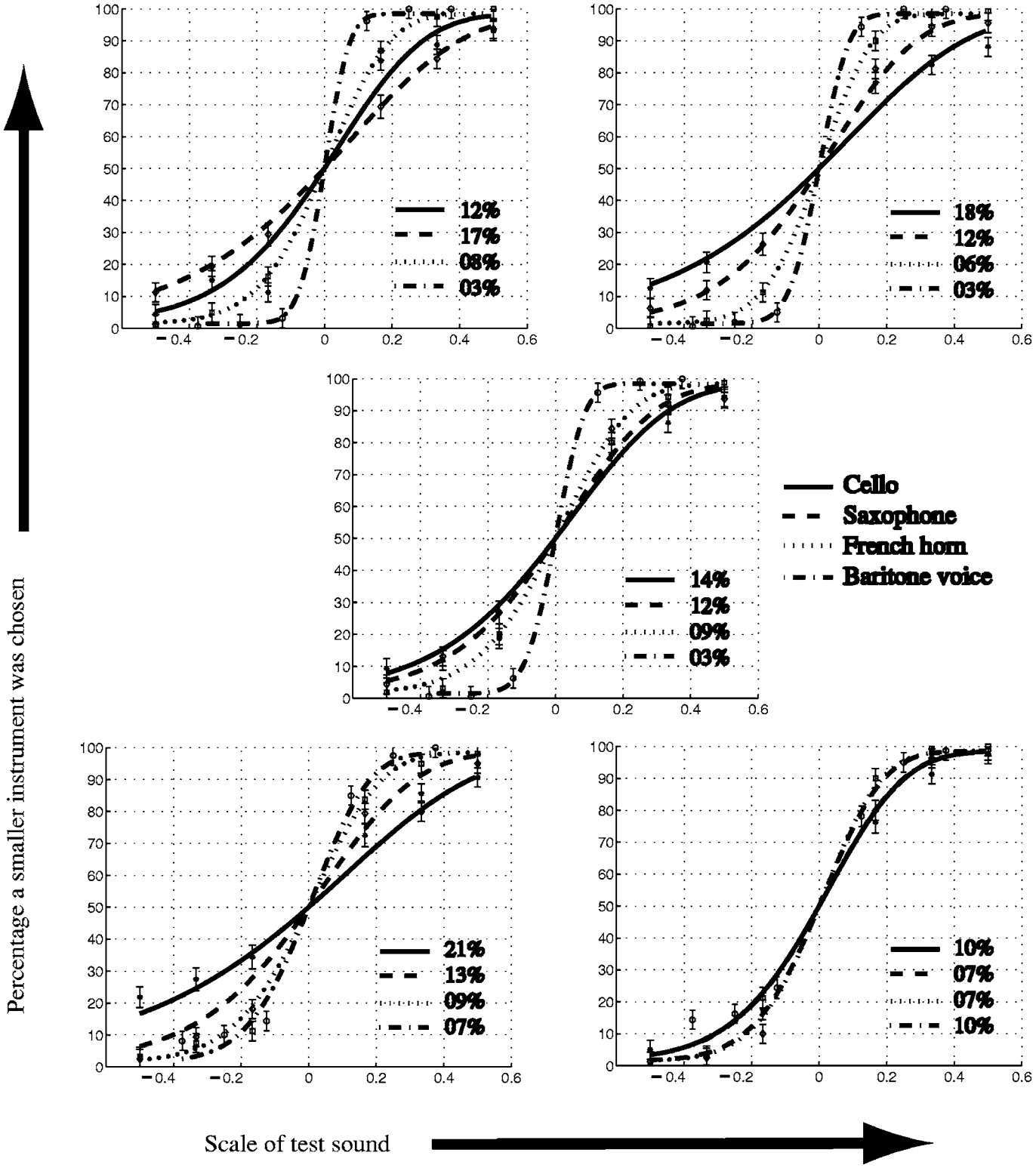


FIG. 8. Psychometric functions showing average percent correct for the five conditions in Table I. The positions of the five panels correspond to the positions in PR-RS space shown in Fig. 4. The abscissa is a base-2 logarithmic scale. The solid, dashed, dotted, and dashed-dotted lines are the psychometric functions for the cello, saxophone. French horn, and baritone voice, respectively. The JNDs are indicated separately in each of the panels.

large) and they increase to around 20% when the pitch is low and the instrument is large (or the pitch is high and the instrument is small). In the central condition, the JND is about 15%. The rise in the JND along the positive diagonal from about 15%–20% seems not unreasonable; in these conditions, the pitch rises as the instrument gets smaller (and *vice versa*) in the usual way, but these notes might be a little less familiar than those in the central condition. However, along the negative diagonal, the JND decreases from about 15% to about 10%, in conditions where pitch rises as the instrument gets *larger* (and *vice versa*). We did not find any particular reason for this reversal of what might have been expected.

TABLE II. The 16 instruments used for the identification experiment. The family name is presented at the top of each column.

Register	Strings	Woodwind	Brass	Voice
High	Violin	Soprano sax	Trumpet	Alto voice
Mid-High	Viola	Alto sax	Trombone	Tenor voice
Low-Mid	Cello	Tenor sax	French Horn	Baritone voice
Low	Contra bass	Baritone sax	Tuba	Bass voice

The RS discrimination experiments of Smith *et al.* (2005) included conditions with a baritone voice and the JNDs are comparable to those observed in the current experiment. This includes the condition in the bottom right-hand panel of Fig. 8 where the JND rises to about 10%. The JNDs for sensory dimensions are typically 10% or more; in vision, the JND for brightness is around 14%. (Cornsweet and Pinsker, 1956); in hearing, the JND for loudness is around 10% (Miller, 1947) and the JND for duration is around 10% (Abel, 1972). The small JNDs associated with pitch and visual acuity are the exception. So, the JNDs for RS appear to be as good as, or slightly better than, those for other auditory properties, which in turn supports the hypothesis that RS is a property of auditory perception.

One of the listeners was an amateur musician who plays the viola da gamba. His JNDs for the cello were much smaller than those for the other listeners, whereas his JNDs for the other instruments were about the same, and so, familiarity with an instrument may improve performance.

In summary, the results show that listeners are able to discriminate RS in instrument sounds. They can specify which is the smaller of two instruments from short melodies that differ in RS, and for the most part, they do not need feedback to support the discrimination. Within a family, the JND is fairly consistent, varying by no more than a factor of 2 across conditions, except for the baritone voice where it is a factor of 3 in one condition. Overall, listeners have slightly more difficulty when the instrument is large and plays a low-pitched melody.

IV. RECOGNITION OF INSTRUMENTS SCALED IN PR AND RS

The purpose of this experiment was to demonstrate that listeners can recognize versions of instrument sounds with a wide range of combinations of PR and RS; that is, listeners are robust to changes in RS as well as to changes in PR. The experiment includes PR and RS values within the normal range and beyond. Whereas Exp. I showed that listeners can discriminate changes in RS, Exp. II shows that listeners can recognize an instrument independent of its RS over quite a wide range of RS values.

A. Method

1. Stimuli and design

The same four instrument families were used in this experiment: string, woodwind, brass, and voice. Four members with different sizes were chosen in each family; the specific instruments are presented in Table II. The instruments in each row were chosen to have pitch ranges that largely overlap. For convenience, the four sizes are labeled by pitch range, or "register," as "High," "Mid High," "Low Mid" and "Low." The instruments were selected from the RWC database (Goto *et al.*, 2003), as before. Each note was extracted with its initial onset intact; the waves were truncated to produce a total duration of 350 ms, and a cosine-squared gate was applied to reduce the amplitude to zero over the last 50 ms of the sound. The use of natural sounds means that the notes contain cues such as vibrato and bow noise which can be used to recognize an instrument, in addition to the stationary pitch and timbre cues. In the selection of the sounds, we attempted to keep these cues to a minimum; however, one of the 16 instruments, the tenor voice, had a small amount of vibrato, which might be used as a recognition cue.

The vocoder STRAIGHT was used to modify the PR and the RS of the notes for all 16 instruments. The PR factors are presented in Table III, along with the RS factors for each register in Table II. The starting keys for the PR factors are indicated in the second column for each register. There were five PRs and five RSs, for a total of 25 conditions per instrument. Table III shows that the range of RS values is from $2^{-2/3}$ to $2^{2/3}$, that is, each instrument was resynthesized as an instrument that was from 0.7 to 1.6 times the original RS. The PR range was from $C_1 \approx 33$ Hz to $C_5 \approx 522$ Hz. The key note, C_1, is very close to the lower limit of melodic pitch, which is about 32 Hz (Krumbholz *et al.*, 2000; Pressnitzer *et al.*, 2001).

TABLE III. The conditions used in the experiment for each group given in Table II.

	Pulse rate						1/Resonance scale				
Register	Key	Factor					Factor				
High	C_4	2^{-1}	$2^{-5/12}$	2^{0}	$2^{7/12}$	2^{1}	$2^{-2/3}$	$2^{-1/3}$	2^{0}	$2^{1/3}$	$2^{2/3}$
Mid-High	G_3	2^{-1}	$2^{-7/12}$	2^{0}	$2^{5/12}$	2^{1}	$2^{-2/3}$	$2^{-1/3}$	2^{0}	$2^{1/3}$	$2^{2/3}$
Low-Mid	G_2	2^{-1}	$2^{-7/12}$	2^{0}	$2^{5/12}$	2^{1}	$2^{-2/3}$	$2^{-1/3}$	2^{0}	$2^{1/3}$	$2^{2/3}$
Low	C_2	2^{-1}	$2^{-5/12}$	2^{0}	$2^{7/12}$	2^{1}	$2^{-2/3}$	$2^{-1/3}$	2^{0}	$2^{1/3}$	$2^{2/3}$

2. Listeners and procedure

Four male listeners, aged between 20 and 35, participated in the experiment. Three of the four listeners also participated in Exp. I. They had normal hearing and they reported no history of hearing impairment. The stimuli were presented to the listeners over headphones at a level of approximately 60 dB SPL in a sound-attenuated booth.

A 16-alternative, forced-choice procedure was used to measure recognition performance. On each trial, a note from one instrument was played three times; the note had one of five PRs and one of five RSs as indicated in Table III. The listeners were presented with a graphical interface having 16 buttons labeled with instrument names in the layout shown in Table II. The family name was presented above each column of buttons. The listeners' task was to identify the instrument from one of the 16 options. Feedback was presented at the end of each trial but only with regard to the family of the instrument.

A run consisted of one trial for each instrument, of every combination of PR and RS; so there were $4\times4\times5\times5=400$ trials, and they were presented in a random order. Each listener completed ten replications over a period of several days. At the start of each replication, the 16 instrument sounds were presented to the listeners twice; and then again after every 100 trials.

Training was provided before the main experiment, to familiarize the listeners with the instrument sounds, and to see whether the listeners could identify the instruments prior to the experiment. The training was performed with the *original notes* rather than the notes manipulated by STRAIGHT. The training began with two families, namely, the woodwind and brass families. The notes of the two families were presented to the listeners twice. Then, the listener was tested with a mini run of 32 trials, in which each of the eight instruments was presented four times. The trials had the same form as those in the main experiment, and there was instrument-specific feedback after each note. The train-and-test sessions were repeated until performance reached 90% correct. After one successful test run, the string family was added to the training set, and the train-and-test procedure was continued until performance returned to 90% correct for the three families. Then the final family, voice, was added and the training continued until performance returned to 90% correct for all four families. All of the participants completed the training successfully with three, or fewer, cycles of train and test at the three stages of training.

B. Results and discussion

1. Effects of PR and RS on instrument recognition

The pattern of recognition performance was similar for all four listeners; the mean data are presented in Fig. 9 as performance contours. Each data point is the percent correct instrument recognition for the ten replications of each condition, averaged over the 25 conditions of PR and RS, the 16 instruments, and all four listeners—a total of $4\times16\times10=640$ trials per point. The data are plotted on a base-2, logarithmic axis both for the change in PR (the abscissa) and the change in RS (the ordinate). The data for PRs with multiplication factors of $2^{-7/12}$ and $2^{-5/12}$ were averaged and plotted above the value $2^{-1/2}$ on the abscissa. Similarly, the conditions with multiplication factors $2^{5/12}$ and $2^{7/12}$ were averaged and plotted above $2^{1/2}$. The contour lines show that performance is above 55% correct throughout the PR-RS plane, rising to over 80% correct in the center of the plane. This shows that listeners can identify the instruments reasonably accurately, even for notes scaled well beyond the normal range for that instrument. The chance level for this 16-alternative, instrument-identification task is about 6.25%; the chance level for correct identification of instrument family is about 25%; and when performance on the family-identification task is near 100%, then the chance level for instrument identification is closer to 25%. Either way, performance is well above chance throughout the PR-RS plane. Performance for notes with their original PR and RS values is a little below 90% correct for three of the four listeners, even though performance in the training sessions ended above 90% correct for all listeners. This is not really surprising since there were 400 different notes presented in each run.

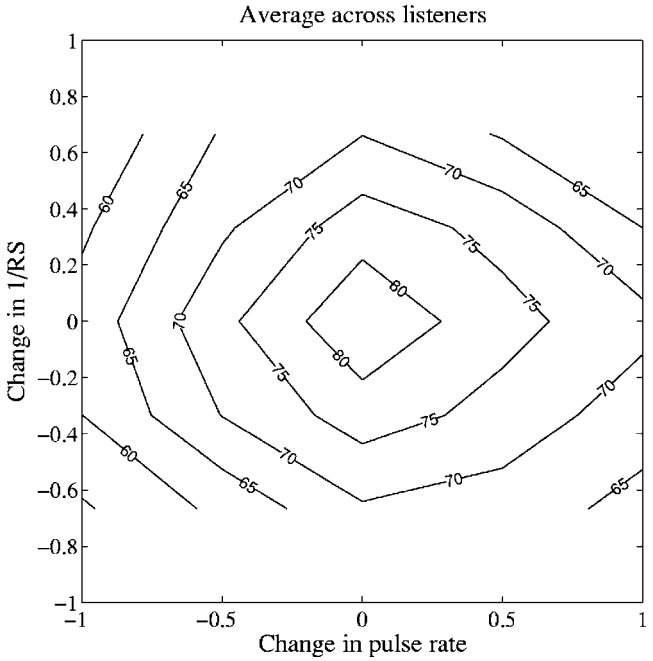


FIG. 9. Contours showing the percent-correct instrument recognition as a function of the change in PR and RS, averaged over all conditions, instruments and listeners. The data are plotted on a base-2 logarithmic axis for both the change in PR (the abscissa) and the change in RS (the ordinate).

The effects of PR and RS were not uniform across the four registers, and so contour plots for the ***individual registers*** are presented in Fig. 10. The figure shows that the contour plots for the Mid-High and Low-Mid registers are similar to the contour plot of overall performance (Fig. 9). But for the High register, peak performance is shifted to a higher PR and a smaller RS, and for the Low register, the effect is reversed—peak performance is shifted to a lower PR and a larger RS. These results show that listeners are likely to choose the smallest instruments for combinations with a high PR and a small RS, and they are likely to choose the largest instruments for combinations with a low PR and a large RS.

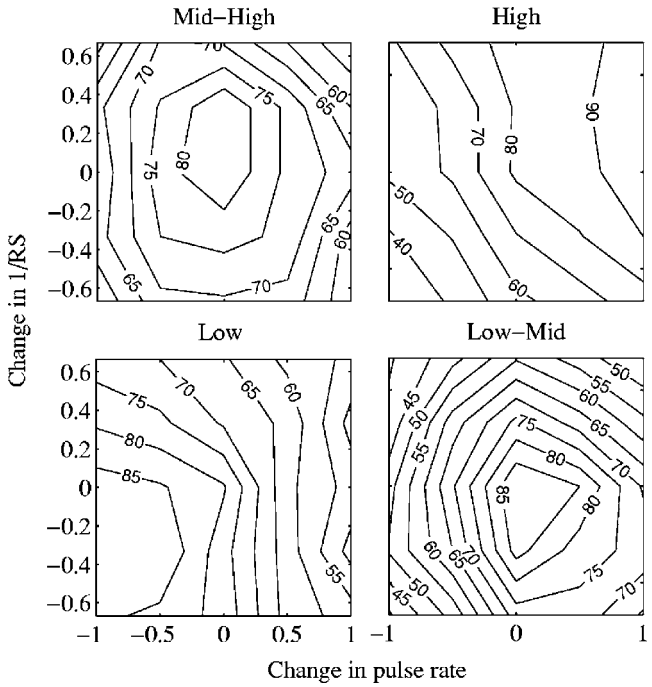


FIG. 10. Percent-correct contour plots for the four registers presented in Table II. The results are averaged across the 25 conditions and four listeners for each register.

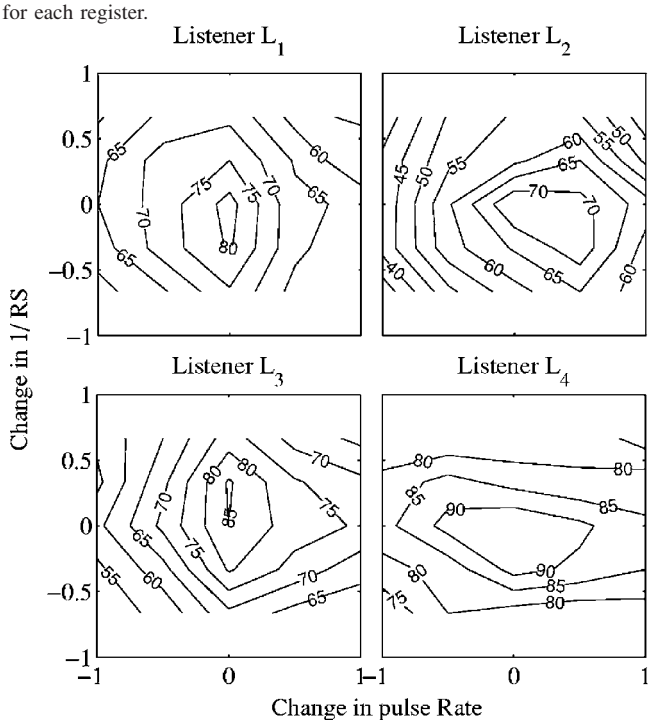


FIG. 11. Percent-correct contour plots for listeners L_1, L_2, L_3 and L_4. The results are averaged across the 25 conditions and 16 instruments in the experiment. The data are plotted on a base-2 logarithmic axis for both the change in PR (the abscissa) and the change in RS (the ordinate).

Contour plots for the ***individual listeners***, averaged across conditions and instruments, are presented in Fig. 11. Performance is well above chance for all of the listeners throughout the plane, and the contours show that the surface is a smooth hill with its peak in the center for all listeners. Nevertheless, there are distinct differences between the listeners. The performance of listeners L_1 and L_3 is roughly comparable, with scores above 80% correct for the original notes in the center, falling to around 60% for the most extreme combinations of PR and RS values. For both listeners, the manipulation of PR produced a greater reduction in performance than the manipulation of RS. Listener L_4 produced the best performance with greater than 90% correct over a large region around the original PR and RS values, and greater than 80% correct over most of the rest of the plane. Although this is probably due to musical training, it should be noted that this listener was the first author who prepared the stimuli for the experiment. For this listener, the manipulation of RS produced a greater reduction in performance than the manipulation of PR. Listener L_2 produced the worst performance which was, nevertheless, above 70% for a substantial region near the center of the plane, and only fell to below 50% for the lowest PRs. For this listener, the manipulation of PR and RS have roughly similar effects.

Two of the listeners were amateur musicians (L_3 and L_4) and the other two were nonmusicians (L_1 and L_2). Table IV provides a short description of each listener's musical involvement including their instrument, where appropriate, and their average percent correct for the original notes. The listener with the worst performance is a nonmusician (L_2) and the listener with the best performance is an amateur musician (L_4) which suggests that there is a link between performance and musicality. However, L_1 is a nonmusician and this listener's performance was comparable to that of L_3 who was an amateur musician, indicating that the distributions are probably more overlapping than separate.

2. Effects of PR and RS on family recognition

When listeners made a mistake in instrument identification, they typically chose a larger or smaller member of the same family whose PR and RS were compatible with those of the note they were presented. For example, when the PR of the viola was decreased and its RS increased, if the listener made an error, it was very likely that they would choose the cello as the instrument. Instrument-family recognition is analogous to vowel recognition; the four instrument families are like four vowel types. Within a family (e.g., brass) the notes of different instruments (e.g., the trumpet and tuba) are like tokens of one vowel (e.g., /i/) produced by

TABLE IV. Musical association of the listeners. The average correct percentage of the natural sounds, i.e., no pulse-rate or resonance-scale modification of the sounds, are also given for each listener.

Listener	Musical association	Affinity with instrument/interest	% Correct
1	Nonmusician	High interest in listening to music	82
2	Nonmusician	Average interest in listening to music	75
3	Amateur musician	Viola da gamba	85
4	Amateur musician	Singing, piano	95

TABLE V. Family-recognition performance in percent for the 25 conditions. The conditions are rearranged in the same order as the conditions presented in Figs. 9–12. The numbers in boldface represent the instruments in the normal range.

	Family recognition [%]					Mean of rows
	81	90	94	96	96	92
	88	92	96	97	97	94
↑ 1/RS	91	**96**	**99**	**98**	**97**	96
	94	97	98	98	94	96
	93	97	97	96	93	95
			PR →			
Mean of columns	89	94	97	97	96	95

people of different sizes (e.g., a small child and a large man, respectively). Accordingly, we reanalyzed the data scoring a response correct if the instrument family was correct. The pattern of family recognition was similar for the four instrument families and so the data were averaged over family. The data are presented in Table V, where family recognition is expressed as percent correct, averaged over listeners as well as families for the 25 combinations of PR change and RS change.

The rightmost column shows the mean RS values averaged over PR and the bottom row shows the mean PR values averaged over RS. Overall performance is *95% correct* on family identification as shown in the bottom right-hand corner of the table. Thus, the vast majority of instrument errors are within-family errors. Moreover, performance is uniformly high at around 95% correct over much of the PR-RS plane. Thus, instrument-family recognition is similar to vowel recognition in the sense that performance is very high over a large area of the PR-RS plane. Values below 95% correct are concentrated in the upper left section of the table where the PR and RS factors are both small. In this region, the instruments sound buzzy and the pitch for the low-register instruments falls below the lower limit of melodic pitch (Krumbholz *et al.*, 2000; Pressnitzer *et al.*, 2001). There were small differences between the instruments in this region; the average performance was roughly 70, 80, 90 and 100% correct, respectively, for the strings, brass, woodwinds and voice.

3. Trade-off between PR and RS in within-family errors

An analysis of the within-family, instrument-recognition errors is presented below with the aid of confusion matrices. The confusion data are highly consistent, so we begin by presenting a summary of the error data in terms of a surface that shows the trading relationship between PR and RS for within-family errors. The question is: Given that the listener has made an error, in what percentage of these cases does the listener choose a larger member of the family, and how does this percentage vary as a function of the difference in PR and RS between the scaled and unscaled versions of the note? Figure 12 shows the results averaged over instrument family as a contour plot of ***within-family errors*** where the score is the percentage of cases where the listener chose a larger member of a family, given a specific combination of PR and RS.

Consider first the 50% contour line. It shows that there is a strong trading relation between a change in PR and a change in RS. When we increase PR on its own, it increases the probability that the listener will choose a smaller member of the family; however, this tendency can be entirely counteracted by an increase in RS (making the instrument sound larger). Moreover, the contour is essentially a straight line in these log-log coordinates and the slope of the line is close to −1; that is, in log units, the two variables have roughly the same effect on the perception of which family member is producing the note. The same trading relationship is observed for all of the contours between about 20% and 80%, and the spacing between the lines is approximately equal. Together these observations mean that the errors are highly predictable on the basis of just two numbers, the logarithm of the change in PR and the logarithm of the change in RS.

In an effort to characterize the trading relationship, we fitted a two-dimensional, third-order polynomial to the data of Fig. 12 using a least-squares criterion; the surface is shown with the data points in the top panel of Fig. 13. The

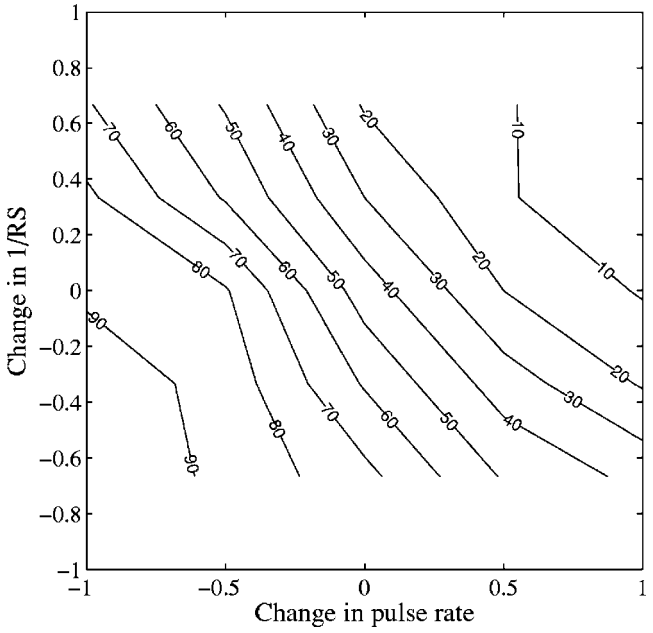


FIG. 12. Contours showing the percentage of within-family errors where the listener chose a larger member of the family as a function of the difference in PR and RS between the scaled and unsealed versions of the note. The data are plotted on a base-2 logarithmic axis for both the change in PR (the abscissa) and the change in RS (the ordinate).

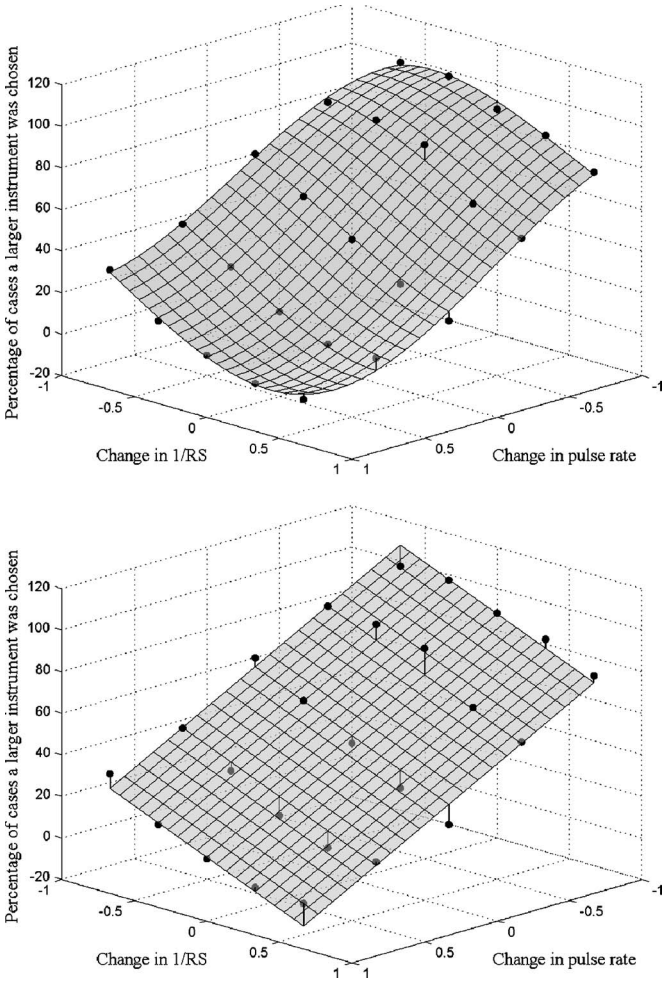


FIG. 13. Third-order polynomial and planar surfaces fitted to the within-family error data. The surfaces show the percentage of cases where the listener chose a larger instrument as a function of the difference in PR and RS between the scaled and unsealed versions of the note. The data are plotted on a base-2 logarithmic axis for both the change in PR (the abscissa) and the change in RS (the ordinate).

data points are shown by black spheres and their deviation from the surface is shown by vertical lines. The surface fits the data points very well. The panel shows that the central section of the surface is essentially planar; the corners bend up at the bottom and down at the top due to floor and ceiling effects. The fact that the central part of the surface is planar means that we can derive a simple expression for the trading relationship that characterizes the data, except at the extremes, by fitting a plane to the data. The plane is shown with the data points in the bottom panel of Fig. 13. The fit is not much worse than that provided by the surface in the upper panel. The rms error increases a little from 13 to 30, but this is still relatively small, and the plane is described by three coefficients, whereas the curved surface requires ten.

The equation for the plane is $z=-38x-30y+50$, where z is the "percentage of cases that a larger instrument was chosen," x is $\log_2$ (change in PR), and y is $-\log_2$(change in RS). The plane shows that, for any point in the central range, (1) an increase in PR of −0.5 log units (six semitones) will increase the probability of choosing a larger instrument within the family by about 15%, and (2) a change in PR of PR log

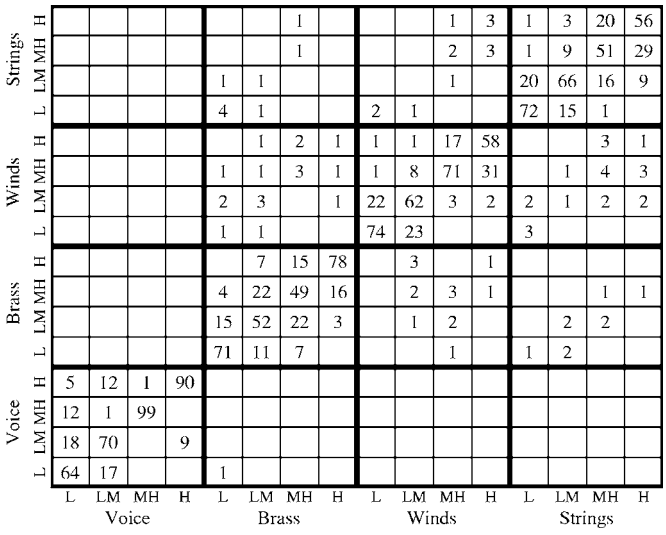


FIG. 14. Confusion matrix showing recognition performance for the 16 instruments in the experiment. The abscissa shows the instrument presented to the listener by family name and register within family. The entries in each column show the percentage of times each of the 16 instrument names was chosen in response to the instrument sound presented.

units can be counteracted by a change in RS of 1.3 PR log units. This means that, when measured in log units, the effect of a change in PR on the perception of size is a little greater than the effect of a change in RS. If we express the relationship in terms of JNDs instead of log units, the relative importance of RS increases. The JND for RS was observed in Exp. I to be about 10%. The JND for PR is more like 1% (Krumbholz *et al.*, 2000; Fig. 5). So, one JND in RS has about the same effect on the perception of size as eight JNDs in PR.

4. Confusion matrices

The confusion matrix for the 16 instruments in the experiment is presented in Fig. 14. The instrument presented to the listener is indicated on the abscissa by family name and register within family. The entries in each column show the percentage of times each instrument name was used in response to a given instrument on the abscissa (again by family and register). The entries are averaged over the 25 PR × RS conditions and over all four listeners. The figure shows that the majority of the responses are correct (68%); they appear on the positive diagonal. Moreover, the diagonals immediately adjacent to the main diagonal contain 68% of the remaining errors. So the most common error by far is a within family error to one of the instruments that is closest in size. The figure also shows that, away from the main diagonal, the errors still occur largely within family blocks. There are essentially no confusions between the human voice and the other instrument families; "voice" is never used as a response when another instrument is presented, and "voice" is always the response when a voice is presented. There is a low level of confusion between the remaining three instrument families, but it does not appear that any family confusion is more likely than any other. These family confusions also appear to be reasonably symmetric.

5. Source information for the recognition experiment

The stimuli in these experiments are natural sounds, and as such they could contain cues other than timbre, PR, and RS which could be used to recognize an instrument. Although we chose the instruments and notes to minimize these extra cues, one of the 16 instruments, the tenor voice, had a small amount of vibrato. An analysis of the individual instrument data showed that performance for the tenor voice was somewhat better than for the other instruments (between 95% and 100% for all but the most extreme combinations of PR and RS), which suggests that the listeners probably did use the vibrato cue to assist in identifying this instrument. This was the only instrument for which performance was largely independent of PR and RS. For the other instruments, performance was lower and graded as illustrated in Figs. 9 and 10. Figure 9 shows that performance decreases as PR and RS are manipulated from their initial values. Figure 10 shows that the effects of PR and RS were not uniform across the four registers. The contour plots for the High register show peak performance is shifted away from the center to a higher PR and a smaller RS. For the Low register, the effect is reversed. The results indicate that listeners' judgments are strongly affected by the specific values of PR and RS, indicating that the extra timbre cues associated with the use of natural sounds did not dominate the judgments.

V. DISCUSSION

The purpose of the current study was to extend previous research on the perception of scaled speech sounds to the perception of musical notes. The question arose following Cohen's (1993) development of a transform to describe the scale information in sounds, such as the Doppler effect in echolocation, and the size related changes that occur in speech sounds as the vocal tract grows in length. Cohen has argued that "scale" is a physical attribute of sound just like time and frequency. If this is the case, and if the auditory system has a general mechanism for processing acoustic scale, we might expect to find that (a) the fine discrimination of resonance scale observed with vowel sounds is also possible with the notes of sustained-tone instruments, since they also produce pulse-resonance sounds, and (b) listeners are able to recognize the family of an instrument from notes scaled in PR and RS over a wide range of PR and RS values.

A. Discrimination of RS in sustained-tone instruments

Smith *et al.* (2005) showed that listeners can discriminate a small RS difference between two vowel sequences, and Ives *et al.* (2005) showed that listeners can discriminate a small RS difference between two syllable phrases. They both interpreted the fact that the JND is small over a large portion of the PR-RS plane as supporting Cohen's postulate that scale is a property of sound. Experiment I of the current paper showed that listeners are also able to discriminate a change in the RS of musical notes when presented two short melodies with slightly different RSs. The JND values are slightly larger than those obtained with speech sounds (Smith *et al.*, 2005; Ives *et al.*, 2005), and the JNDs are greater for some instruments than others, but they are on the same order as those for speech sounds (about 10%). Thus, discrimination of RS in the notes of sustained-tone instruments provides further support for the postulate that scale is a property of sound and that the auditory system has a mechanism for processing it.

B. Recognition of scaled instruments

Smith *et al.* (2005) showed vowel recognition is robust to changes in PR and RS over the normal range of experience, *and* well beyond the normal range. They argued that their data support the hypothesis of Irino and Patterson (2002) that the auditory system has general purpose mechanisms for normalizing the PR and RS of sounds before vowel recognition begins. This was contrasted with the hypothesis of Assmann *et al.* (2002) that listeners learn vowel categories by experience; they had shown that a neural net could learn to recognize scaled vowels from within the range presented during training. The focus of the discussion was the pattern of recognition performance in the region of the PR-RS plane where the combination of PR and RS is beyond normal experience. Neural nets have no natural mechanism for generalizing to stimuli whose parameter values are beyond the range of the training data (LeCun and Bengio, 1995; Wolpert, 1996a, b); their success is largely attributable to interpolating between values of the training data. The performance of automatic speech recognition systems improves if the system is adapted to the speech of individual speakers by expanding or contracting the frequency dimension to fit the VTL of the speaker (Welling and Ney, 2002). Thus, human performance on scaled vowels would be expected to deteriorate in the region beyond normal experience if they were using a neural net for learning and recognition without prior normalization. A general purpose scaling mechanism does not depend on training, and any observed limitations on performance are assumed to arise from other constraints. In support of the statistical learning hypothesis, Assmann *et al.* (2002) noted that performance does drop off somewhat for stimuli with combinations of PR and RS beyond normal experience; in response, Smith *et al.* (2005) pointed out the fall off in performance only occurs for extreme combinations of PR and RS, and that performance remains high in regions of the PR-RS plane well beyond normal experience. The purpose of this section of the Discussion, however, is not to try and decide between these two hypotheses concerning the recognition of scaled vowels and scaled musical notes. It seems likely that, even if there is a general mechanism that normalizes sound patterns before the commencement of recognition processing, there is also a statistical learning mechanism at the recognition level. Thus, the question is not "Which mechanism do we have?" but rather "How do the two mechanisms work together to produce the performance we observe." The purpose of this part of the Discussion, then, is to review the pattern of musical note recognition with respect to everyday musical experience, and to compare it with vowel recognition and vowel experience.

There is a notable difference between the human vocal tract and other musical instruments in terms of our experi-

ence of resonance scale. Humans grow continuously, whereas instruments come in a limited number of fixed sizes. Constraints on the production and playing of musical instruments mean that the instruments of one class (e.g., French horn or violin) are all pretty similar in size. Thus, the distribution of resonance scales that we experience is concentrated on a small number of widely spaced values. We experience the voice from a distribution of vocal-tract lengths that is relatively smooth and continuous. With regard to the PR-RS plane, our experience of an instrument family is limited to examples from a small number of horizontal bands that are relatively narrow in the RS dimension, one band for each instrument within the family. In Table V, the bands that provide our experience of the 12 brass, string, and woodwind instruments are all represented by the four bold numbers in the center row. Strictly speaking, when the RS of one of these instruments is scaled up or down, it leads to a new instrument (within the same family) whose notes are not part of everyday experience. The lowest PR is not in bold font because this PR was below the normal range for all of the instruments. Recognition for the voice family was essentially uniform over the plane, so it can simply be neglected in this discussion.

The table shows that increasing or decreasing RS has very little effect on performance. There is a small reduction in performance for notes with the smallest RS (top row) but it is limited to notes with low PRs. Overall, performance for notes from novel instrument sizes is 95.6% correct, compared with 97.5% correct for the traditional instruments. Thus, the instrument data do not show the pattern that would be expected for a system based solely on statistical learning. They are more compatible with a general normalization mechanism.

There is an additional aspect to this argument: When the notes of the highest instrument in each of the brass, string, and woodwind families are scaled down in RS to simulate a smaller instrument, the RS of the notes is beyond the normal range for that family, *for all of the PR conditions*. Similarly, when the notes of the lowest instrument in each family are scaled up in RS to simulate a larger instrument, the RS is beyond the normal family range *for all PR conditions*. The extent to which the notes were scaled beyond the normal range of the family in Exp. II, is about the same as the extent to which the vowels were scaled beyond the normal range of human speech in Smith *et al.* (2005). The pattern of recognition performance for the brass, string, and woodwind instruments is like that of the vowels in Smith *et al.* (2005), inasmuch as near ceiling performance extends well beyond the range of normal experience. To achieve this level of recognition performance, a statistical learning mechanism would have to extrapolate well beyond its experience, which is not something that they can typically do. It seems more likely that the learning mechanism is assisted by a general normalization mechanism which makes the family patterns similar before learning and recognition.

C. Interaction of PR and RS in the perception of instrument size

When listeners made a mistake during instrument identification, they typically chose a larger or smaller member of the same family—a member whose PR and RS are compatible with those of the note they were presented. This suggests that much of the distinctiveness of instruments within a family is due to the combination of PR and RS in the notes they produce. Analysis of the confusion data indicated that there is a strong trading relationship between PR and RS; an increase in PR can be counteracted by an increase in RS and *vice versa*, in judgments of instrument size. Similar effects were observed by Fitch (1994) and Smith and Patterson (2005), both of whom found that the PR and RS of speech sounds interact in the estimation of speaker size and speaker sex. Interaction of PR and RS is also evident in the data of Feinberg *et al.* (2005), who investigated the influence of PR and RS on the size, masculinity, age, and attractiveness of human male voices. They found that decreasing PR and increasing RS both increased the perception of size, and that a combination of a decrease in PR and an increase in RS has an even larger effect on the perception of size. Together these results suggest that the perception of size information in musical sounds is similar to that observed with speech sounds.

D. Size/scale information in other sources

The idea that sounds from physical sources contain scale information, and that the mammalian auditory system uses some form of Mellin transform to normalize sounds for scale, was first proposed by Altes (1978). He was particularly interested in echolocation by bats and dolphins, and the fact that the Mellin transform would provide for optimal processing of linear, period-modulated signals. The magnitude information of the Mellin transform provides a representation of the source that is independent of the speed of the source relative to the observer. The phase information in the Mellin transform can be used to estimate the rate of dilation of the signal for echolocation. But Altes (1978) also recognized that there was scale information in speech sounds and that the Mellin transform might be useful for producing a scale-invariant representation of speech sounds.

Recent studies have shown that there is RS information in a range of vertebrate communication sounds: for example, birds (Fitch, 1999), deer (Fitch and Reby, 2001; Reby and McComb, 2003), lions (Hast, 1989), dogs (Riede and Fitch, 1999) and macaques (Fitch, 1997). Several of these papers also demonstrate that animals are sensitive to differences in RS and they interpret RS as size information. For example, Fitch and Kelly (2000) showed that cranes attend to changes in the formant frequencies of species-specific vocalizations, and they hypothesized that the formants provide cues to body size. Reby and McComb (2003) showed that male red deer with a low fundamental frequency and a long vocal tract had a greater chance of reproductive success. Finally, Gazanfar *et al.* (2006) have recently reported that adult macaques presented with silent videos of a large and a small macaque, and

a simultaneous recording of the call of a large or small macaque, look preferentially to the video that matches the sound in terms of macaque size.

There are also studies to show that there is RS information in the sounds produced by inanimate objects other than musical instruments, and that humans perceive the RS information in terms of source size. Houben *et al.* (2004) showed that listeners can discriminate changes in the size of wooden balls from the sound of the balls rolling along a wooden surface. Grassi (2005) dropped wooden balls of different sizes (that acted as pulse generators) on baked clay plates (that resonated), and asked listeners to estimate the size of the *ball*. The size estimates were correlated with ball size but the form of the clay plate influenced the size estimates, so it is not really clear in this case how the resonance scale of the sound was controlled or perceived. It was also the case that the larger balls produced louder sounds which probably influenced the judgments as well. Finally, Ottaviavi and Rocchesso (2004) synthesized the sounds of spheres and cubes of different sizes and demonstrated that listeners can discriminate changes in the volume of the object from the synthesized sounds.

In summary, studies with animal calls and the sounds of inanimate sources support the hypothesis that the mammalian auditory system has a general purpose mechanism for processing the acoustic scale information in natural sounds.

E. Resonance scale: A property of timbre or a property of sound?

In Sec. III A 2, it was noted that the auditory images and spectral profiles of the baritone voice notes in Fig. 6 and the French horn notes in Fig. 7 suggested that RS is a property of timbre, according to the standard definition. That is, two instruments from one family playing the same note (same pitch) with the same loudness would nevertheless be distinguishable by the difference in RS, and so RS is a component of timbre perception. It was also noted that, although the mathematics of acoustic scale supports the hypothesis that RS is a property of auditory perception, the mathematics indicates that RS is a property of sound itself, independent of the definition of timbre. We pursue this distinction briefly in this final subsection of the Discussion.

In mathematical terms, the transformation of a sound wave as it occurs in air into basilar membrane motion on a quasi-logarithmic frequency scale, is a form of wavelet transform involving a warping operator that has the effect of segregating the RS information from the remaining information in the sound, which is represented by the *shape* of the magnitude distribution in the spectral profile. In mathematical terms, the spectral profile is a *covariant* scale-timbre representation of the sound (Baraniuk and Jones, 1993, 1995); that is, a representation in which the timbre information about the structure of the source is coded in the shape of the distribution, and the RS information is coded, separately, in terms of the position of the distribution along the warped-frequency dimension. The importance of the covariance representation is the demonstration that the two forms of information can be segregated, because once segregated, the RS can be separated from the magnitude information using standard Fourier techniques. Moreover, the resulting magnitude distribution is a scale *invariant* representation of the timbre information (Baraniuk and Jones, 1993, 1995). For example, Irino and Patterson (2002) have shown how the two-dimensional auditory image can be transformed into a two-dimensional, scale-covariant, size shape image (SSI), and then subsequently, into a two-dimensional, scale invariant, Mellin Image. They illustrated the transforms with vowel sounds, but the sequence of transformation would be equally applicable to the notes of sustained-tone instruments.

The details of these alternative mathematical representations of sound, and their potential for representing auditory signal processing, are beyond the scope of this paper. The important point for the current discussion about timbre is that the mathematics indicates that acoustic scale is actually a property of sound itself (Cohen, 1993). This suggests that acoustic scale might better be regarded as a separate property of auditory perception, rather than an internal property of timbre. Just as the repetition rate of a sound is heard as pitch and the intensity of the sound is heard as loudness, so it appears that the acoustic scale of a sound has a major effect on our perception of the size of a source. Acoustic scale is not the only contributor to the perception of source size; clearly, the average pitch of the voice contributes to speaker size, and the average pitch of an instrument contributes to the perception of its size. Nevertheless, isolated changes in resonance scale are heard as changes in source size.

In order to interpret acoustic scale information in terms of speaker size or instrument size, the listener has to have some experience with people and instruments, but the mapping between acoustic scale and resonator size is very simple in speech and music. The relationship between acoustic scale and VTL is essentially linear, and speaker size is very highly correlated with speaker height (Fitch and Giedd, 1999). Similarly, acoustic scale is linearly related to resonator size in musical instruments which is directly related to overall instrument size. For practical reasons, instrument makers choose to achieve some of the resonance scaling by means of changes in mass and thickness, but there remains a very high correlation between RS and instrument size within instrument families. So it is not surprising that the RS appears to function as a property of auditory perception like pitch and loudness, and not just as a component of timbre.

The purpose of this paper was to illustrate that RS provides a basis for size discrimination in instrument sounds, and that appropriate processing of RS information minimizes cross family confusion. It is not crucial to the design of the experiments or the interpretation of the results, whether the mathematics of acoustic scale eventually leads to a modification of the definition of timbre to exclude RS. Such a change in perspective would, however, appear to be worth considering given that we perceive instrument sounds in terms of families and members within families which differ in terms of their register. Part of the register information is average pitch, but another important part is resonance scale, and like pitch, resonance scale appears to function as a property of auditory perception.

VI. CONCLUSIONS

Listeners can detect relatively small changes (about 10%) in the resonance scale of the notes produced by sustained-tone instruments, such as the instruments in the string, woodwind, brass, and voice families. Listeners are also able to recognize instrument sounds scaled over a wide range of pulse rates and resonance scales, including combinations beyond the normal range. Both pulse rate and resonance scale contribute to the perception of instrument size, as expected, and there is a strong trading relationship between pulse rate and resonance scale in instrument identification.

The results of the experiments suggest that pulse rate and resonance scale play similar roles in the perception of speech and music. As such, the results support the hypothesis that the auditory system applies some kind of scale transform to all sounds, to segregate the RS information and produce scale covariant and/or scale invariant representations of the sound source—representations that would be expected to enhance recognition performance (e.g., Welling and Ney, 2002).

ACKNOWLEDGMENTS

We would like to thank the associate editor and the reviewers for their helpful comments. The research was supported by the U.K. Medical Research Council (G9901257, G9900369, G0500221), and ONRIFO (Grant No. N00014-03-1-1023).

[1]The term "pulse resonance" is intended to describe the waves produced by source-filter systems like the voice and sustained-tone musical instruments, independent of the system that produces it. The term pulse-resonance describes this category of sounds in terms of what we believe matters to the auditory system in the analysis of the sounds, rather than in terms of how the sounds are produced physically by the system that produces the sound.

[2]Cohen (1993) argues that scale is a physical attribute of a signal just like time and frequency. It should be noted, however, that scale is not orthogonal either to time or frequency; a change of scale in a signal has an effect both on time and frequency.

[3]http://www.pdn.cam.ac.uk/groups/cnbh/teaching/sounds-movies/melodies PRRS-files/slide0305.htm

Abel, S. M. (**1972**). "Duration discrimination of noise and tone bursts," J. Acoust. Soc. Am. **51**, 1219–1223.

Altes, R. A. (**1978**). "The Fourier-Mellin transform and Mammalian hearing," J. Acoust. Soc. Am. **63**, 174–183.

Assmann, P. F., and Katz, W. F. (**2005**). "Synthesis fidelity and time-varying spectral change in vowels," J. Acoust. Soc. Am. **117**, 886–895.

Assmann, P. F., Nearey, T. M., and Scott, J. M. (**2002**). "Modeling the perception of frequency-shifted vowels," *1CSLP 02*, 425–428.

Baraniuk, R. G., and Jones, D. L. (**1993**). "Warped wavelet basis: Unitary equivalence and signal processing," *Proc. IEEE ICASSP'93*, 320–323.

Baraniuk, R. G., and Jones, D. L. (**1995**). "Unitary equivalence: A new twist on signal processing," IEEE Trans. Signal Process. **43**, 2269–2282.

Benade, A. H. (**1976**). *Fundamentals of Musical Acoustics* (Oxford University Press, Oxford).

Benade, A. H., and Lutgen, S. J. (**1988**). "The saxophone spectrum," J. Acoust. Soc. Am. **83**, 1900–1907.

Chiba, T., and Kajiyama, M. (**1941**). *The vowels, its nature and structure* (Tokyo-Kaiseikan, Tokyo).

Cohen, L. (**1993**). "The scale representation," IEEE Trans. Signal Process. **41**, 3275–3292.

Cornsweet, T. N., and Pinsker, H. M. (**1956**). "Luminance discrimination of brief flashes under various conditions of adaptation," J. Physiol. (London) **176**, 294–310.

Dinther, R. van, and Patterson, R. D. (**2004**). "The perception of size in four families of instruments; brass, strings, woodwind and voice," BSA, **P62**.

Dinther, R. van, and Patterson, R. D. (**2005**). "The perception of size in musical instrument sounds," J. Acoust. Soc. Am. **117**, 2374.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., and Perrett, D. I. (**2005**). "Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices," Anim. Behav. **69**, 561–568.

Fitch, W. T. (**1994**). Vocal tract length perception and the evolution of language, Ph.D. thesis, Brown University.

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. (**1999**). "Acoustic exaggeration of size in birds by tracheal elongation: Comparative and theoretical analyses," J. Zool. (London) **248**, 31–49.

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.

Fitch, W. T., and Kelly, J. P. (**2000**). "Perception of vocal tract resonances by whooping cranes, Grus americana," Ethology **106**, 559–574.

Fitch, W. T., and Reby, D. (**2001**). "The descended larynx is not uniquely human," Proc. R. Soc. London, Ser. B **268**, 1669–1675.

Fletcher, N. H. (**1978**). "Mode locking in nonlinearly excited inharmonic musical oscillators," J. Acoust. Soc. Am. **64**, 1566–1569.

Fletcher, N. H. (**1999**). "The nonlinear physics of musical instruments," Rep. Prog. Phys. **62**, 723–764.

Fletcher, N. H., and Rossing, T. D. (**1998**). *The physics of musical instruments* (Springer-Verlag, New York).

Gazanfar, A. A., Turesson, H. K., Maier, J. X., Dinther, R. van, Patterson, R. D., and Logothetis, N. K. (**2006**). "Formants as cues to body size in a non-human primate: Substrates for the evolution of speech," Anim. Behav. (submitted).

Glasberg, B. R., and Moore, B. C. J. (**2002**). "A model of loudness applicable to time-varying sounds," J. Audio Eng. Soc. **50**, 331–342.

Goto, M., Hashiguchi, H., Nishimura, T., and Oka, R. (**2003**). "RWC music database: Music genre database and musical instrument sound database," *ISMIR*, 229–230.

Grassi, M. (**2005**). "Do we hear size or sound? Balls dropped on plates," Percept. Psychophys. **67**, 274–284.

Hast, M. (**1989**). "The larynx of roaring and non-roaring cats," J. Anat. **163**, 117–121.

Helmholtz, H. L. F. von (**1877**). *On the sensations of tone*, Translated by A. J. Ellis (Dover, New York, **1954**).

Houben, M. M. J., Kohlrausch, A., and Hermes, D. J. (**2004**). "Perception of the size and speed of rolling balls by sound," Speech Commun. **43**, 331–345.

Hutchins, C. M. (**1967**). "Founding a family of fiddles," Phys. Today **20**, 23–27.

Hutchins, C. M. (**1980**). "The new violin family," in *Sound Generation in Winds, Strings, Computers* (Royal Swedish Academy of Music), pp. 182–203.

Irino, T., and Patterson, R. D. (**2002**). "Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-Mellin transform," Speech Commun. **36**, 181–203.

Irino, T., and Patterson, R. D. (**2006**). "A dynamic, compressive gammachirp auditory filterbank," IEEE Trans. Speech and Audio Processing (in press).

Ives, D. T., Smith, D. R. R., and Patterson, R. D. (**2005**). "Discrimination of speaker size from syllable phrases," J. Acoust. Soc. Am. **118**, 3816–3822.

Kawahara, H., and Irino, T. (**2004**). "Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation," in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluer Academic, Massachusetts), pp. 167–180.

Kawahara, H., Masuda-Kasuse, I., and de Cheveigne, A. (**1999**). "Restructuring speech representations using pitch-adaptive time-frequency smoothing and instantaneous-frequency based F0 extraction: Possible role of repetitive structure in sounds," Speech Commun. **27**, 187–204.

Krumbholz, K. Patterson, R. D., and Pressnitzer, D. (**2000**). "The lower limit of pitch as determined by rate discrimination," J. Acoust. Soc. Am. **108**, 1170–1180.

LeCun, Y., and Bengio, Y. (**1995**). "Convolutional networks for images, speech, and time-series," in *The Handbook of Brain Theory and Neural Networks*, edited by M. A. Arbib, MIT Press, Cambridge, MA.

Liu, C., and Kewley-Port, D. (**2004**). "STRAIGHT: A new speech synthesizer for vowel formant discrimination," ARLO, 31–36.
Luce, D., and Clark, M. (**1967**). "Physical correlates of brass-instrument tones," J. Acoust. Soc. Am. **42**, 1232–1243.
Mclntyre, M. E., Schumacher, R. T., and Woodhouse, J. (**1983**). "On the oscillations of musical instruments," J. Acoust. Soc. Am. **74**, 1325–1345.
Meddis, R., and Hewitt, M. J. (**1991**). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification," J. Acoust. Soc. Am. **89**, 2866–2882.
Miller, G. A. (**1947**). "Sensitivity to changes in the intensity of white noise and loudness," J. Acoust. Soc. Am. **19**, 609–619.
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.
Ottaviavi, L., and Rocchesso, D. (**2004**). "Auditory perception of 3D size: Experiments with synthetic resonators," Trans. Appl. Perceptions **1**, 118–129.
Patterson, R. D. (**1994a**). "The sound of a sinusoid: Spectral models," J. Acoust. Soc. Am. **96**, 1409–1418.
Patterson, R. D. (**1994b**). "The sound of a sinusoid: Time-interval models," J. Acoust. Soc. Am. **96**, 1419–1428.
Patterson, R. D., Allerhand, M., and Giguère, C. (**1995**). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890–1894.
Patterson, R. D., and Irino, T. (**1998**). "Modeling temporal asymmetry in the auditory system," J. Acoust. Soc. Am. **104**, 2967–2979.
Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). "Complex sounds and auditory images," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 67–83.
Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.
Reby, D., and McComb, K. (**2003**). "Anatomical constraints generate honesty: acoustic cues to age and weight in roars of red deer stags," Anim. Behav. **65**, 519–530.
Riede, T., and Fitch, W. T. (**1999**). "Vocal tract length and acoustics of vocalization in the domestic dog *Canis familiaris*," J. Exp. Biol. **202**, 2859–2867.
Smith, D. R. R., and Patterson, R. D. (**2005**). "The interaction of glottal-pulse rate and vocal tract length in judgements of speaker size, sex and age," J. Acoust. Soc. Am. **118**, 3177–3186.
Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H., and Irino, T. (**2005**). "The processing and perception of size information in speech sounds," J. Acoust. Soc. Am. **117**, 315–318.
Welling, L., and Ney, H. (**2002**). "Speaker adaptive modelling by vocal tract normalization," IEEE Trans. Speech Audio Process. **10**, 415–426.
Wolpert, D. H. (**1996a**). "The lack of *a priori* distinctions between learning algorithms," Neural Comput. **8**, 1341–1390.
Wolpert, D. H. (**1996b**). "The existence of *a priori* distinctions between learning algorithms," Neural Comput. **8**, 1391–1420.
Yost, W. A., Patterson, R. D., and Sheft, S. (**1996**). "A time-domain description for the pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **99**, 1066–1078.

Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults

Patti M. Johnstone and Ruth Y. Litovsky[a)]
Waisman Center, University of Wisconsin-Madison, Binaural Hearing & Speech Laboratory, 1500 Highland Avenue, Room 523, Madison, Wisconsin 53706

(Received 12 October 2004; revised 24 May 2006; accepted 19 June 2006)

Speech recognition in noisy environments improves when the speech signal is spatially separated from the interfering sound. This effect, known as spatial release from masking (SRM), was recently shown in young children. The present study compared SRM in children of ages 5–7 with adults for interferers introducing energetic, informational, and/or linguistic components. Three types of interferers were used: speech, reversed speech, and modulated white noise. Two female voices with different long-term spectra were also used. Speech reception thresholds (SRTs) were compared for: Quiet (target 0° front, no interferer), Front (target and interferer both 0° front), and Right (interferer 90° right, target 0° front). Children had higher SRTs and greater masking than adults. When spatial cues were not available, adults, but not children, were able to use differences in interferer type to separate the target from the interferer. Both children and adults showed SRM. Children, unlike adults, demonstrated large amounts of SRM for a time-reversed speech interferer. In conclusion, masking and SRM vary with the type of interfering sound, and this variation interacts with age; SRM may not depend on the spectral peculiarities of a particular type of voice when the target speech and interfering speech are different sex talkers. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2225416]

PACS number(s): 43.66.Lj, 43.71.Ft, 43.71.Gv, 43.66.Pn [GDK] Pages: 2177–2189

I. INTRODUCTION

In a "cocktail party" environment (Cherry, 1953), a listener must be able to follow a specific conversation and ignore all the other interfering voices and sounds. Considerable research has been done to study how "cocktail party" environments affect adult listeners (for review, see Bronkhorst, 2000), however, very few studies have examined how children perform in a comparable situation. This area of research is rather important, given that children often find themselves having to hear and learn in very challenging acoustic settings.

The ability of children to extract information from auditory signals has been compared with that of adults in other measures. Children as young as 4 years of age demonstrate adult-like performance on frequency resolution tasks (Hall and Grose, 1994) and by 5 years of age minimum audible angle thresholds for simple sound configurations are adult-like (Litovsky, 1997). Neurophysiological studies support behavioral data in that children demonstrate adult-like frequency-specific maturation of the auditory periphery within the first year of life (Eggermont *et al.*, 1996; Ponton *et al.*, 1992).

In contrast, maturation is protracted past 10 years of age, for some temporal resolution tasks, thought to reflect central auditory processes (Hall and Grose, 1994; Hartley *et al.*, 2000). Temporal resolution abilities are likely important for sound source segregation in "cocktail party" environments. Other abilities that might similarly be important are also not fully developed in young children. Examples include echo suppression (e.g., Morrongiello *et al.*, 1984; Litovsky, 1997; for a review, see Litovsky and Ashmead, 1997) and detection of movement of a fused auditory image (Cranford *et al.*, 1993). Similarly, the ability to detect signals in a fluctuating noise background is reduced in children when compared to adults (Grose *et al.*, 1993; Veloso *et al.*, 1990).

Studies on the "cocktail party" in adults, in which spectral cues vary, have shown that speech intelligibility depends on the number and type of interferers present (Culling *et al.*, 2004; Hawley *et al.*, 2004). When a single interfering sound is present, adults experience less masking in the presence of speech-based sounds such as speech and time-reversed speech than they do with modulated noise. When multiple interferers are present, however, speech reception thresholds (SRTs) with speech or time-reversed speech interferers become worse than for modulated noise. The trend toward increased masking with multiple speech interferers does not continue beyond two interferers. As the numbers of speech interferers is increased beyond two or more the difference between speech interferers and noise interferers decreases.

Another important factor is the advantage for speech intelligibility typically observed when the interfering sounds are spatially separated from the target, known as spatial release from masking (SRM) (Arbogast *et al.*, 2002; Bronkhorst and Plomp, 1988, 1992; Culling *et al.*, 2004; Dirks and Wilson, 1969; Freyman *et al.*, 1999; Hawley *et al.*, 1999; Hawley *et al.*, 2004; Peissig and Kollmeier, 1997; Plomp and Mimpen, 1981). In adults, SRM can be as large as 12 dB in the free field when multiple linguistically relevant interferers (such as speech or time-reversed speech) are present. SRM is smaller for adults in the presence of multiple

[a)]Author to whom correspondence should be addressed; electronic mail: litovsky@waisman.wisc.edu

noise-type interferers that do not contain linguistic content or context (Bronkhorst, 2000; Hawley *et al.*, 2004). This suggests that when informational masking occurs, spatial cues become particularly important for source segregation (e.g., Bronkhorst, 2000; Culling *et al.*, 2004). The role of informational masking in studies that focus on free field spatial segregation remains to be better understood. Certainly for children, this remains an unexplored area.

When considering how children negotiate their acoustic space some important factors to bear in mind include: what *types* of sounds are interfering with the signal, where those interfering sounds are positioned relative to the target source, and *how many* interfering sounds are present. While studying these questions in 4–7 years-old children and adults Litovsky (2005) found that masking is greater with interfering sounds that do not contain linguistic content (e.g., modulated noise) than for speech, but SRM is similar for the two interferer types. The modulated noise (MN) interferer differs from speech in two ways. First, although amplitude fluctuations are similar to those found in speech, the spectrum of MN does not vary in time, while speech contains both amplitude and frequency fluctuations. Second, MN has no linguistic content or context. From that work, the relative effects of variation in spectral overlap and the presence or lack of linguistic content/context were difficult to tease apart. The current study therefore extended previous work to examine masking and SRM with time-reversed speech. Reversed speech carries linguistic *context* (listeners perceive it as being speech-like) but does not carry linguistic *content* (listeners cannot understand what is being said).

Like speech, ongoing frequency variations result in variable overlap between the first and second formants in the target and interfering sounds. It is also important to consider whether the amount of masking that is produced by a specific voice influences the size of SRM. Specifically, female voices can be quite variable with regard to the spectra (Stelmachowicz *et al.*, 1993; Hazan and Markahm, 2004), such that some can produce significantly more masking than others.

In the present study, three issues were examined. First, in one set of subjects three types of interfering sounds (modulated noise, speech, and time-reversed speech) were used in order to understand how stimuli that have linguistic content and/or context, or neither, affect masking and SRM in children and adults. Second, in two new groups of subjects the issue of female voice type was teased apart by comparing masking and SRM for two female voices. Both voices are pleasant to the ear and intelligible; while one has reduced energy at 1–3 kHz (female A), the other has a more "average" spectral shape and hence greater overlap in energy with the target (female B). Third, the effect of task difficulty on masking and SRM in adult listeners was examined by comparing the performance of the first group of adults using a 25-AFC task with published data from a group of adults using a 4-AFC task.

II. METHODS

A. Listeners

Twenty volunteer children in kindergarten and first grade were recruited from the Madison area public schools. Twenty adult volunteers were recruited from the University of Wisconsin-Madison student and staff population. All of the listeners were native speakers of English and had normal hearing sensitivity, as indicated by pure-tone, air-conduction thresholds of 20 dB HL or less (ANSI, 1989) at octave frequencies between 250 and 8000 Hz. No asymmetry in hearing exceeded 10 dB HL between the two ears. Since middle ear problems are common in young children, tympanometry was performed before each visit using a screening tympanometer calibrated to ANSI specifications (ANSI, 1987). Children were included in the study only when their peak-compensated static admittance was normal.

The children comprised two groups with ten subjects in each group: Group 1A (ages 5.0—6.11) were tested using three different interfering stimuli (MN, female-A speech, female-A reversed speech), and completed testing in two 1 -h sessions. Group 1B (ages 6.0—7.0) were tested using only the female-B speech interferer during a single session. The adults also comprised two groups with ten subjects in each group: Groups 2A (ages 18—42) and 2B (ages 18—30) were tested on parallel conditions to those of the corresponding children groups, and each subject completed testing in a single session. The children groups did not entirely overlap in age, but this is not a concern since Litovsky (2005) showed that masking and SRM do not vary under these conditions between 4.5 and 7 years of age.

B. Stimuli

The target stimuli consisted of a closed set of twenty-five, two-syllable children's spondees recorded with a male voice and obtained from Auditec. The root-mean-square values of all words were equalized.

Four types of interfering stimuli were used. (1) forward speech spoken by *female A* with relatively little energy at high frequencies; (2) modulated white noise (MN); (3) time-reversed speech spoken by *female A*; (4) forward speech spoken by *female B* with greater energy at high frequencies, resembling the "average" spectra from previously published reports (e.g., Stelmachowicz *et al.*, 1993). Figure 1 shows the long-term spectra of the two female voices and the MN, each beside the target voice; these were obtained by taking the FFT of the signals in 1/3 octave bands.

The content of the speech interferers for both female voices consisted of digitized sentences from the Harvard IEEE list (Rothauser *et al.*, 1969). Examples of sentences include: "Tea served from the brown jug is tasty" or "A dash of pepper spoils beef stew." To generate the MN interferer, the speech envelope was extracted from speech interferers created using the female-A voice and was used to modulate white noise tokens, giving the same coarse temporal structure as the speech. The envelope of running speech was extracted using a method similar to that described by Festen and Plomp (1990), in which a rectified version of the wave form is low-pass filtered. A first-order Butterworth low-pass filter was used with a 3-dB cut-off at 40 Hz. The time-reversed interferers consisted of the speech interferers, but simply reversed in time end to end (e.g., Hawley *et al.*, 2004). Though they shared the same temporal-spectral struc-

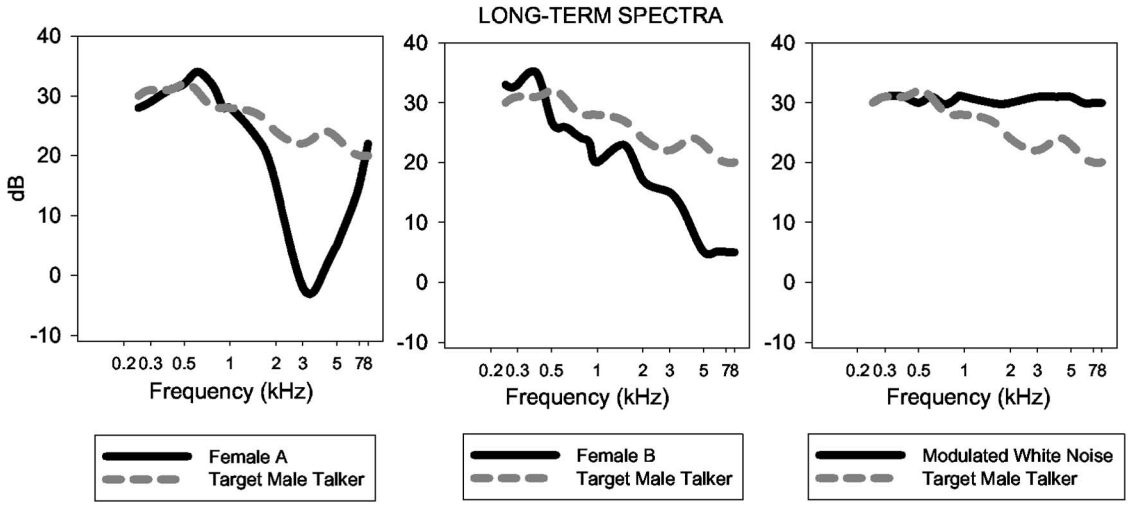


FIG. 1. The long-term speech spectra of two female talkers and MN are each shown beside the target male talker stimulus.

ture of the speech interferers using the voice of female A, the reversed speech interferers were unintelligible.

Testing was conducted in a standard IAC sound-proof booth with an inside dimension of 2.75 m × 3.25 m. Targets and interferers were presented via two separate loudspeakers (Cambridge Soundworks, Center/Surround IV) positioned at a distance of 1.5 m from the listener's head.

C. Design and procedure

The three basic conditions were: Quiet (target 0° front, no interferer), Front (target and interferer both 0° front), and Right (interferer 90° right, target 0° front). In group 1A, children visited the lab twice. During their first visit testing included three conditions: Quiet, reversed speech-Front and reversed speech -Right. During the second visit testing included five conditions: Quiet, MN-Front, MN-Right, Speech-Front, Speech-Right. Children in group 1B were tested on three conditions: Quiet, Speech-Front, Speech-Right. During each session, the order of conditions was randomized for each child, and testing was conducted once per condition. Adults in group 2A completed two repetitions of each condition within a single visit, and adults in group 2B were tested once on each condition; order of presentation was always randomized.

Prior to testing, each child participated in a brief familiarization task to determine if s/he could readily identify every target word. The experiment was designed to measure the effect of speech and noise interferers on word recognition, rather than vocabulary. During the familiarization task the child was asked to identify the pictured spondees. For each participant that was recruited, within minutes, it was clear that the child was comfortable and familiar with every target word that could be used during the experiment.

All listeners sat at a small table facing a computer screen that was positioned under the loudspeaker at 0°.

In conditions containing interfering sounds, the interferer was turned on first followed by the target presentation, and continued after the target was turned off for approximately 1–2 s. Subjects were instructed to ignore the female voice and to listen carefully to the male voice.

The task for children consisted of a 4-alternative-forced-choice procedure (Litovsky, 2004; 2005). On every trial a word from the Children's Spondee List, spoken by a male talker, was chosen randomly from a closed set of 25 targets. The randomization process ensured that for every subject, on average, all 25 words were selected an equal number of times. The target word was preceded by a leading phrase "Ready? Point to the…." also spoken by a male talker. Each child was then asked to identify a picture matching that word from an array of four pictures that appeared together on the computer screen, only one of which matched the target. Correct responses were followed by positive feedback (animation); incorrect responses were followed by "negative" feedback from the computer such as "let's try another one" or "that must have been difficult."

The task for adults in group 2A consisted of a 25-alternative-forced-choice (25-AFC) procedure. Subjects used the computer mouse to select the written word from a list of 25 words that appeared on the computer screen, only one of which matched the target. No feedback regarding performance was provided. The 25-AFC task in the current experiment was invoked as a means of equating task difficulty as much as possible between adults and children. In the previous study, adults tested on the 4-AFC task (Litovsky, 2005) had very low SRTs compared with children. In addition, there was little or no SRM on some conditions, suggesting that the ease with which the task was performed may have eliminated some of the most interesting effects. In the present study, an effort was made to balance age-dependent difficulty of the task by increasing the number of alternatives in the forced choice paradigm.

D. SRT estimation

At the start of each condition measurement, the level of the target was initially 60 dB SPL. When interferers were present (Front or Right conditions), the interferer level was fixed at 60 dB SPL. SRTs were estimated using a method described by Litovsky (2004, 2005). An adaptive tracking method was used to vary the level of the target signal, such that correct responses result in level decrement and incorrect

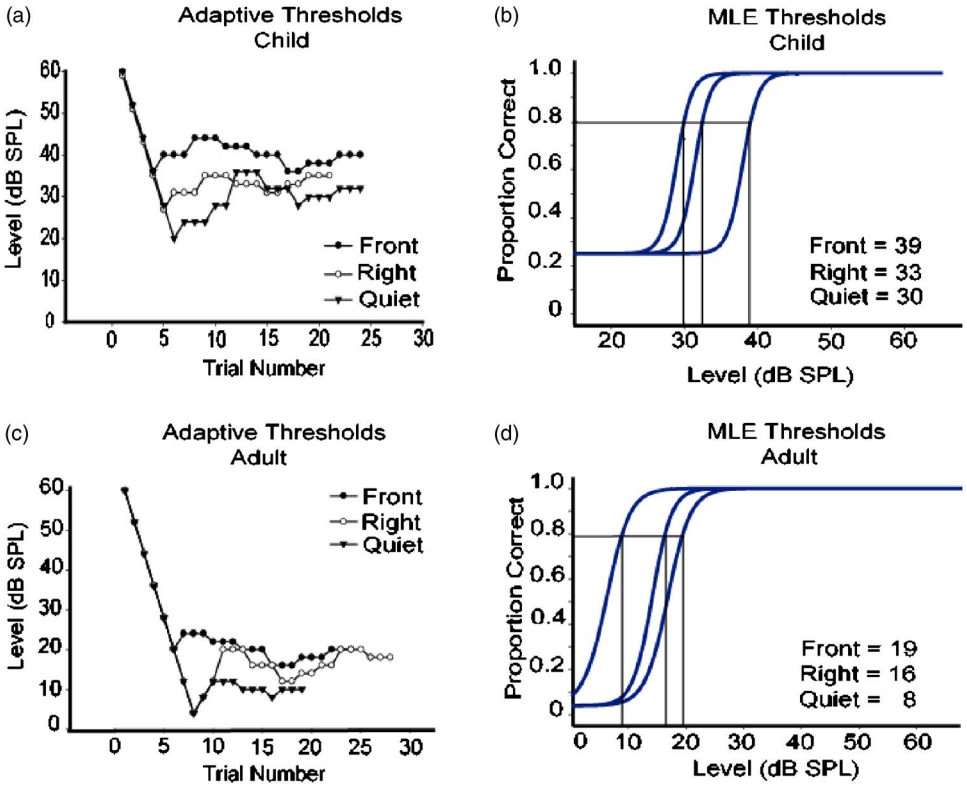


FIG. 2. Examples of adaptive tracts (3-down-1-up) for one child (A) and one adult (C) are shown. Interferer level remained fixed at 60 dB SPL. A sigmoidal function [(B), (D)] was fit to the raw data from each adaptive track, and a maximum likelihood estimate (MLE) procedure was used to estimate threshold at a performance level of 80% correct. The legend reports the associated estimates.

responses result in level increment. The algorithm includes the following rules: (1) Level is initially reduced in steps of 8 dB, until the first incorrect response. (2) Following the first incorrect response a 3-down/1-up rule is used, whereby level is decremented following three consecutive correct responses and level is incremented following a single incorrect response. (3) Following each reversal the step size is halved. (4) The minimum step size is 2 dB. (5) A step size that has been used twice in a row in the same direction is doubled. For instance if the level was decreased from 40 to 36 (step =4) and then again from 36 to 32 (step=4), continued decrease in level would result in the next level being 24 (step =8). (6) After three consecutive incorrect responses a "probe" trial is presented at the original level of 60 dB. If the probe results in a correct response the algorithm resumes at the last trial before the probe was presented. If more than three consecutive probes are required, testing is terminated and the subject's data are not included in the final sample. (7) Testing is terminated following four reversals [e.g., Figs. 2(a) and 2(c)].

SRTs were estimated using a constrained maximum-likelihood method of parameter estimation (MLE) outlined by Wichmann and Hill (2001a; 2001b). All the data from each adaptive track were fit to a logistic function and the inverse of the function at a specific probability level was taken. Slopes were calculated by taking the derivative of the function with respect to threshold. Psychometric functions for the children's data, which were collected with a 4-AFC task, were set to a lower bound level of 0.25, which was the level of chance performance. Give that an adaptive 3-down/1-up procedure was used, threshold corresponded to the stimulus value point on the psychometric function where performance was approximately 79.4% correct (as estimated by Levitt, 1971).

It is important to note that biased estimates of threshold can be introduced by the sampling scheme used and lapses in listener attention. The upper bound of the psychometric function was constrained within a narrow range (0.05), as suggested by Wichmann and Hill (2001b), who demonstrated that bias associated with attention lapses was overcome by introducing a highly constrained parameter to control the upper bound of the psychometric function. As the authors suggest, under some circumstances, bias introduced by sampling scheme may be more problematic to avoid even when a hundred trials are obtained per level visited.

SRTs obtained using MLE were compared with SRTs calculated from the last three reversals in each experimental run. A repeated measures t-test revealed no statistical difference between the two threshold estimates [$t(179)=1.832$, $p=0.0686$), two tailed]. Although the MLE procedure produces similar group mean thresholds, its advantage is that group variances are typically smaller. The MLE approach offers an advantage especially when few reversals can be obtained, such as when working with young children, in which case a single trial can have a disproportionate weight. The MLE approach reduces the effect of individual trials by placing greater weight on levels at which the total number of trials is largest.

E. Data analysis

Average SRTs and standard deviations for children and adults for all conditions (Quiet, Front, Right) and interferer types were computed. Since the children in group 1A visited

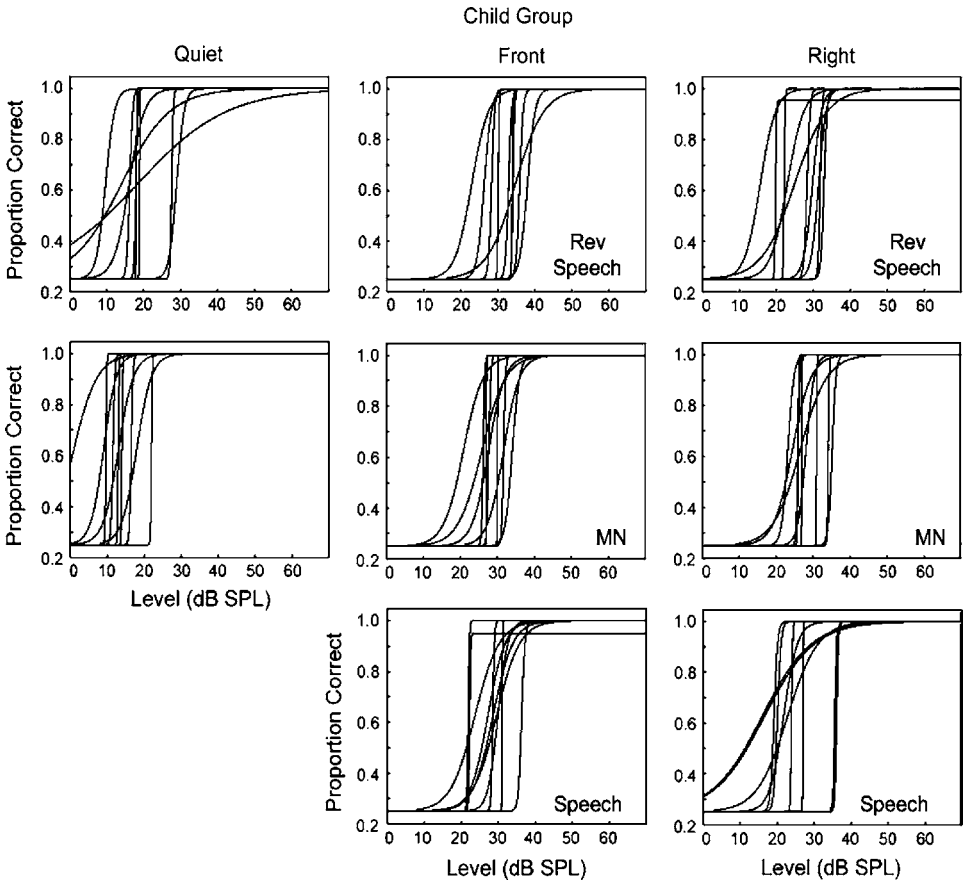


FIG. 3. Sigmoidal functions are shown for each individual measurement in the child group. Each panel shows functions for one condition (Quiet, Front, Right) and one interferer type.

the lab twice, the possibility that a learning effect occurred was considered upon comparison of thresholds obtained during the two visits. It appeared that the children's SRTs were generally higher for all conditions tested on the first visit, including Quiet, compared with those tested on the second visit. Nevertheless, the two Quiet thresholds were not statistically different ($p > 0.05$).

The higher thresholds obtained during the first visit would not have posed a problem if the type of interferer encountered during the first visit was randomized among the children in group 1A. This approach was not taken because the study was initially designed for the purpose of measuring performance with the reversed speech only. When the results appeared to be quite different between the child and adult groups a post-hoc decision was made to retest the same children using the speech and MN interferers. The children returned to the lab for a second visit to collect the speech and MN interferer data. This same step was not performed for adults in group 2A, because they were tested on all conditions during the initial visit.

In order to reduce the potential biasing effect that learning in a test-retest situation might create, all thresholds from the second visit were normalized by subtracting the average difference between the two visits for the Quiet condition (3.7 dB). It is important to note that, although the reduction in the Quiet condition may overestimate changes due to retesting that might occur in the Front and Right conditions, it was the only condition that was retested during the second visit and could be directly compared for threshold changes. While future studies should examine test-retest issues across conditions more systematically, for the purpose of the present study, the decision to normalize using Quiet thresholds is the most conservative approach, effectively reducing the age difference in the reversed speech conditions. All statistical analyses were performed using the normalized data.

III. RESULTS

Results are first discussed for groups 1A and 2A, who were tested on three types of interfering sounds (MN, speech, and time-reversed speech). Second, results from the speech conditions are compared for the groups that were tested with two different female voices (female A, female B). Finally, results with adult subjects (group 2A) using the 25-AFC procedure are compared to published data from adult subjects using the 4-AFC procedure (Litovsky, 2005). The voice of female A was used as the speech interferer for both groups.

A. Estimating SRTs and goodness of fit

Because results presented here essentially consist of SRTs that are based on estimates of psychometric functions, individual fits to the data are shown in Figs. 3 and 4, for groups 1A and 2A. Quiet psychometric functions are generally shifted toward lower signal levels relative to data collected during the Front and Right conditions. Also noticeable is that the functions in adult subjects (Fig. 4) ascend at lower levels and appear to be less variable within each condition than the children's functions (Fig. 3). The psychometric functions also clearly vary in steepness. It is difficult to know

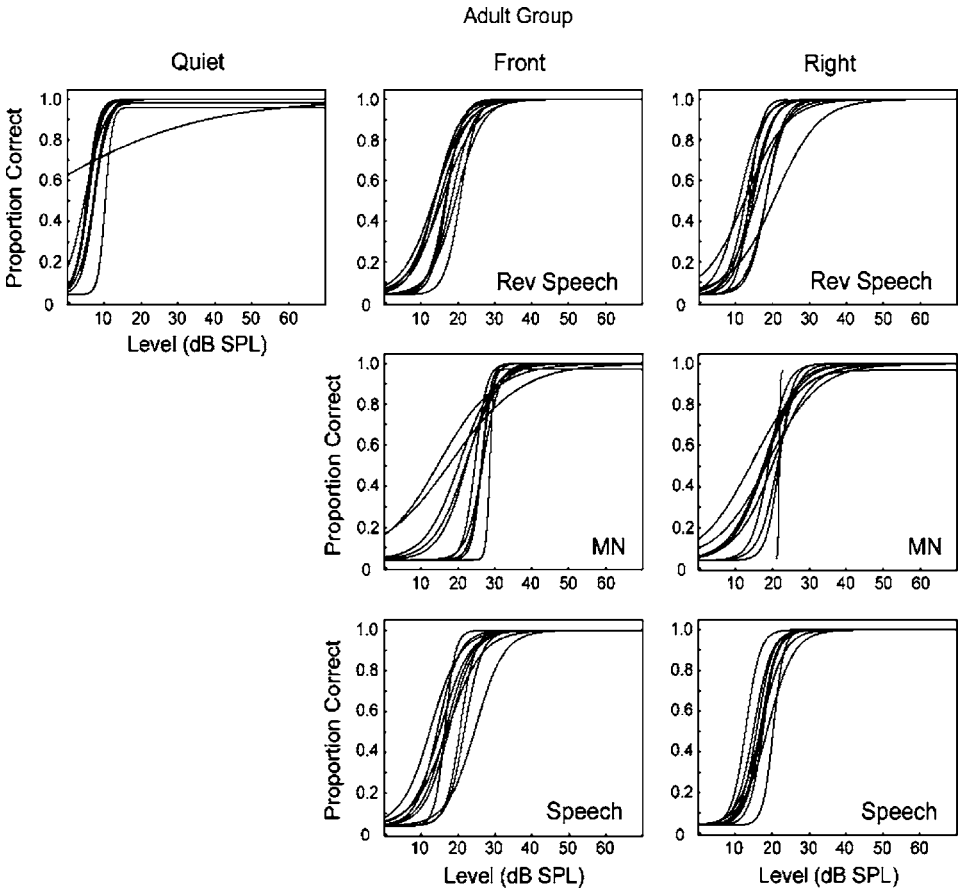


FIG. 4. Sigmoidal functions are shown for each individual measurement in the adult group. Each panel shows functions for one condition (Quiet, Front, Right) and one interferer type.

if the slope of the psychometric function is a true reflection of a sensory process or is due to averaging high variability in the underlying track (Leek *et al.*, 1991b). Individual tracks that "wandered" did result in shallow functions for both adults and children. The high variance occasionally seen in individual tracks obtained in this study did result in shallow psychometric functions as would be expected. Allen and Wightman (1995) attributed shallow psychometric functions obtained in their study to inattention. It is possible that subjects in our study experienced lapses in attention and that contributed to reduction in slopes of some of the psychometric functions.

The unusually steep functions obtained in this study were associated with dramatic changes in performance over a very small change in level. This pattern was seen more frequently with the children, whose performance often went from 100% correct to 0% correct over a very small decrement in level. It is possible that over a small change in level the target went from being unintelligible to being clearly intelligible. It is also possible that children are not particularly good at guessing the correct response when only a portion of the target word is heard and understood.

As is often the case with populations of subjects whose performance can vary dramatically, an objective analysis of the goodness of fit of the psychometric functions is appropriate. This was determined using deviance as described by Wichmann and Hill (2001a, 2001b) and as applied to informational masking data by Lutfi *et al.*, (2003). [1] Slightly less than 9% of the fitted functions were estimated to have a likelihood of less than 5%. Only three of the psychometric functions had estimated parameters that identified them as clear outliers. Two of these outliers were in the Quiet condition (one child and one adult) and one was for a child in the Right, MN condition.

In addition to measuring the deviance for each psychometric function, MLEs of confidence intervals were obtained for each threshold estimate following the procedure outlined by Lutfi *et al.*, (2003), and using the method devised by Wichmann and Hill (2001a, 2001b). A bootstrapping technique was used to estimate the sampling distributions of each threshold and determine confidence limits for the estimated distributions. Using proportion correct (assuming a binomial distribution) at each signal level from each fitted function, a simulated proportion correct at each signal level was randomly drawn and a logistic was fit to these values. This procedure was repeated 10 000 times to provide 10 000 estimates of the logistic function. Confidence limits were set at 0.025 and 0.975 so the points from the sampling distribution of the thresholds provided a 95% confidence interval. Individual thresholds fell within the confidence interval.

B. Effect of age on SRTs

In general, children had higher SRTs than adults. These results are consistent with previous research using the same methods and procedures (Litovsky, 2005). Figure 5 shows average SRTs (±s.d.) for children and adults for the various conditions and interferer types. For both children and adults, SRTs were higher in the presence of a single interfering sound source compared with the Quiet condition. Overall the

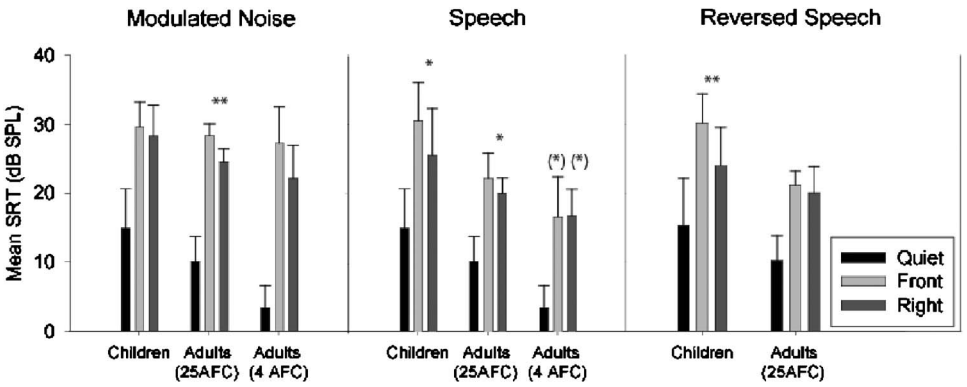


FIG. 5. Mean (±s.d.) SRTs are shown for children and adults, for the different types of interferers and conditions. Asterisks indicate significant difference between group means in the Front and Right conditions: $^{*}p<0.05$, $^{**}p<0.01$. Tables I and II identify the other significant differences. Data using a 4-AFC procedure from Litovsky (2005) are also included for adults for MN and speech interferers. Asterisks in parentheses indicate significant differences between the 4-AFC procedure and the 25-AFC procedure.

increase was 6–15 dB for children and 10–17 dB for adults. A mixed, nested analysis of variance (ANOVA) was performed to test for age differences in SRTs. Two within-subjects variables [interferer-type (MN, speech, reversed speech) and condition (Quiet, Front, Right)] and a single between-subjects variable [age (child, adult)] were used. Results revealed no significant main effects, but a significant three-way interaction ($F[4,72]=3.097$, $p<0.01$).

Tables I and II list significant differences obtained with post hoc Scheffé tests for group SRTs. Children had significantly higher SRTs than adults when the interferer was speech ($p<0.01$) or reversed speech ($p<0.001$) in the Front condition (at the same location as the target). In the Right condition, children had significantly higher SRTs than adults for all interferer types ($p<0.05$). In the Quiet condition, there was no difference between SRTs of the two age groups.

Post hoc Scheffé tests of adult data also revealed that in the Front condition SRTs were higher (greater masking) for the MN interferer compared with speech ($p<0.001$) or reversed speech ($p<0.0001$). This pattern of results was also found in the Right condition ($p<0.0005$ and <0.0001 for speech and reversed speech, respectively). SRTs did not differ between the two speech-type interferers. Finally, SRTs were significantly higher in the Front than Right conditions, for the MN ($p<0.01$) and speech ($p<0.05$) interferers, but not for reversed speech. Children's performance differed from adults in the Front condition; a lack of significant post hoc effects suggests similar amounts of masking for all interferer types. In contrast, in the Right condition children's performance mirrored that of adults in that SRTs were significantly higher with the MN than with the speech ($p<0.01$) or reversed speech ($p<0.0001$) interferers.

One of the hallmarks of auditory development is individual variability. While some children reach adult-like performance at relatively young ages, others do not. In addition, since the task focuses on listening to speech in the presence of maskers, the actual signal-to-noise ratio (SNR) at which SRTs are achieved is an important feature of the data that can then be used to compare with other masking studies. Therefore, Fig. 6 shows individual results on all conditions, as a function of the SNR at which SRTs were obtained. The extremely low SNR values (−30 to −40) reported here are likely due to several factors: the type of target used (simple spondee), the number of interferers (a single one), the difference in F0 between the target (male voice) and interferer (female voice), and the spectral characteristics of the interferer (female voice) used in our experiments. These findings are consistent with results obtained in a previous study using a similar test procedure (Litovsky, 2005).

TABLE I. Comparisons made between differences in mean SRTs for groups 1A and 2A and the level of significance obtained using post hoc Scheffé test.

Condition	Interferer type	SRT comparison	Significance
Front	MN	Children > Adults	$p<0.05$
Front	Speech	Children > Adults	$p<0.01$
Front	Reversed speech	Children > Adults	$p<0.001$
Right	MN	Children vs Adults	Not significant
Right	Speech	Children > Adults	$p<0.05$
Right	Reversed speech	Children > Adults	$p<0.05$
Quiet	No interferer	Children vs Adults	Not significant

C. Age effects and the ability to utilize interferer-dependent cues

Masking was quantified from the measured SRTs in two ways: Quiet SRTs were subtracted from Front (F-Q) or Right (R-Q) SRTs, resulting in values corresponding to "Front Masking" and "Right Masking." Masking was evaluated using a mixed ANOVA with two within-subjects variables [interferer-type (MN, speech, reversed speech) and masking condition (F-Q, R-Q)] and one between-subjects variable [age (child, adult)].

TABLE II. Comparisons made between differences in mean SRTs for the different interferer types and the level of significance obtained using post hoc Scheffé test.

Condition	Group	SRT comparison	Significance
Front	Adults	MN > Speech	$p<0.001$
Front	Adults	MN > Rev Sp	$p<0.0001$
Front	Adults	Speech vs Rev Sp	Not significant
Right	Adults	MN > Speech	$p<0.0005$
Right	Adults	MN > Rev Sp	$p<0.0001$
Right	Adults	Speech vs Rev Sp	Not significant
Front	Children	MN vs Speech	Not significant
Front	Children	MN vs Rev Sp	Not significant
Front	Children	Speech vs Rev Sp	Not significant
Right	Children	MN > Speech	$p<0.01$
Right	Children	MN > Rev Sp	$p<0.0001$
Right	Children	Speech vs Rev Sp	Not significant

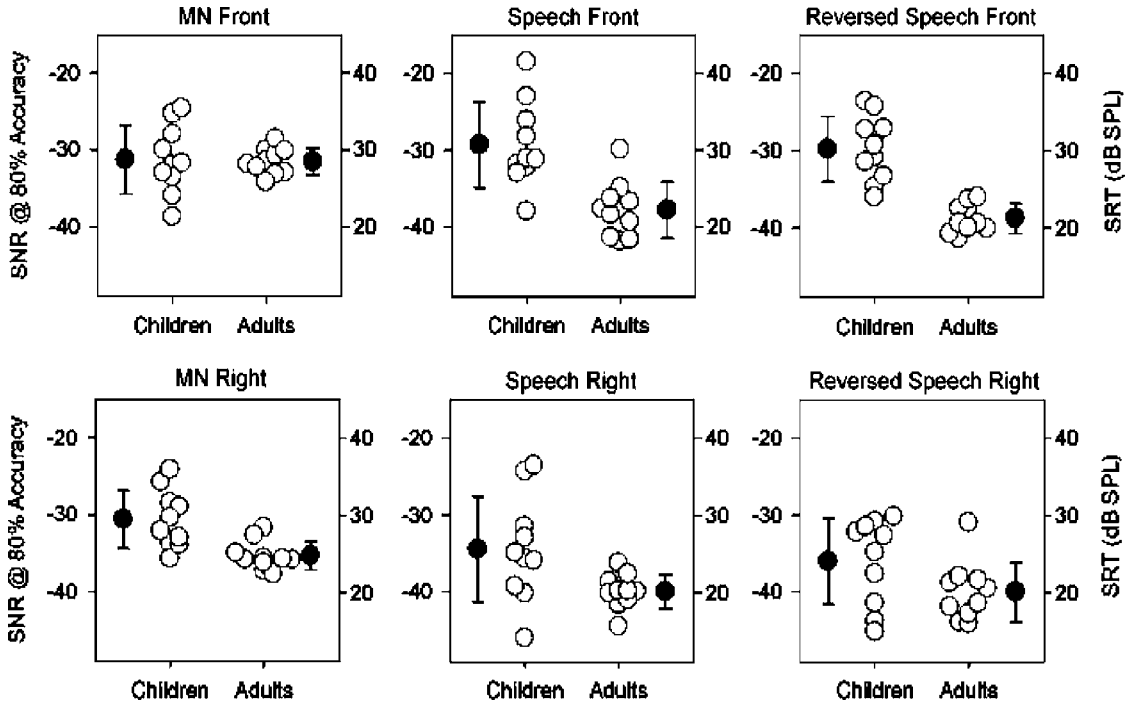


FIG. 6. Performance of children and adults is compared for the three types of interferers. Data are plotted such that either SNR at which performance reached 80% correct (left axis value) or SRT (right axis value) can be seen. Individual data (white symbols) and group means ±s.d. (black symbols) are shown.

No significant main effects for masking were found. However, there was a significant 3-way interaction ($F[2,36]=5.645$, $p<0.01$). Post hoc Scheffé tests revealed that, in the adult group F-Q masking was significantly larger with the MN than speech ($p<0.0001$) or reversed speech ($p<0.0001$). Children, however, did not show any significant differences in the amount of F-Q masking across interferer types. Interesting age effects were found for F-Q masking. Children experienced significantly greater masking than adults with reversed speech ($p<0.01$), but significantly less masking than adults with MN; there were no age effects in Front masking for the speech interferer.

In the Right conditions, both children and adult groups showed more masking with MN than speech ($p<0.01$) or reversed speech ($p<0.01$). Children and adults did not differ significantly on the amounts of R-Q masking for any interferer type.

D. Spatial release from masking (SRM)

Spatial release from masking (SRM) is defined as the difference in SRT between the Front and Right conditions (F-R). A mixed ANOVA was applied to SRM values treating interferer type (MN, speech, reversed speech) as the within-subjects variable and age (child, adult) as the between-subjects variable. There were no significant main effects, but a significant two-way interaction was found ($F[2,36]=21.369$, $p<0.0005$). Post hoc Scheffé tests revealed that SRM was significantly *larger* in children than adults with reversed-speech ($p<0.01$), but significantly *smaller* in children than adults with MN ($p<0.05$). In addition, the amount of SRM for children was significantly less with MN compared with reversed speech ($p<0.01$) or speech ($p<0.05$), and there was no difference between speech and reversed speech. In adults, SRM was similar for all three interferer types. These results suggest that SRM is interferer-type dependent for children but not for adults. In addition, SRM in children is best measured with a speech-type interferer rather than a noise-type interferer.

SRM is a measure that varies greatly across individuals. This can be seen in Fig. 7, where SRM becomes visible when R-Q values are plotted as a function of F-Q values for each subject, by condition. The line bisecting each panel represents unity, such that symbols falling along the diagonal indicate absence of SRM, while symbols appearing below the line indicate that SRM occurred. It is clear that although SRM occurred for the majority of subjects, especially in the speech-type conditions, there was great variability across individuals, and that the adult data are generally clustered more tightly than the children's data.

E. Effect of type of female voice

In an effort to understand how masking and SRM could be affected by variability in the amount of high-frequency energy present in female talkers, two female voices with very different voice spectra were used in the present study on some of the measures. Results were compared for the two groups of children (1A and 1B) and two groups of adults (2A and 2B). Plotted in Fig. 8(a) are values for F-Q and R-Q masking obtained with the two different female talkers, for children and adults. A mixed ANOVA was performed with one within-subjects variable [masking condition (F-Q, R-Q)] and two between-subjects variables [age (Child, Adult) and talker (Female A, Female B)]. Significant main effects were found for both masking condition [$F(1,36)=14.918$, $p<0.0001$] and talker [$F(1,36)=65.890$, $p<0.00001$], with no significant interactions. These results show that amount of

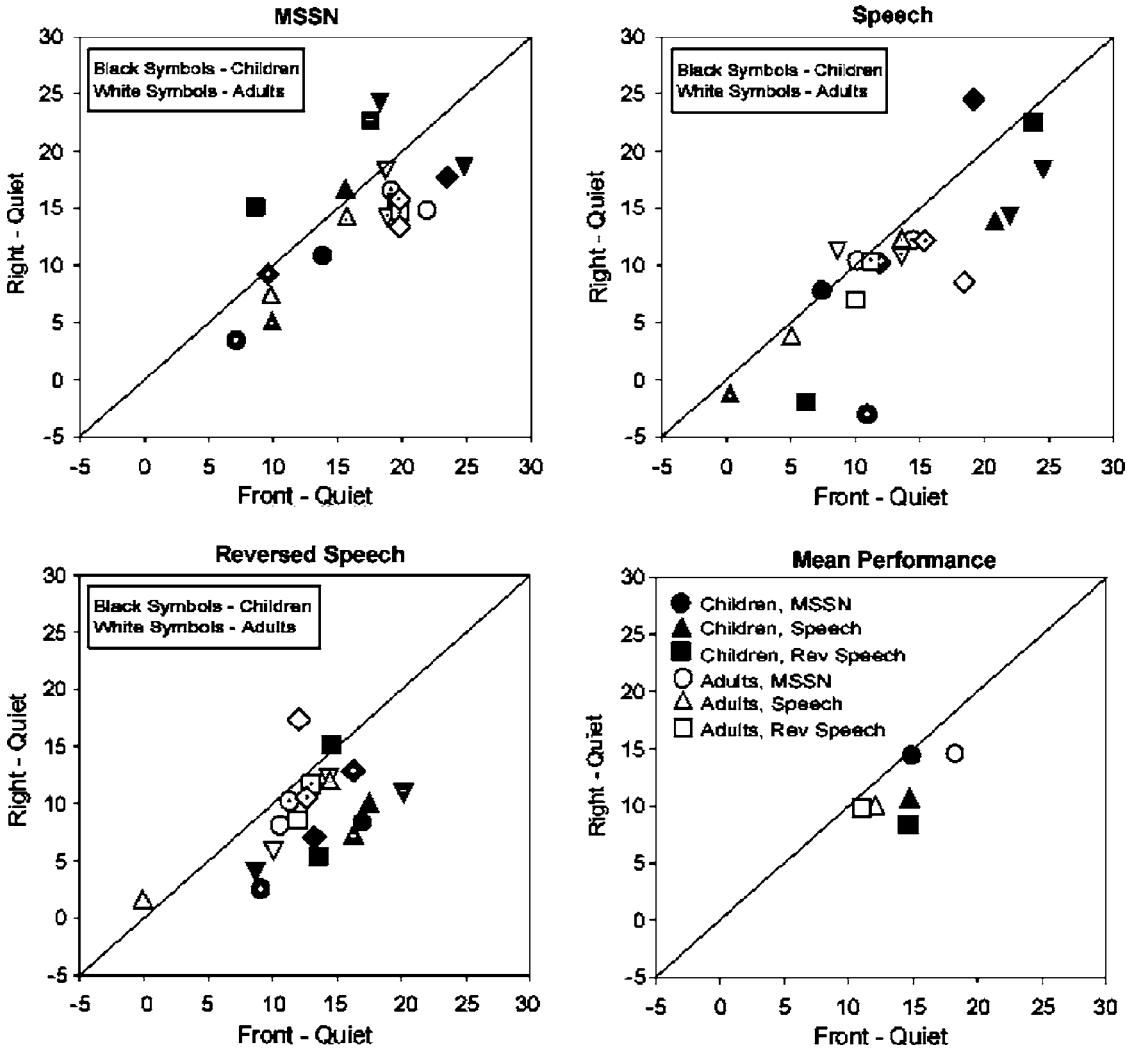


FIG. 7. Individual data points for children (black symbols) and adults (white symbols) are plotted for the R-Q condition as a function of values obtained in the F-Q condition. Each plot shows values for a different type of interferer. The plot on the bottom-right shows average values for these conditions, for the two age groups.

masking was higher for Female B than Female A, for both adults and children. In addition, masking was greater in the F-Q than R-Q condition, for both age groups, and both voices. These results are not surprising given that the voice of female B contains more spectral energy in the high frequencies. It is interesting to note that F-Q and R-Q masking increased equally, by about 11 to 12 dB, for both children and adults. The equal increase in masking in the Front and Right conditions [see Fig. 8(b)] suggests that the increase may be due to energetic masking.

SRM was evaluated with a two-way between-subjects ANOVA [age (Child, Adult); and talker (Female A, Female B)]. No significant main effects or interactions were found. The amount of SRM obtained using the two different female voices did not differ for either age group (see Fig. 9), suggesting that the effect being explored in this study remains constant regardless of the two types of voices used here.

F. Task difficulty in the forced choice paradigm for adults

The use of the 25-AFC task in the current experiment was a deliberate attempt to make the task more difficult for the adults compared with the 4-AFC task. It was hypothesized that, by increasing the level of difficulty for the adults, SRTs would increase. Independent-samples *t*-tests were used to compare the adult data from the 4-AFC task (Front and Right, MN and speech interferers; Litovsky, 2005) with adult data obtained here using the 25-AFC. No statistically significant difference for the MN interferer was found in the Front or Right ($p > 0.05$). SRTs were significantly higher for the speech interferer with the 25-AFC task than the 4-AFC task in both Front [$t(18)$ one tailed$=-2.586$, $p < 0.05$] and Right [$t(18)$ one tailed$=-2.280$, $p < 0.05$]. Task difficulty seems to increase masking for a speech interferer, but not for a noise interferer. There was no effect of task difficulty on SRM, suggesting that with either method the adult and child data are comparable.

IV. DISCUSSION

This study investigated the ability of young children and adults to understand speech in the presence of three different types of interferers (MN, speech, and time-reversed speech), for two different female voices, and the extent to which lis-

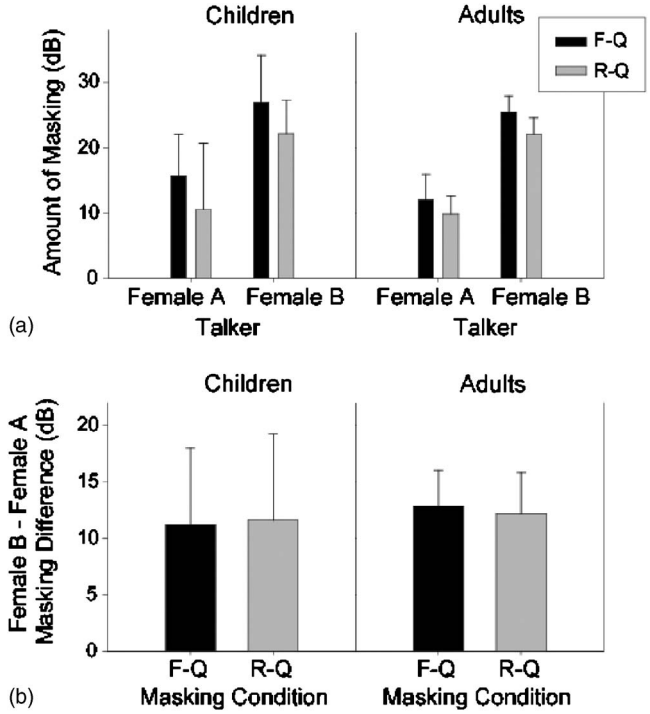


FIG. 8. (A) Mean (±s.d.) amount of masking in the F-Q and R-Q conditions, shown for children and adults. (B) Mean (±s.d.) increase in masking (F-Q and R-Q) for the Female B interferer compared with Female A. Values are shown for children and adults in dB.

teners benefit from spatial separation of the target and interferer. Effects of noise-based, language content-based and language-context based interference were compared. A child-friendly task was employed in which children could perform well enough to evaluate their performance under conditions that vary in difficulty. This approach requires some thought with regard to previous work conducted with adults, in which typically harder, open-set sentence materials are used as target stimuli.

A. Age effect on speech reception threshold

Results suggest that, compared with the Quiet condition, adding a single interferer leads to decreased speech intelligibility, as seen by elevated SRTs for both children and adults. Children's SRTs tended to exceed those of adults regardless of the type of interferer, location of the interferer, or level of task difficulty (4-AFC task for children versus 25-AFC task for adults), see Fig. 4; this finding supports previous reports on speech recognition performance in adults and children in the presence of speech interferers (Litovsky, 2005; Papso and Blood, 1989) and speech-shaped noise (Hall *et al.*, 2002). These results can be explained by monaural masking differences, since young children usually experience higher thresholds than adults in the context of complex listening tasks (e.g., Grose *et al.*, 1993; Veloso *et al.*, 1990).

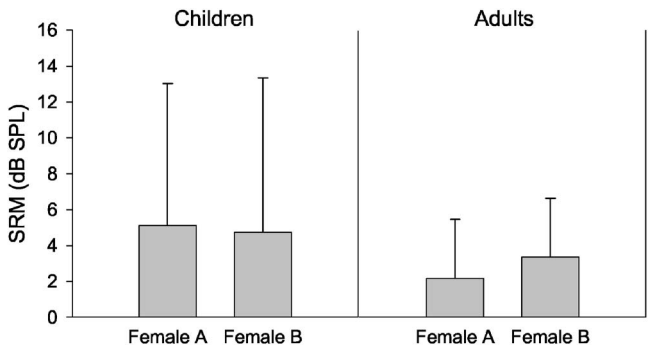


FIG. 9. Mean (±s.d.) amount of SRM for children and adults obtained using two different female voices.

B. Interferer-dependent masking

The results suggest that the amount of masking varied by location, interferer type, and age. In adults, the MN interferer produced more masking than either speech or reversed speech. Effect of interferer type on masking measured here for adults is consistent with previous work of Hawley *et al.* (2004) in which masking with a single interferer was greater with MN than speech or reversed speech; the authors attributed this difference to an energetic masking effect. In the presence of a single speech or reversed-speech competing signal, listeners can take advantage of both the temporally modulated amplitude gaps and the spectral variation to hear out the target speech, but the lack of spectral modulation in MN results in greater energetic masking and greater overall SRTs.

Energetic masking, which is thought to occur when the signal is degraded by energetic spectral overlap from the interfering sounds, is presumed to occur in the peripheral auditory system (e.g., von Békésey, 1960; Green and Swets, 1974). This contrasts with informational masking which occurs when the signal is "lost" in a background of interference that shares similar, often nonoverlapping acoustic patterns with the signal (Watson *et al.*, 1976; Lutfi, 1990; Leek *et al.*, 1991a, b; Brungart, 2001; Kidd *et al.*, 2002; Arbogast *et al.*, 2002; Durlach *et al.*, 2003). Informational masking is thought to result from processes within the central auditory system that cannot be accounted for simply by peripheral mechanisms. The speech and reversed-speech interferers are good examples of maskers that produce a combination of energetic and informational masking, although the relative contributions of each effect are difficult to tease apart.

Masking also varied by location and type, especially in children: MN produced higher SRTs than speech or reversed speech when the target and interferers were spatially separated, but there was no effect of interferer type when the target and interferer were spatially coincident. In other words, in absence of spatial cues all interferers were equally difficult to ignore, but when spatial cues were available they were more useful with speech-based interferers than the noise-based interferer. Similar interactions between interferer type and location have been reported when spatial separation was induced using the precedence effect paradigm (Freyman *et al.*, 2001).

C. Age- and task-dependent masking

One explanation for the age difference in amount of Front masking produced by different interferers is that it may reflect maturation of frequency resolution. That is, children may be less able to use differences in the F0 between the

target and maskers in the speech and reversed-speech conditions, an ability that adults are known to be very good at when a single interferer is used (Festen and Plomp, 1990; Brungart *et al.*, 2001; Hawley *et al.*, 2004). This explanation is not compelling because basic abilities such as frequency resolution and tone detection in noise are adult-like by age 4 (e.g., Hall and Grose, 1994). More likely is the explanation that speech intelligibility in the presence of any masker, especially in absence of spatial cues, is more challenging for children than adults due to immature central auditory processes that are required for source segregation.

Our findings are consistent with reports on developmental changes in auditory attention in children (for reviews see Gomes *et al.*, 2000; Plude *et al.*, 1994). Younger children are more likely to process both relevant and irrelevant stimulus streams (Berman and Freidman, 1995; Doyle, 1973; Macoby, 1969), and seem to have difficulty selecting the appropriate stimulus channel and maintaining attentional focus over time (Berman and Friedman, 1995; and Macoby, 1969). In addition, older children (and adults) seem better able to engage in tasks and to employ useful strategies to solve problems that require selective auditory attention (Guttentag and Ornstein, 1990). It is possible that the underlying inefficient allocation of auditory attention often seen in younger children is directed not only by the physical characteristics of the stimuli but by the interests, motives, and cognitive strategies of the child perceiving them (Gomes *et al.*, 2000). This could likely affect performance during the most challenging conditions in the present study, such as when the target and interferer arise from similar locations.

Finally, it is important to note that the SNR at which SRTs are estimated were fairly low in the present study compared with related work. For instance, here the SNR for 80% accuracy was between −30 and −40 dB as compared to −10 dB reported by Hawley *et al.*, (2004). This can be explained by considering several methodological differences between the two studies. Here we used target and interferers that differed in gender (and fundamental frequency), while the other study used the same talker for both target and interferer. Second, we used a small corpus of targets in a closed-set forced-choice paradigm, whereas in the other study open-set sentences were employed. Third, we only had the single-interferer condition compared with single-, double-, and triple-interferer conditions in the Hawley *et al.* study. Additional interferers would have likely added more masking and reduced the threshold SNR. These issues are important to consider in future developmental studies. Finally, the spectral characteristics of the interferer (female voice) used in our experiments reduced the amount of masking experienced by listeners.

D. Spatial release from masking (SRM)

Results in the current study suggest that both children and adults experience SRM. Adults showed slightly larger SRM here than in the Litovsky (2005) study, likely due to the increased difficulty of the 25-AFC task compared with the 4-AFC task in the previous study. In general though, the average amount of SRM reported here for adults (about 2 dB) is at the lower range of previous reports of 2–4 dB for a single interferer (Hawley *et al.*, 1999, 2004). The small amount of SRM reported here may be attributed to several methodological differences. First, another level of task difficulty is introduced when open-set sentences are used (e.g., Hawley *et al.*, 2004) compared with the closed-set spondees, even in a 25-AFC task. Second, different gender voices were used for the target (male) and interferers (female-based) in the current study. Previous research has shown that adults can experience a 4–5 dB increase in SRM when the target and interferer are the same gender voices versus different gender voices and with a more difficult task (e.g., Brungart and Simpson, 2002a). Finally, a single interferer was used in the current study. Other research has shown that when multiple linguistically relevant interferers are present the amount of SRM also increases (Hawley *et al.*, 2004). One of the greatest benefits of SRM may be the improvement in speech intelligibility afforded under conditions in which the listening environments contain the most challenging content and/or contexts.

The children in this experiment demonstrate greater amounts of SRM with interferers that have speech-like properties (an informational masker) compared with the interferer that has the same long-term spectrum as the target (an energetic masker). In addition, children showed largest SRM for the time-reversed speech interferer (6.7 dB) compared with speech (3.4 dB) or MN (0.5 dB). We hypothesize that the novelty of a reversed speech signal affected how children allocated their attention. Anecdotally, one child asked the experimenter what language the lady was speaking. It may be that the children gave the reversed-speech interferer special attention because it sounded like a real, albeit foreign language to them. The adults, on the other hand, having had more extensive real-world knowledge would either not have mistaken it for a foreign language, or would have treated it like any other speech-like sound.

As an interferer, time-reversed speech is unusual. It has no semantic content, and reversal of speech in the time domain renders discrimination difficult if not impossible (Ramos *et al.*, 2000), in part because the reversal dramatically alters the onset and offset patterns for voicing and the patterns of closure duration for stop consonants (Rosen, 1992). However, reversal of speech wave forms in time does not eliminate their having a speech-like quality; many aspects of forward speech are preserved, such as F0, formant transitions, and frication.

While adults do show impaired discrimination of time-reversed stop syllables (Li and Boothroyd, 1996) when perception is measured, the use of time-reversed speech as a masker does not markedly alter performance as compared with a forward speech interferer (Brungart and Simpson, 2002b; Hawley *et al.*, 2004). In contrast, for children who participated in the current study, reversed speech produced different findings than the forward speech. This may be because the reversed speech masker provides more informational masking for children than adults. The general idea that SRM is greater with maskers that have more informational masking has been discussed in recent years by others (Arbogast *et al.*, 2002; Freyman *et al.*, 2001). In those studies the

focus was on comparison of speech and noise-type maskers. Results presented here suggest that informational masking contributes to greater SRM in children, but the difference is borne out of comparisons between reversed-speech and noise type maskers.

E. SRM is independent of female voice spectra

This study examined the effect of two different female voices on masking and SRM in children and adults. The spectra of female voices are notoriously variable, in particular with regard to energy components that are high frequency. Therefore, the question of whether variability in energetic masking affects SRM is highly relevant to studies on the cocktail party effect. We have demonstrated that, although the amount of masking produced by a female voice containing spectral energy approximating that of an "average" female voice (Stelmachowicz *et al.*, 1993) is greater than that for a female voice with reduced spectral energy in the high frequencies, the ability of listeners to utilize spatial cues for source segregation remains the same. This is because the amount of masking produced by Female B leads to increase in the Front and Right SRTs that are virtually identical, for both children and adults. SRM is therefore not dependent directly on the type of female voice used, rendering this measure quite robust under several conditions.

V. CONCLUSIONS

Under conditions studied here, children generally perform worse than adults in the presence of a single interferer. Overall, children have higher SRTs and experience greater masking. In addition, masking and SRM vary with the type of interfering sound, and this variation interacts with age. Compared with adults, children experience greater amounts of masking, and also SRM, especially in the presence of a time-reversed speech interferer. This is quite interesting, because, although this interferer is probably not encountered in everyday listening situations, it may be akin to novel sounds such as foreign languages, which are indeed a part of children's realistic auditory environments. Finally, the exact voice that is used for measuring these effects contributes to the amount of energetic masking, but does not affect the benefit of spatially segregating target speech from interferers. One of the long-term goals of this work is to identify situations that enable some level of prediction about children's ability to function in complex environments. Ultimately, such measures can potentially be applied in clinical settings to assess performance of children with hearing impairments and hearing prosthetic devices.

ACKNOWLEDGMENTS

This research was supported by the NIH-NIDCD (Grant Nos. R01-DC003083 and R21-DC05469 to R.Y.L.). The authors would like to thank Allison Olson and Shelly Godar for helping with data collection, Dr. Mary Lindstrom for assistance with statistical analyses, and Dr. Robert Lutfi for suggestions regarding deviance measurements for psychometric functions. We are very grateful to the children and adults who participate in our "listening games."

[1]The deviance measure is $D=2\Sigma_{i=1}^{K}\{n_i p_i \log(p_i/p_i)+n_i(1-p_i)\log((1-p_i)/(1-p_i))\}$ where K equals the number of data points of each psychometric function, and n_i is the number of trials for the ith data point.

Allen, P., and Wightman, F. (**1995**). "Effects of signal and masker uncertainty on children's detection," J. Speech Hear. Res. **38**, 503–511.

ANSI (**1987**). ANSI S3.9–1987, *American National Standards Specification for Instruments to Measure Aural Acoustic Impedance and Admittance* (American National Standards Institute, New York).

ANSI (**1989**). ANSI S3.9–1989, *American National Standards Specification for Audiometers* (American National Standards Institute, New York).

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

Berman, S., and Friedman, D. (**1995**). "The development of selective attention as reflected by event-related potentials," J. Exp. Child Psychol. **59**, 1–31.

Bronkhorst, A. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acoustica **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. S., and Simpson, B. D. (**2002a**). "The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal," J. Acoust. Soc. Am. **112**, 664–676.

Brungart, D. S., and Simpson, B. D. (**2002b**). "Within-ear and across-ear interference in a cocktail-party listening task," J. Acoust. Soc. Am. **112**, 2985–2995.

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

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and two ears," J. Acoust. Soc. Am. **25**, 975–979.

Cranford, J. L., Morgan, M., Scudder, R., and Moore, C. (**1993**). "Tracking of 'moving' fused auditory images by children," J. Speech Hear. Res. **36**, 424–430.

Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**, 1057–1065.

Dirks, D. D., and Wilson, R. H. (**1969**). "The effect of spatially separated sound sources on speech intelligibility," J. Speech Hear. Res. **12**, 650–664.

Doyle, A. (**1973**). "Listening to distraction: A developmental study of selective attention," J. Exp. Child Psychol. **15**, 100–115.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr. (**2003**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**(1), 368–379.

Eggermont, J. J., Brown, D. K., Ponton, C. W., and Kimberley, B. P. (**1996**). "Comparison of distortion product otoacoustic emission (DPOAE) and auditory brain stem response (ABR) traveling wave delay measurements suggests frequency-specific synapse maturation," Ear Hear. **17**, 386–394.

Feston, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition." J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Halfer, 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.

Gomes, H., Malholm, S., Christodoulou, C., Ritter, W., and Cowan, N. (**2000**). "The development of auditory attention in children," Front. Biosci. **5**, d108–d120.

Green, D. M., and Swets, J. (**1974**). *Signal Detection Theory and Psychophysics* (Kreger, New York).

Grose, J. H., Hall, J. W., III, and Gibbs, C. (**1993**). "Temporal analysis in children," J. Speech Hear. Res. **36**, 351–356.

Guttentag, R. E., and Ornstein, P. A. (**1990**). "Attentional capacity and chil-

dren's memory strategy use," in *Development of Attention: Research and Theory*, edited byJ. T. Enns (Elsevier, Amsterdam). pp. 305–320.
Hall, J. W., III, and Grose, J. H. (**1994**). "Development of temporal resolution in children as measured by the temporal modulation transfer function," J. Acoust. Soc. Am. **96**, 150–154.
Hall, J. W., III, Grose, J. H., Buss, E., and Dev, M. B. (**2002**). "Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children," Ear Hear. **23**, 159–165.
Hartley, D. E., Douglas, E. H., Wright, B. A., Hogan, S. C., and Moore, D. R. (**2000**). "Age-related improvements in auditory backward and simultaneous masking in 6- to 10-years-old children," J. Speech Lang. Hear. Res. **43**, 1402–1415.
Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). "Speech intelligibility and localization in a multi-source environment," J. Acoust. Soc. Am. **105**, 3436–3448.
Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **105**, 3436–3448.
Hazan, V., and Markham, D. (**2004**). "Acoustic-phonetic correlates of talker intelligibility for adults and children" J. Acoust. Soc. Am. **116**, 3108–3118.
Kidd, G., Mason, C. R., and Arbogast, T. L. (**2002**). "Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **111**, 1367–1376.
Leek, M. R., Brown, M. E., and Doorman, M. F., (**1991a**). "Informational masking and auditory attention," Percept. Psychophys. **50**, 205–214.
Leek, M. R., Thomas, E. H., and Marshall, L. (**1991b**). "An interleaved tracking procedure to monitor unstable psychometric functions," J. Acoust. Soc. Am. **90**, 1385–1397.
Levitt, H. (**1971**). "Transformed up-down methods in psychophysics," J. Acoust. Soc. Am. **49**, 467–477.
Li, A. C., and Boothroyd, A. (**1996**). "Speech perception of temporally reversed syllables by normally hearing adults" (in Chinese), Zhonghua Yi Xue Za Zhi, (Taipei) **57**, 1–6.
Litovsky, R. Y. (**2005**). "Speech intelligibility and spatial release from masking in young children," J. Acoust. Soc. Am. **117**, 3091–3099.
Litovsky, R. Y. (**2004**). "Method and system for rapid and reliable testing of speech intelligibility in children," J. Acoust. Soc. Am. **115**, 2699.
Litovsky, R. Y. (**1997**). "Developmental changes in the precedence effect: Estimates of minimum audible angle," J. Acoust. Soc. Am. **102**, 1739–1745.
Litovsky, R. Y., and Ashmead, D. (**1997**). "Developmental aspects of binaural and spatial hearing," in *Binaural and Spatial Hearing*, edited by R. H. Gilkey and T. R. Anderson (Earlbaum, Hillsdale, NJ) pp. 571–592.
Lutfi, R. A., (**1990**). "How much masking is informational masking?" J. Acoust. Soc. Am. **88**, 2607–2610.
Lutfi, R. A., Kistler, D. J., Callahan, M. R., and Wightman, F. L. (**2003**). "Psychometric functions for informational masking," J. Acoust. Soc. Am. **114**, 3273–3282.
Macoby, E. (**1969**). "The development of stimulus selection," Minn. Symp. Child Psych. **3**, 68–96.
Morrongiello, B. A., Kulig, J. W., and Clifton, R. K. (**1984**). "Developmental changes in auditory perception," Child Dev. **55**, 461–71.
Papso, C. F., and Blood, I. M. (**1989**). "Word recognition skills of children and adults in background noise," Ear Hear. **10**, 235–236.
Pessig, 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., 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.
Plude, D. J., Enns, J. T., and Brodeur, D. (**1994**). "The development of selective attention: A life-span overview," Acta Psychol. **86**, 227–272.
Ponton, E. W., Eggermont, J. J., Coupland, S. G., and Winkelaar, R. (**1992**). "Frequency-specific maturation of the eighth nerve and brain-stem auditory pathway: Evidence from derived auditory brain-stem responses (ABRs)," J. Acoust. Soc. Am. **91**, 1576–1586.
Ramos, F., Hauser, M. D., Miller, C., Morris, D., and Mehler, J. (**2000**). "Language discrimination by human newborns and by cotton-top tamarin monkeys," Science **288**, 349–352.
Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbigert, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). "IEEE Recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.
Stelmachowicz, P., Mace, A. L., Kopun, J. G., and Carney, E., (**1993**). "Long-term and short-term characteristics of speech: Implications for hearing aid selection for young children," J. Speech Hear. Res. **36**, 609–620.
Veloso, K., Hall, J. W., III, and Grose, J. H. (**1990**). "Frequency selectivity and comodulation masking relsease in adults and in 6-years-old children," J. Speech Hear. Res. **33**, 96–102.
von Békésy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).
Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). "Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty," J. Acoust. Soc. Am. **60**, 1176–1186.
Wichmann, F. A., and Hill, J. (**2001a**). "The psychometric function. I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1290–1313.
Wichmann, F. A., and Hill, J., (**2001b**). "The psychometric function. II. Bootstrap-based confidence intervals and sampling," Percept. Psychophys. **63**, 1314–1329.

Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing[a)]

Piotr Majdak[b)] and Bernhard Laback
Acoustics Research Institute, Austrian Academy of Sciences, Reichsratsstrasse 17, A-1010 Wien, Austria

Wolf-Dieter Baumgartner
ENT-Department, Vienna University Hospital, Währinger Gürtel 18-20, A-1097 Wien, Austria

(Received 30 November 2005; revised 30 June 2006; accepted 30 June 2006)

Bilateral cochlear implant (CI) listeners currently use stimulation strategies which encode interaural time differences (ITD) in the temporal envelope but which do not transmit ITD in the fine structure, due to the constant phase in the electric pulse train. To determine the utility of encoding ITD in the fine structure, ITD-based lateralization was investigated with four CI listeners and four normal hearing (NH) subjects listening to a simulation of electric stimulation. Lateralization discrimination was tested at different pulse rates for various combinations of independently controlled fine structure ITD and envelope ITD. Results for electric hearing show that the fine structure ITD had the strongest impact on lateralization at lower pulse rates, with significant effects for pulse rates up to 800 pulses per second. At higher pulse rates, lateralization discrimination depended solely on the envelope ITD. The data suggest that bilateral CI listeners benefit from transmitting fine structure ITD at lower pulse rates. However, there were strong interindividual differences: the better performing CI listeners performed comparably to the NH listeners. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2258390]

PACS number(s): 43.66.Pn, 43.66.Ts, 43.66.Qp, 43.66.Mk [AJO] Pages: 2190–2201

I. INTRODUCTION

An important cue for the localization of sound sources is the interaural time difference (ITD). It is well established that ITD information in unmodulated signals can only be processed up to about 1500 Hz (Zwislocki and Feldman, 1956; for reviews see Blauert, 1997, and Wightman and Kistler, 1997). At higher frequencies a slow modulation of the carrier transmits the ITD information (e.g., Bernstein, 2001). Using modulated signals, like speech, at least two different types of ITD can be defined: ITD in the envelope (ITD ENV) and ITD in the fine structure (ITD FS). Signals with equal ITD ENV and ITD FS can be considered as a special case in that the whole waveform of one channel is delayed relative to the other channel. This case is most often found in natural signals and is referred to as waveform delay (WD).

Several studies have examined ITD perception in cochlear implant (CI) listeners. Lawson *et al.* (1998) showed that lateralization discrimination using ITD only is possible. Using a pulse rate of 480 pulses per second (pps) they obtained a just noticeable difference (JND) of 150 μs. More detailed studies were performed by van Hoesel and Clark (1997), van Hoesel *et al.* (2002), and van Hoesel and Tyler (2003). In general, the performance was much worse than that of normal hearing (NH) listeners and had high intersubject variability. For unmodulated stimuli the JNDs increased for higher pulse rates and could not be determined at a pulse rate of 800 pps. However, when using low-frequency amplitude-modulated stimuli at this pulse rate, JNDs were on the order of JNDs for unmodulated stimuli with carrier pulse rate equal to that of the low-frequency modulation. Unfortunately, they did not separate the relative contribution of ITD ENV and ITD FS, which may be important for amplitude-modulated stimuli like speech. Laback *et al.* (2004) investigated the effects of ITD ENV manipulation (the ITD FS was random and uncontrolled) in electric hearing by presenting acoustic stimuli via unsynchronized speech processors. They showed that JNDs differed between NH subjects (19 μs) and CI listeners (259 and 384 μs, best JNDs for CI listener S2 and S1, respectively) and depended on the type of stimulus (lowest for click trains, highest for speech or noise bursts).

Current cochlear implant systems use a variety of stimulation strategies to transmit the acoustic information; an overview of which is given in Wilson (2004). Almost all strategies were designed for monaural use and do not include any binaural synchronization: the electric stimulation is controlled by two independently running speech processors. As a result, the ITD information is coded in the envelope only. The strategies have one other aspect in common: according to the specification, they use a constant stimulation pulse rate at both ears. Due to the lack of synchronization between the two ears, the stimulation pulses have an interaural delay, which can be regarded as an ITD FS. This depends on the switch-on delay between the processors and has a random value between 0 μs and the interpulse interval (IPI). If CI listeners are sensitive to ITD FS, it will interact with other lateralization cues like ITD ENV or interaural level differences.

[a)]Portions of this work were previously presented at the 28th Midwinter Research Meeting of the Association for Research in Otolaryngology, New Orleans 2005 and at the Conference on Implantable Auditory Prostheses, Asilomar 2005.

[b)]Electronic mail: piotr@majdak.com

TABLE I. Clinical data of CI listeners.

Subject	Aetiology	Age at implant L yr	Age at implant R yr	Deafness duration L	Deafness duration R	Binaural electric stimulation experience
CI1	Meningitis	14	14	5.5 months	1.5 months	6 yr
CI2	Skull trauma	54	48	21 yr	25 yr	4 yr
CI3	Meningitis	21	21	2 months	2 months	1 month
CI8	Osteogenesis imperfekta	41	39	3 yr	12 yr	2 months

Due to the manufacturing tolerances, the time bases deviate between the speech processors at the two ears, resulting in different IPI. Therefore, the pulse rates cannot be assumed to be equal at both ears. This leads to a dynamically changing ITD FS, which varies between 0 and the IPI. The period of this "ITD beat" increases with decreasing deviation in the pulse rates. If subjects are sensitive to ITD FS, the dynamically changing ITD FS will result in a movement of the auditory image.

It was shown in a recent study (Laback *et al.*, 2005) that ITD FS contributes to lateralization discrimination for lower pulse rates. In this case, a controlled ITD FS may support the effect of ITD ENV and improve the lateralization of sound sources. Coding ITD FS information also may be advantageous for speech perception in noise (Licklider, 1948; Hirsh, 1950; Dirks and Wilson, 1969; Bronkhorst and Plomp, 1988; Hawley *et al.*, 1999) or speech segregation (Drennan *et al.*, 2003; Culling *et al.*, 2004). One study, performed with anesthetized cats, indicates that ITD FS in a low-frequency carrier may be a much stronger cue than ITD ENV in an amplitude-modulated high-frequency carrier: Smith and Delgutte (2005) showed that the neuronal tuning curves in the inferior colliculus are sharper for ITD FS in a low-frequency stimulus (tested up to 320 pps) than for ITD ENV in a high-frequency carrier (tested 1000 pps carrier and modulation frequencies up to 160 Hz).

The goal of this study is to systematically investigate the effects of fine structure ITD manipulation on lateralization discrimination in electric stimulation using amplitude-modulated stimuli. It was expected that CI listeners would be sensitive to ITD FS at lower pulse rates. In addition, the same experiments were performed with normal hearing (NH) subjects using a simulation of electric stimulation to compare their performance with that of the CI listeners. The results allow the assessment of the need for the synchronization of speech processors, taking some synchronization methods into account.

II. METHODS

A. Subjects and apparatus

Four NH subjects participated in this study, of whom one (NH3) was female. All subjects were between the ages of 25 and 35 years old and had no indication of hearing abnormalities. Two of them were the authors of this study (NH2, NH4).

Four cochlear implant (CI) listeners were tested. Three of them were implanted bilaterally with the C40+ implant system manufactured by MED-EL Corp. This system provides pulsatile, nonsimultanous biphasic current pulses on up to 12 electrodes with a minimum phase duration of 26.7 μs. One CI listener (CI2) used the C40+ in the left ear and an older implant, the C40, in his right ear. The C40 provides current pulses on up to eight electrodes with a minimum phase duration of 40 μs. Clinical data of CI listeners can be found in Table I. The subjects were selected from a total of seven CI listeners invited for participation in the study. These four listeners fulfilled the selection criterion, as defined by the ability to reproducibly perform left/right discrimination on the basis of waveform ITD in a pulse train with a pulse rate of 100 pps in a reasonable amount of time.

A personal computer system was used to control electric and acoustic stimulation. Each implant was controlled by a Research Interface Box (RIB), manufactured at the University of Technology Innsbruck, Austria. The two RIBs were synchronized, providing an interaural accuracy of stimulation timing better than 2.5 μs. Prior to the experiment, the stimuli were verified using a pair of dummy implants (Detektorbox, MED-EL). The stimuli for acoustic stimulation were output via a 24-bit stereo A/D-D/A converter (ADDA 2402, Digital Audio Denmark) using a sampling rate of 96 kHz per channel. The analog signals were sent through a headphone amplifier (HB6, TDT) and an attenuator (PA4, TDT) and presented to the subjects via a circumaural headphone (K501, AKG). Calibration of the headphone signals was performed using a sound level meter (2260, Brüel & Kjær) connected to an artificial ear (4153, Brüel & Kjær).

B. Stimuli

The stimuli were amplitude-modulated pulse trains which were designed as pulse trains multiplied by a predefined envelope. ITD FS and ITD ENV were introduced by delaying the temporal position of the pulses and of the envelope, respectively, at one ear relative to the other ear. The following ITD conditions were specified: ITD in envelope only (ENV), ITD in fine structure only (FS), no ITD at all, which is the reference condition (REF), and the identical ITD in both the envelope and the fine structure, referred to as waveform delay (WD).

The envelope consisted of four trapezoids with durations of 60 ms, each repeated at a period of 80 ms, resulting in 20 ms gaps between two successive trapezoids and a total stimulus duration of 300 ms (Fig. 1). The trapezoid period of 80 ms yields an amplitude modulation frequency of 12.5 Hz. Since the envelope modulation is trapezoidal, the modulation

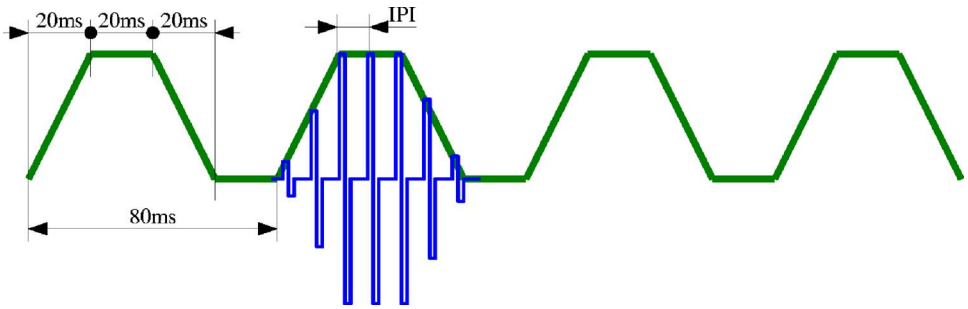


FIG. 1. A schematic representation of the stimulus used in this study. For readability purposes the fine structure characteristics are shown in one trapezoid only. The ramps slope down to the absolute threshold of each subject. Between the trapezoids the amplitude was set to zero. In acoustic stimuli, pulses with positive amplitude were applied.

spectrum contains multiples of the 12.5 Hz as well. There were several reasons for selecting a relatively slow modulation rate. Although sensitivity to ITD increases with growing modulation frequency up to approximately 125 Hz (Henning, 1974; Bernstein, 2001) values in that range interfere with the lower limit of the pulse rates used here (100 pps). Even modulation rates as low as tens of hertz reduce the information available in the fine structure. Furthermore, pulse trains with modulation rates on the order of 12.5 Hz more closely resemble real-world signals than pulse trains modulated with higher rates. In particular, speech has a modulation spectrum peak of approximately 5 Hz (Greenberg *et al.*, 2003). Considering these aspects, these values of modulation frequency lead to signals providing sufficient ITD information in both the fine structure and the envelope. Finally, the rise and release time of each trapezoid was set to 20 ms. This value was chosen to emphasize the onset and the offset effects, which are assumed to enhance sensitivity to ITD ENV. The level of the first and last pulse of each trapezoid was set at the subject's absolute threshold, which was determined in pretests (see Sec. II C). Between the trapezoids, the amplitude was set to zero. The acoustic amplitudes were interpolated logarithmically and the electric currents were interpolated linearly.

Both van Hoesel (2003) and Laback *et al.* (2005) found that subjects differed strongly in their sensitivity to ITD as a function of pulse rate. In the two studies the sensitivity was generally highest at lowest pulse rates tested, which were 50 and 100 pps, respectively. On the other hand, current stimulation strategies use pulse rates up to about 1600 pps (Wilson, 2004), thus, testing these pulse rates is important for real world applications. Consequently, the pulse rates to be tested must be selected individually for each subject: three to four pulse rates between 100 and 1600 pps, corresponding to IPI between 10 ms and 625 μs, were chosen for each subject on the basis of lateralization discrimination pretests described in Sec. II D.

In the case of electric stimulation, the pulse trains were composed of biphasic current pulses. Each phase of a pulse had a duration of 26.7 and 40 μs for the C40+ and the C40 devices, respectively. An interaurally pitch-matched electrode pair, selected in pretests (see Sec. II C), was used for all experiments.

To allow a direct comparison of the results from NH subjects with those from CI subjects, the electric stimulation was simulated in the NH subjects using a method developed by McKay and Carlyon (1999) and further successfully applied by Carlyon *et al.* (2002). Pulse trains were composed of monophasic pulses with a duration of 10.4 μs, corresponding to one sampling interval at a sampling rate of 96 kHz. The pulse trains were passed through a digital eighth-order Butterworth filter with a geometric center frequency of 4590 Hz and −3 dB bandwidth of 1500 Hz.

Due to the filtering of amplitude-modulated pulse trains, a possible naming clash might have been introduced. In the NH literature, the "fine structure" of the acoustic stimulus refers to the carrier frequency, which is 4590 Hz in our case. Following this definition, every filtered pulse has an "envelope," which is the envelope of the impulse response of the bandpass filter. Furthermore, the envelope of the pulse train appears as a second-order envelope. The carrier frequency arises from the filtering procedure, which is not the object of interest in this study. Thus, the definitions from the CI literature have been adopted to the acoustic signals. In reference to acoustic signals, the term "fine structure" defines the total impulse response of the bandpass filter, not the carrier only, and "envelope" refers to the slow trapezoidal amplitude modulation of the filtered pulse trains. Keeping in mind that the acoustic stimuli represent a simulation of electric stimulation, the same terms can be used to describe electric and acoustic stimulation effects.

Given that the sound pressure level (SPL) depends on the pulse rate, stimulation amplitudes were adjusted to maintain a constant SPL of 59 dB, measured at the headphones, at all rates for all NH subjects. Despite the filtering of the pulse trains, some artifacts like harmonic distortions or intermodulation at the basilar membrane can cause stimulation outside the desired frequency band. To prevent these artifacts from being heard, a binaurally uncorrelated pink noise with a spectrum level of 15.2 dB SPL at 4.6 kHz was continuously played throughout the testing.

Eight ITD FS values were chosen for each pulse rate, which corresponded to values from 0 μs up to seven-eighths of the IPI in steps of eighth IPI. These values covered the range of ITD FS which would occur in a setup of unsynchronized speech processors and included ITDs exceeding the natural head-width delay for lower pulse rates. The investigations on effects of ITD ENV were secondary in this study; thus, only two values were used. The intended values of 400 and 625 μs represent large ITD values with respect to the head size, and correspond to ITD ENV cues as they occur in real-world situations. Unfortunately, in the lateralization pretests the CI listeners showed no sensitivity at 400 μs and very low sensitivity at 625 μs. Thus, intending to produce as much effect as possible, larger ITD ENV values were chosen for the CI listeners: 625 and 800 μs.

C. Pretests

In the experiments with CI users, pretests were performed to determine a binaurally loudness balanced, pitch-matched electrode pair for each listener. The pretests used pulse trains of 300 ms duration with zero ITD, 100 pps pulse rate, no amplitude modulation, and consisted of a manual up/down procedure to estimate each listener's threshold (THR), comfortable level (CL), and maximum comfortable level (MCL); a balancing procedure to iteratively determine levels of binaurally equal loudness for each electrode pair; a monaural pitch estimation procedure to reduce the number of candidates for pitch matching for both ears; and a pitch ranking procedure to determine the pitch discriminability for the pair candidates and finally select one pitch matched pair.

To determine the THR, CL, and MCL for each electrode the perceived loudness was indicated by the subjects by pointing to the appropriate position on a continuous scale, ranging from "not audible" to "just uncomfortably loud." The CL corresponded to the subject's response "comfortable." The same procedure was then applied to determine the binaural CL, i.e., the comfortable level when both ears were stimulated simultaneously. Starting at 80% of the monaural CLs, levels were varied simultaneously in equal steps at the two ears. Subjects were instructed to attend to the overall loudness in the binaural case rather than to "hear out" a left-ear or right-ear contribution. Following the initial adjustment of the binaural CL, centralization of the perceived stimulus was checked and monaural levels were adjusted if necessary. All subjects required a reduction of current levels in the binaural condition relative to the monaural conditions to achieve the same loudness.

A magnitude estimation procedure was applied to obtain an estimate of the perceived pitch across the electrodes at both ears, similar to the procedures applied by Busby *et al.* (1994) and Collins *et al.* (1997). Stimuli were presented randomly between both ears and at each of the electrodes 1–8, using the binaural CLs determined before. Subjects were instructed to assign numbers according to the perceived pitch of each stimulus. No restrictions on the range and type of numbers were given. Each stimulus was presented ten times. The distribution of pitch judgments across the electrodes and the two ears allowed selection of about 16 interaural electrode pairs supposed to elicit similar pitch sensation at the two ears. These pairs were evaluated further in the pitch-ranking task.

An automated procedure was applied to obtain interaurally loudness-balanced levels for each of the electrode pairs used further in the pitch ranking task. The members of each electrode pair were presented in two subsequent intervals. By pressing one of two buttons the subjects adjusted the relative level of the signals between the two ears in steps corresponding to the smallest amplitude changes realizable by the implants to arrive at an interaurally matched loudness. The sum of the two levels within a trial was held constant and corresponded to the sum of the binaural CLs determined for the respective electrodes. The level difference at the beginning of each run was randomly roved. The mean value resulting from four runs was defined as the loudness-balanced levels for the members of the respective electrode pair.

TABLE II. Stimulation levels for each CI listener as parameter of pulse rate. "…" shows not tested pulse rates.

	Stimulation current left/right in μA			
Pulse rate in pps	CI1	CI2	CI3	CI8
100	…	358/1045	…	…
150	…	355/1045	…	…
200	478/486	362/1031	…	…
400	470/401	393/909	478/524	376/586
600	…	…	…	…
800	…	…	440/470	376/586
938	…	…	…	371/532
1600	501/424	…	347/370	…

In the pitch-ranking procedure the members of each of the electrode pairs were directly compared with respect to the perceived pitch difference, using a two-interval, two-alternative forced-choice (2-AFC) procedure. The pair members were presented randomly either in the first or second observation interval. Subjects were required to indicate which of the two stimuli sounded higher in pitch while concentrating on pitch rather than on other attributes such as timbre or loudness. Electrode pairs with an average discriminability across 25 repetitions within the range of chance (50±18%) were considered as pitch-matched. For subjects with more than one pitch-matched electrode pair, the pair at medial tonotopic position was chosen. The selected electrode pairs were (left/right): 4/1 (CI1), 2/3 (CI2), 4/3 (CI3), and 7/5 (CI8).

Using the automated loudness balancing procedure described earlier, the levels for each pulse rate were determined with the goal of obtaining a binaurally balanced, comfortable loudness level for the selected pitch-matched electrode pair. Table II depicts the subject-dependent stimulation currents for each pulse rate.

D. Procedure

A two-interval, 2-AFC procedure was used in the lateralization discrimination tests. The first interval contained a reference stimulus with zero ITD, evoking a centralized auditory image. The second interval contained the target stimulus with the ITD tested. The subjects were requested to indicate whether the second stimulus was perceived to the left or to the right of the first one by pressing an appropriate button. All stimuli were repeated at least 60 times, in a balanced format with 30 targets on the left and 30 targets on the right. Thus, a subject with no ITD sensitivity could get 50% responses correct by guessing. A score of 100% correct responses would indicate that all stimuli were discriminated, with lateralization corresponding to the ear receiving the leading signal. In contrast, a score of 0% implies perfect discrimination as well, but with lateralization at the ear receiving the delayed signal.[1] To avoid biasing the subjects toward a particular manner of responding, no feedback was

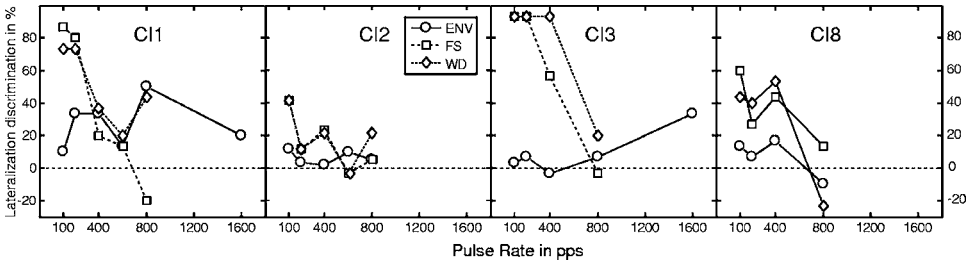


FIG. 2. Pretest results as lateralization discrimination (LD) for different pulse rates and four CI listeners. Conditions: ENV: ITD ENV=625 μs; FS: ITD FS=625 μs; WD: ITD FS=ITD ENV=625 μs.

given. To simplify the interpretation of the results, scores ranging from 0% to 100% were mapped to a range from −100% to +100%, referred to as "lateralization discrimination" (LD). Lateralization discrimination of 0% means that the target could not be discriminated from the reference stimulus with respect to the lateral position and represents 50% correct responses.

Lateralization discrimination pretests were performed to select, for each listener, the pulse rates to be used in the main experiment. Discarding conditions with very low sensitivity to ITD kept the test time as short as possible. In the pretests, one ITD value of 625 μs was presented in three different ITD conditions: ENV, FS, and WD. The results for each CI listener are shown in Fig. 2. Based upon these results and the availability of the subjects three to four pulse rates were chosen for each subject to be tested in main experiments (CI1: 200, 400, 1600 pps; CI2: 100, 150, 200, 400 pps; CI3: 400, 800, 1600 pps; CI8: 400, 800, 938 pps). The NH subjects were tested at 400, 600, 800 and 938 pps.

III. RESULTS

The LD data of the individual CI listeners are shown in Fig. 3 (CI1), Fig. 4 (CI2), Fig. 5 (CI3), and Fig. 6 (CI8). The results of the NH subjects were more homogeneous; as a result, the mean scores of all four NH listeners are provided in Fig. 7. For all listeners there was a common pattern of LD as a function of ITD FS. At the lowest pulse rates (different for each subject) in the conditions ITD ENV=0 μs, LD increased monotonically with ITD FS for ITD FS less than 0.25 IPI with a maximum at about 0.25 IPI. For ITD at approximately 0.5 IPI, LD was at chance (=0%), confirming the ambiguity in the lateralization task using ITD FS only. As ITD FS exceeded 0.5 IPI, the magnitude of LD as a function of ITD FS was similar to that for ITD FS$<$0.5 IPI but with the opposite sign. This indicates that LD upon ITD FS is periodic and that stimuli with ITD FS$>$0.5 IPI effectively represent stimuli with negative ITD FS. At the highest pulse rates tested (different for each subject) the dependence of LD on ITD FS disappeared. Introducing a nonzero ITD ENV resulted in a lateralization shift toward the ear receiving the stimulus with the leading envelope. This effect seems to increase with increasing pulse rate.

Although most trends were easily distinguishable by visual inspection, a statistical analysis was used to determine the significance of the trends. The statistical method employed was multidimensional contingency table analysis (Lienert, 1978; Agresti, 1984, 1996) implemented in "stats" package of R (R Development Core Team, 2004). This is a useful method for intersubject comparisons of results obtained by a 2-AFC task, for which the variance analysis is not available,[2] although it is an unusual method in psychoacoustics. A general description of this method would exceed the scope of this paper. Thus, only a summary of the tests and models applied in the context of the data analysis is provided.

The significance of the differences between two conditions was tested by obtaining the two-tailed probability p of the Pearson χ^2 statistic for the null hypothesis that the logarithmic odds ratios[3] for both conditions are equivalent (Agresti, 1984). Log-linear models were fitted to determine the interactions between different factors. The goodness-of-fit of a model to the data is described by the G^2, df, and p values. In these cases, the significance of an effect is given by the significance of the corresponding model parameter. In cases of fitted data, the calculation of odds ratios was done using estimated response frequencies, according to Agresti (1984, pp. 47–69). Some conditions were tested with a higher number of repetitions ($>$60) due to differences in the availability of subjects. Thus, the investigations of marginal associations, collapsing[4] the data over the variable tested with different number of repetitions, were done using regularization of the data to the smallest common number of repetitions (=60). To increase the test power, pulse rate and ITD ENV were treated as ordinal factors.

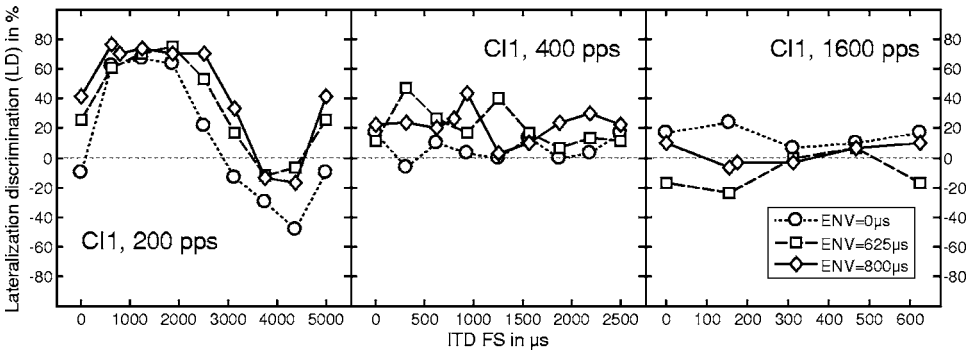


FIG. 3. Lateralization discrimination for CI1 and different pulse rates. To point out the periodicity of the ITD FS the data points for the ITD FS=IPI are copies of the data points for ITD FS =0 μs. Note the different scaling of the X axes.

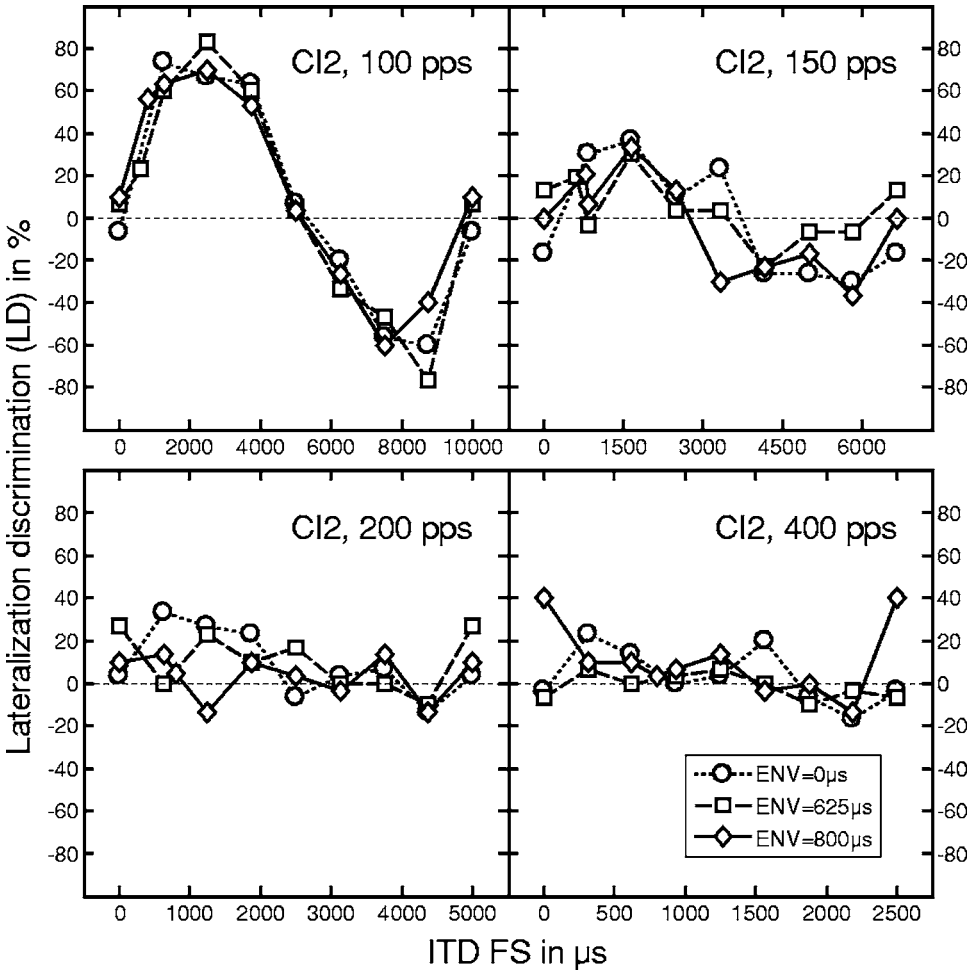


FIG. 4. Lateralization discrimination for CI2 and different pulse rates. To point out the periodicity of the ITD FS the data points for the ITD FS=IPI are copies of the data points for ITD FS $=0$ μs. Note the different scaling of the X axes.

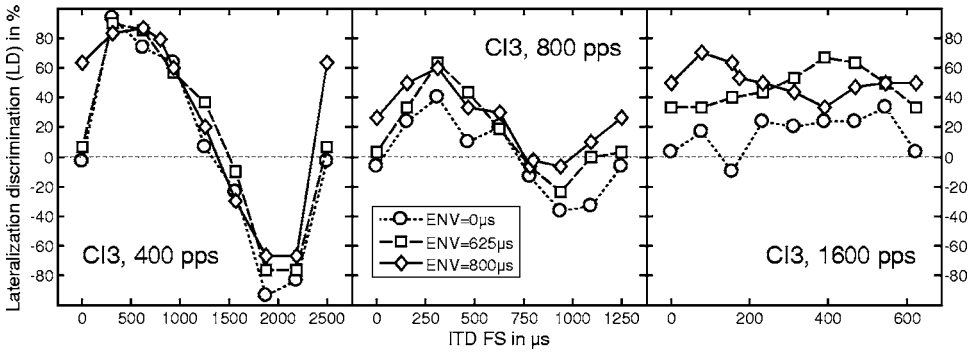


FIG. 5. Lateralization discrimination for CI3 and different pulse rates. All other conventions are as in Fig. 4.

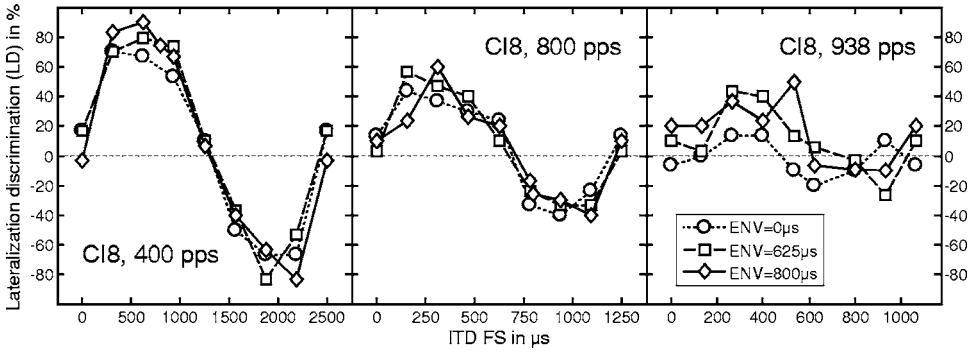


FIG. 6. Lateralization discrimination for CI8 and different pulse rates. All other conventions are as in Fig. 4.

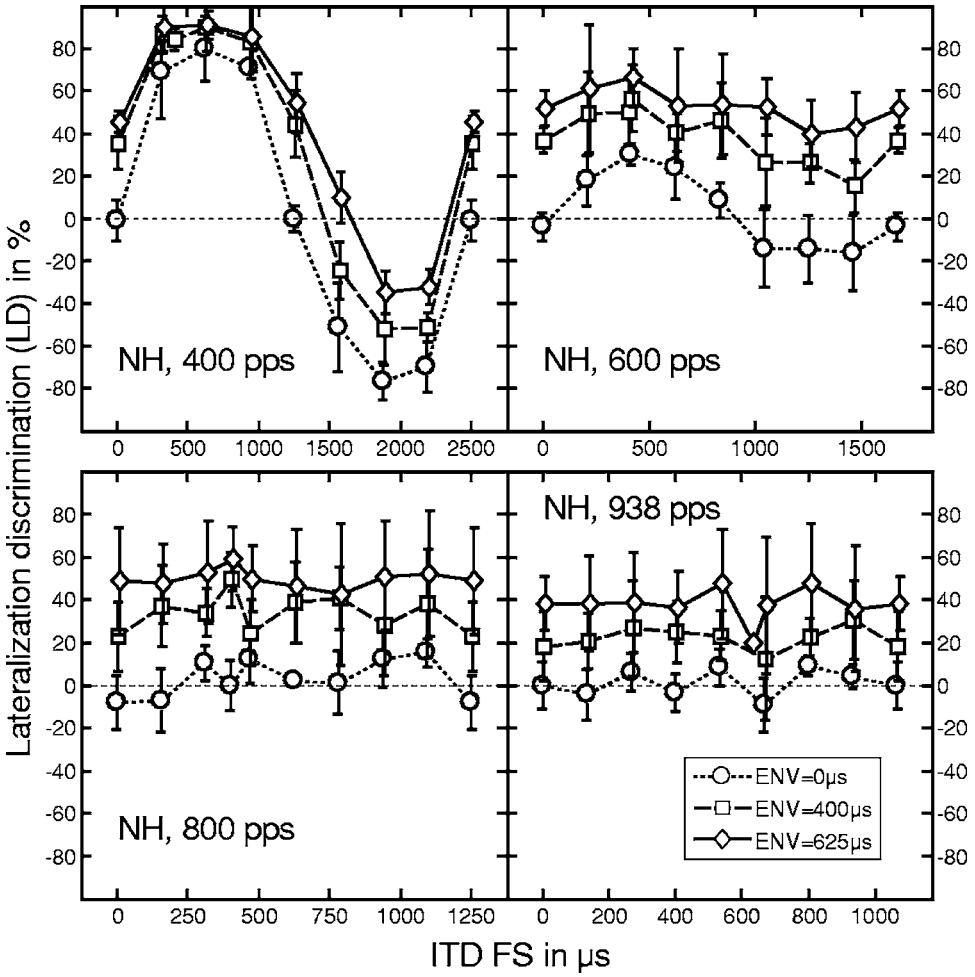


FIG. 7. Average lateralization discrimination for all NH subjects and different pulse rates. To point out the periodicity of the ITD FS the data points for the ITD FS=IPI are copies of the data points for ITD FS=0 μs. The bars shows the standard deviation. Note the different scaling of the *X* axes.

The statistical analysis is structured as follows: in Sec. III A the differences between subjects will be analyzed and a classification will be done to show the homogeneity in the results of the CI listeners and how it compares to the results of the NH listeners. Section III B considers effects of interaural synchronization on lateralization discrimination and shows some improvements which can be achieved by specific coding of the ITD. In Sec. III C the effects of ITD ENV will be analyzed in relation to the factors pulse rate and subject.

A. Groups of subjects

The pulse rate of 400 pps and ITD ENV of 0 and 625 μs were used for the analysis of differences between subjects, since they were the only values available for all subjects. Using these data, the odds ratios for the categorical factor subject were calculated and analyzed. The subjects were clustered according to the type of stimulation and the analysis was performed on each group separately. The results showed that the group of NH subjects was homogeneous ($\chi^2_3=0.07$, $p=0.995$) and the group of CI listeners was heterogeneous ($\chi^2_3=20.0$, $p<0.001$). Therefore, further analyses were performed on the NH listeners as a group and for each CI listener individually. However, the better performing CI listeners (CI3 and CI8) performed sufficiently similar to the NH subjects that they could have been clustered with the NH subjects to form a larger homogeneous group ($\chi^2_5=1.34$, $p=0.93$).

B. Interaural synchronization

To address the question of the need for interaural pulse synchronization, the dependence of LD on ITD FS was investigated. If a subject lateralizes the stimuli for ITD FS <0.5 IPI to one side and for ITD FS >0.5 IPI to the opposite side, that implies that LD depends on ITD FS. Therefore, the data for all conditions fulfilling ITD ENV=0 μs were grouped as follows: the first group contained all LDs for 0 $<$ITD FS<0.5 IPI and the second group all LDs for 0.5 IPI$<$ITD FS$<$IPI. Results for ITD FS=0 and 0.5 IPI were discarded as there is no lateralization information, and they should be at chance rate. It was hypothesized that if there is a significant difference between the direction of LD for both ITD FS groups, then LD depends on ITD FS, thus indicating the necessity of interaural pulse synchronization.

For the NH listeners the data pool (subject $\times$ ITD FS group $\times$ pulse rate $\times$ response) was collapsed across subjects because of the homogeneity of their performance. A log-linear model was fitted to this data pool with the result that only the saturated model[5] could give an accurate fit, showing a strong interaction between the factors pulse rate and ITD FS group. Thus, the dependency on ITD FS was investigated

TABLE III. Probability of no dependence of LD on ITD FS for ITD ENV=0 μs conditions. Conditions with the highest pulse rate with significant sensitivity on ITD FS are shown bold; df was 1 for all results.

Pulse rate in pps	NH		CI1		CI2		CI3		CI8	
	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p
100	⋯	⋯	⋯		117	<0.001	⋯	⋯	⋯	⋯
150	⋯	⋯	⋯		25.6	<0.001	⋯	⋯	⋯	⋯
200	⋯	⋯	165.1	**<0.001**	7.65	**0.006**	⋯	⋯	⋯	⋯
400	711.9	<0.001	0.1	0.752	1.61	0.21	185.4	<0.001	139.4	<0.001
600	77.89	**<0.001**	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
800	2.186	0.139	⋯	⋯	⋯	⋯	16.48	**<0.001**	42.73	**<0.001**
938	0.1361	0.712	⋯	⋯	⋯	⋯	⋯	⋯	0.571	0.45
1600	⋯	⋯	0.549	0.459	⋯	⋯	2.59	0.108	⋯	⋯

for each pulse rate separately, analyzing the odds ratios in partial contingency tables with fixed pulse rate. For the NH subjects 600 pps was the highest pulse rate with significant sensitivity on ITD FS. For the CI listeners the data were analyzed separately for each subject in the same way as for the NH listeners: odds ratios in partial contingency tables were analyzed for each pulse rate. The better performing CI listeners (CI3, CI8) showed significant sensitivity to ITD FS for pulse rates up to 800 pps, while for the poorer performing CI listeners (CI1, CI2) a significant sensitivity could be found only for pulse rates up to 200 pps. Detailed results of the analysis are shown in Table III.

It was further hypothesized that for conditions showing a dependence of LD on ITD FS, the synchronization of the ITD FS to the ITD ENV would result in a better LD than synchronizing ITD FS to zero. This was evaluated by keeping the ITD constant and comparing the LD between two synchronization conditions: one in which the ITD is carried both by the envelope and the fine structure (WD) or alternatively by the envelope only, keeping ITD FS at zero (ENV). In the statistical analysis, the synchronization conditions were regarded as a factor with two levels (WD, ENV) and the ITD as a factor with two levels, which depended on the subject group (NH: 400 and 625 μs; CI: 625 and 800 μs). The analysis was performed for the NH group and for each CI listener separately. As before, log-linear models were fitted to investigate the interactions. The hypothesis of no interaction between the factors ITD and synchronization condition could not be rejected for CI1 ($G^2=8.3867$, $df=10$, $p=0.591$) or CI8 ($G^2=8.4617$, $df=10$, $p=0.584$). Thus, the data were collapsed over ITD for these subjects. For all other subjects, separate partial tables were used for each ITD value. The probabilities for the hypothesis of equal LDs in both synchronization conditions for each subject and pulse rate showing dependence on ITD FS ($p<0.05$ in Table III), are given in Table IV.

The NH subjects showed an improvement using the WD condition for pulse rates up to 600 pps for an ITD of 400 μs ($p=0.04$). Increasing the ITD to 625 μs, the improvement due to synchronization decreased, and could be found for 400 pps only ($p<0.001$). This revealed an interesting effect of combining ITD FS and ITD ENV: assuming a dependence of LD on ITD FS, it can be expected that increasing the ITD in both the envelope and fine structure (WD) improves LD up to about 0.25 IPI. Above this point, up to ITD=0.5 IPI, LD is expected to decrease because at ITD=0.5 IPI the ITD FS cue provides ambiguous information. Increasing the ITD further, depending on the relative perceptual contribution of ITD ENV, the stimulus may even be lateralized toward the opposite side. This actually happened for CI8 (800 pps, ITD FS=800 μs, see Fig. 6). Thus, the synchronization of the fine structure to the envelope gives an improvement for ITD values smaller than half IPI only.

For the CI listeners, improvements due to synchronization were observed for the following pulse rates: 200 pps

TABLE IV. Probability for equal LD in conditions ENV and WD. Conditions with the highest pulse rate with significantly higher LD for WD than for ENV are shown in bold.

Pulse rate in pps	NH		CI1	CI2		CI3		CI8
	400 μs	625 μs	⋯[a]	625 μs	800 μs	625 μs	800 μs	⋯[a]
100	⋯	⋯	⋯	0.356	**0.007**	⋯	⋯	⋯
150	⋯	⋯	⋯	0.681	0.143	⋯	⋯	⋯
200	⋯	⋯	**<0.001**	0.141	0.729	⋯	⋯	⋯
400	<0.001	<0.001	⋯	⋯	⋯	**<0.001**	0.091	**<0.001**
600	**0.04**	0.696	⋯	⋯	⋯	⋯	⋯	⋯
800	⋯	⋯	⋯	⋯	⋯	0.266	0.043[b]	0.17

[a]ITD was marginalized in these cases.

[b]LD for ENV condition was higher than for WD condition. χ^2 values have been omitted for readability purposes.

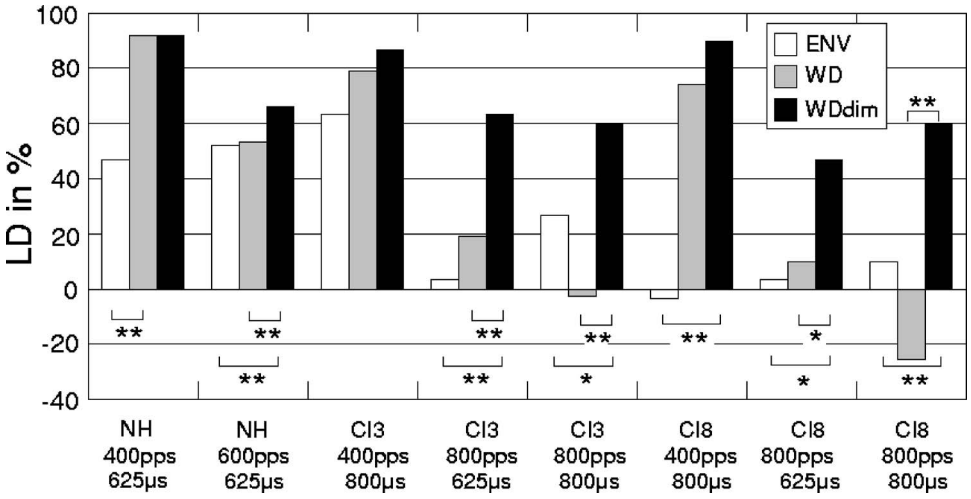


FIG. 8. Comparison of lateralization discrimination for conditions ENV, WD and WD_{DIM}. Significance codes: $^{*}p<0.05$; $^{**}p<0.01$.

(CI1), 100 pps (CI2), and 400 pps (CI3 and CI8). For CI3, at a pulse rate of 800 pps and ITD of 800 μs, there was a significant difference ($p=0.043$), but the LD was higher for ENV than for WD. Since the ITD exceeded 0.5 IPI, this is an example of the effect of combining conflicting ITD FS and ITD ENV, which was also seen in the NH subject' results. This effect seems to reduce the positive effects of synchronizing ITD FS to ITD ENV, but it allows an optimization of coding ITD FS: for ITD values greater than 0.25 IPI the WD condition was modified, diminishing the ITD FS to 0.25 IPI. This condition is termed *diminished waveform delay* (WD_{DIM}) and a new formula for ITD FS coding is proposed:

$$\text{ITD FS} = \min\left(\text{ITD ENV}, \tfrac{1}{4}\text{IPI}\right). \quad (1)$$

To obtain an improvement using WD_{DIM}, the ITD must be greater than 0.25 IPI. To fulfill this requirement for low ITDs the pulse rate must be as high as possible. On the other hand, the effect of ITD FS is weak for higher pulse rates. Thus, the WD_{DIM} optimization is efficient only for medium pulse rates showing sensitivity to ITD FS. Figure 8 compares lateralization discrimination between the conditions ENV, WD, and WD_{DIM}, for each of the ITDs and pulse rates measured and fulfilling the specified requirements. In most cases LD increased using WD_{DIM} optimization; in one case (CI8, 800 pps, 800 μs) even a reversal of lateralization into the correct direction could be achieved.

C. Effects of ITD ENV

To determine the effects of envelope delay, a comparison of the sensitivity to ITD ENV between the subjects was performed for stimuli with 400 pps and ITD ENV values of 0 and 625 μs. For subjects showing no effect of ITD FS at this pulse rate and ITD ENV held constant at either value, all ITD FS values can be averaged to increase the power of the test. Even for subjects showing dependence of ITD FS on LD at this pulse rate, the marginalization of ITD FS is justified, based on the finding of independence between ITD ENV and ITD FS. Thus, log-linear models including the factors subject, response and ITD ENV for results collapsed over ITD FS were fitted to the data. For both CI and NH listeners only the saturated model could be fitted, showing strong intersubject variability ($p<0.001$ for the three-way interaction term). Thus, the NH and CI listener groups were analyzed separately. For the NH listener group the model with the marginalized factor subject fits well ($G^2=3.432$, $df=12$, $p=0.992$), revealing homogeneity of subjects within this group and high sensitivity to ITD ENV ($p<0.001$). For the CI listener group the model including interactions [subject × response] and [ITD ENV × response] gave the best fit ($G^2=6.121$, $df=12$, $p=0.41$) showing a significant overall sensitivity to ITD ENV of the group ($p=0.031$); however, a significant deviation of the performance of listeners CI1 ($p=0.039$) and CI2 ($p=0.032$) from the group of CI subjects could be found. Therefore the ITD ENV sensitivity of the CI listeners was analyzed separately in a contingency table analysis, revealing a significant sensitivity for listener CI1 ($p=0.007$) and no sensitivity for the rest ($p>0.1$ for CI2, CI3, CI8) at 400 pps.

Finally, the effect of ITD ENV (0, 625, 800 μs) on LD was investigated with pulse rate as parameter. The three-way interaction parameters of saturated log-linear models including the factors pulse rate, ITD ENV, and subjects' response for each subject were examined. An interaction between ITD ENV and pulse rate could be found for CI3: increasing the pulse rate within the range of values tested or raising the ITD ENV increased the odds ratio by the factor of 1.129 ($p=0.011$), indicating a greater sensitivity to ITD ENV with increasing pulse rate. For the NH subjects, sensitivity was independent of the pulse rate ($p=0.66$) but strongly associated with ITD ENV (raising the ITD ENV from 0 to 400 μs or from 400 to 625 μs increased the odds ratio by the factor of 1.558, $p<0.001$).

IV. DISCUSSION

The results of this study show that the tested CI listeners are sensitive to ITD in the fine structure for pulse rates up to 800 pps, depending on the individual listener. The NH subjects listening to a simulation of electric stimulation showed sensitivity up to 600 pps for the same conditions. This is qualitatively consistent with the results of Laback *et al.* (2005) who used unmodulated trains with four pulses and found sensitivity to ITD FS up to 800 pps in two out of three CI listeners and up to 400 pps in the NH listeners, depending on the individual listener. Furthermore, in both studies the

data of the NH subjects show little intersubject variability, as opposed to the results of the CI listeners, who in this and most other studies show a wide range of performance. Such strong intersubject variability implies that at least one factor influencing sensitivity to fine structure ITD has not been taken into account in this study and furthermore that statistical evaluation of CI listeners as a group may yield misleading conclusions.

Recovery from forward masking (Chatterjee, 1999; Nelson and Donaldson, 2001; Nelson and Donaldson, 2002) could be an explanation for the large subject dependence among CI listeners. Forward masking in CI listeners may imply longer decay in the excitation pattern, which would smear the fine structure resulting in a lower sensitivity to fine structure ITD.

The strong intersubject variability, combined with the small number of subjects, limits the interpretation of the results to case studies. It is interesting that the better performing CI listeners (CI3, CI8) had 1–2 months of bilateral experience, while the worse performers (CI1, CI2) had years of bilateral experience. One may be tempted to hypothesize that CI1 and CI2 lost their ability to utilize fine structure cues because of the longer period of stimulation with signals that are uncorrelated in the fine structure. We had the opportunity to test CI8 one year after the main tests under similar conditions. In contrast to the hypothesis, the performance of CI8 was not significantly different than the previous results, which suggests that either the time constant of the unlearning process is much longer than one year or the origins underlying the individual differences in performance are much more complex. Actually, testing one subject to validate a hypothesis based on evidence derived from two groups of two subjects appears to be very speculative. Thus, more extensive tests with more subjects are required to validate this hypothesis.

There are at least two possible explanations for the higher maximum rate showing significant effects of ITD FS on LD in the better performing CI listeners compared to the NH listeners. First, one of the limiting parameters in acoustic hearing could be the smearing of the temporal information by auditory filtering in the cochlea. Filtering the acoustic signals with a simulation of the auditory filter (center frequency: 4590 Hz) shows that by increasing the pulse rate, the modulation depth of the stimuli decreases, but is still present for pulse rate as high as 938 pps, which was the highest pulse rate used in the experiments with the NH listeners. In electric hearing, the auditory filters are bypassed. Second, in electric hearing, the degree of neural phase locking is known to be stronger than in acoustic hearing due to bypassing the synaptic mechanism at the hair cell (Abbas, 1993). An appropriate characteristic of both effects with respect to the pulse rate might lead to higher ITD FS sensitivity in electric hearing at higher pulse rates, and may account for these results.

ITD FS was varied in the range between zero and IPI, which corresponds to a setup of unsynchronized speech processors, in which ITD FS varies periodically between 0 and IPI. The results obtained for these stimuli show that ITD FS can cause a lateral shift in the perceived position up to a pulse rate of 800 pps (CI listeners) and 600 pps (NH subjects). Therefore, to control the lateral position of the auditory image, the fine structure of the stimulus should be encoded for stimulation pulse rates up to 800 pps.

Comparing the subjects' sensitivity to ITD ENV at 400 pps, three out of four CI listeners showed no sensitivity, as opposed to the results of CI1 and the NH listeners. It appears to be contradictory that CI listeners, who performed comparably to NH listeners with respect to ITD FS, showed much worse sensitivity to ITD ENV. One possible explanation for that is the effect of the amplitude modulation shape. In our study the trapezoidal modulation was a compromise of providing strong envelope and fine structure cues. As an example, the rectangular modulation is expected to provide a stronger ITD ENV cue, but, it allows ITD ENV values in integer multiples of the IPI only and therefore, it is not adequate for this study. On the other hand, it is expected that extending ramps beyond 20 ms results in, besides a higher ITD ENV resolution, an attenuation of the onset effects in each trapezoid. Furthermore, by applying different ramps or changing the duty factor, the amount of information in the fine structure changes, which has an effect on ITD FS sensitivity. A "nice" alternative may be a speech-shaped pulse train, providing information on lateralization discrimination sensitivity to ITD ENV for real-world stimuli.

Considering all pulse rates tested, one subject (CI3) showed a consistent improvement of sensitivity to ITD ENV with increasing pulse rate. The NH listeners showed a positive effect of ITD ENV but no significant effect of pulse rate. This is in agreement with the results of Henning (1974) showing no monotonous effect of rate. In general, the sensitivity to fine structure ITD was higher than to envelope ITD for all subjects in this study.

In the real world, most stimuli carry coherent ITD information in both the fine structure and envelope corresponding to the waveform delay condition (WD) tested in this study. A comparison of the WD and ENV conditions is important with respect to practical applications. These results show that the WD condition results in better LD for pulse rates up to 400 pps (CI listeners) and 600 pps (NH subjects) relative to condition ENV. It was also shown that for the combination of higher pulse rates and higher ITD values, the WD condition leads to a deterioration of LD as a result of ITD FS cues pointing to the wrong side and ITD ENV cues pointing to the correct side. To avoid this negative effect an optimized WD condition called *diminished waveform delay* (WD_{DIM}) was introduced, in which ITD FS was limited to 0.25 IPI. Using WD_{DIM} resulted in an improvement of LD relative to WD for pulse rates up to 800 pps for CI listeners CI3 and CI8. There are practical constraints with regard to implementing the WD_{DIM} rule in bilateral CI systems. Whether to use WD_{DIM} or not should be based on the pulse rate applied in the stimulation strategy: if ITD values greater than 0.25 IPI are expected, WD_{DIM} will improve lateralization discrimination. The efficacy of WD_{DIM} is restricted to pulse rates with sensitivity to ITD FS, i.e., to pulse rates up to 800 pps only. Furthermore, WD_{DIM} requires a bilateral processor with the ability to extract and control ITD cues in the envelope and fine structure, which may be difficult to implement.

There is also another way to provide ITD FS cues to CI listeners: coding the temporal information in the envelope of a very high pulse rate carrier (several thousands of pps) such as "HiRes" (Wilson, 2004). HiRes was investigated in several monaural studies (e.g., Frijns *et al.*, 2002; Filipo *et al.*, 2004; Bosco *et al.*, 2005), showing some improvements, particularly a better speech recognition in noise, compared to pulse rates in the region of 1500 pps. However, it is difficult to interpret these results in the context of bilateral stimulation and effects on fine structure ITD sensitivity.

Improving ITD FS perception requires using lower pulse rates, which may influence the performance with respect to monaural speech perception in quiet. Several studies have compared speech intelligibility performance by varying the pulse rate. For example Fu and Shannan (2000) and van Hoesel (2002) tested different pulse rates, however they did not consider listener accommodation to a new stimulation rate. In contrast, Vandali *et al.* (2000), Holden *et al.* (2002), and Galvin and Fu (2005) tested different pulse rates and did consider listener accommodation. In general these studies provide no contraindication to using pulse rates as low as 250 pps with respect to speech intelligibility in stimulation strategies optimized for ITD FS coding.

One strategy which encodes timing information in fine structure is peak derived timing (PDT) introduced by van Hoesel and Tyler (2003). PDT takes into consideration the fine structure of acoustic signals and provides electric signals similar to the WD condition in this paper. In the PDT strategy, the temporal position of an acoustic peak in a subband is determined and an electric pulse is applied to the corresponding electrode at the corresponding time. As a consequence, the pulse rate varies according to the temporal properties of the acoustic signal at each channel and was limited to a maximum of 1400 pps. Van Hoesel and Tyler could not find any clear difference between the PDT strategy and the standard clinical stimulation strategy with respect to sound localization and speech perception in noise. Unfortunately, the comparison between the two strategies was confounded by differences in the experimental setup such as automatic gain control, dynamic range, and number of electrodes. Hence, more detailed investigations into the efficiency of encoding fine structure timing information with various strategies are required to determine the actual extent of lateralization improvement for CI listeners.

V. CONCLUSIONS

This study shows that CI listeners are able to lateralize stimuli using interaural time differences in the fine structure only, up to pulse rates as high as 800 pps. This may affect the lateralization of sounds using speech processors which do not consider the synchronization of the fine structure. Three different synchronization conditions were introduced and tested with four CI listeners, indicating some possible constraints for future stimulation strategies to take greater advantage of the interaural time difference information in the fine structure and envelope.

ACKNOWLEDGMENTS

We are indebted to our test persons, in particular the CI listeners, for their patience while performing the longsome tests. We thank MED-EL Corporation for providing the equipment for direct electric stimulation. We are grateful to Peter Nopp for fruitful discussions and Brian Gygi for helpful comments on an earlier version of this document. This study was supported by the Austrian Academy of Sciences.

[1]Lateralization to the "wrong" side was possible due to ambiguous ITD information in the fine structure in cases where the ITD exceeded 0.5 IPI.
[2]Note that the lateralization data could not be modeled by means of JNDs since the functions are nonmonotonic.
[3]The odds ratio is the ratio of the probabilities of obtaining a correct response for one condition compared to another condition. An odds ratio of 1 shows that there is no difference in the correct responses between the two conditions. The significance of the difference between the two conditions can be calculated using the logarithmic odds ratios and their corresponding confidence intervals. In multi-factorial design, logarithmic odds ratios directly correspond to the interaction terms in the log-linear models allowing investigation of interactions between factors.
[4]Agresti (1984) uses the term "to collapse" to describe the process of averaging a factor in a data pool. This process is also known as marginalization by a factor.
[5]According to Agresti (1984) a saturated model contains all interactions of all factors. In this model, no factor may be averaged and the data must be analyzed for each level of each factor separately.

Abbas, P. J. (**1993**). "Electrophysiology," in *Cochlear Implants: Audiological Foundations*, edited by R. S. Tyler (Singular, San Diego).

Agresti, A. (**1984**). *Analysis of Ordinal Categorical Data* (Wiley, New York).

Agresti, A. (**1996**). *Introduction to Analysis of Categorical Data* (Wiley-Interscience, New York).

Bernstein, L. R. (**2001**). "Auditory processing of interaural timing information: New Insights," J. Neurosci. Res. **66**, 1035–1046.

Blauert, J. (**1997**). *Spatial Hearing*, 2nd ed., (MIT, Cambridge, MA).

Bosco, E., D'Agosta, L., Mancini, P., Traisci, G., D'Elia, C., and Filipo, R. (**2005**). "Speech perception results in children implanted with Clarion devices: Hi-Resolution and Standard Resolution modes," Acta Oto-Laryngol. **125**, 148–158.

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.

Busby, P. A., Whitford, L. A., Blamey, P. J., Richardson, L. M., and Clark, G. M. (**1994**). "Pitch perception for different modes of stimulation using the cochlear multiple-electrode prosthesis," J. Acoust. Soc. Am. **95**, 2658–2669.

Carlyon, R. P., van Wieringen, A., Long, C. J., Deeks, J. M., and Wouters, J. (**2002**). "Temporal pitch mechanisms in acoustic and electric hearing," J. Acoust. Soc. Am. **112**, 621–633.

Chatterjee, M. (**1999**). "Temporal mechanisms underlying recovery from forward masking in multielectrode-implant listeners," J. Acoust. Soc. Am. **105**, 1853–1863.

Collins, L. M., Zwolan, T. A., and Wakefield, G. H. (**1997**). "Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects," J. Acoust. Soc. Am. **101**, 440–455.

Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**, 1057–1065.

Dirks, D. D., and Wilson, R. H. (**1969**). "The effect of spatially separated sound sources on speech intelligibility," J. Speech Hear. Res. **12**, 5–38.

Drennan, W. R., Gatehouse, S., and Lever, C. (**2003**). "Perceptual segregation of competing speech sounds: the role of spatial location," J. Acoust. Soc. Am. **114**, 2178–2189.

Filipo, R., Mancini, P., Ballantyne, D., Bosco, E., and D'Elia, C. (**2004**). "Short-term study of the effect of speech coding strategy on the auditory performance of pre- and post-lingually deafened adults implanted with the Clarion CII," Acta Oto-Laryngol. **124**, 368–370.

Frijns, J. H., Briaire, J. J., de Laat, J. A., and Grote, J. J. (**2002**). "Initial evaluation of the Clarion CII cochlear implant: Speech perception and neural response imaging," Ear Hear. **23**, 184–197.

Fu, Q. J., and Shannon, R. V. (**2000**). "Effect of stimulation rate on phoneme recognition by nucleus-22 cochlear implant listeners," J. Acoust. Soc. Am. **107**, 589–597.

Galvin, J., and Fu, Q. J. (**2005**). "Effects of stimulation rate, mode, and level on modulation detection by cochlear implant users," presented at the 28th Midwinter Research Meeting of the Association for Research in Otolaryngology, New Orleans.

Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (**2003**). "Temporal properties of spontaneous speech—A syllable-centric perspective," J. Phonetics **31**, 465–485.

Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). "Speech intelligibility and localization in a multi-source environment," J. Acoust. Soc. Am. **105**, 3436–3448.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.

Hirsh, I. J. (**1950**). "The relation between localization and intelligibility," J. Acoust. Soc. Am. **22**, 196–200.

Holden, L. K., Skinner, M. W., Holden, T. A., and Demorest, M. E. (**2002**). "Effects of stimulation rate with the Nucleus 24 ACE speech coding strategy," Ear Hear. **23**, 463–476.

Laback, B., Majdak, P., and Baumgartner, W. (**2005**). "Interaural time differences in temporal fine structure, onset, and offset in bilateral electrical hearing," poster presented at the 28th Midwinter Research Meeting of the Association for Research in Otolaryngology, New Orleans.

Laback, B., Pok, S. M., Baumgartner, W. D., Deutsch, W. A., and Schmid, K. (**2004**). "Sensitivity to interaural level and envelope time differences of two bilateral cochlear implant listeners using clinical sound processors," Ear Hear. **25**, 488–500.

Lawson, D. T., Wilson, B. S., Zerbi, M., van den Honert, C., Finley, C. C., Farmer, J. C., Jr., McElveen, J. T., Jr., and Roush, P. A. (**1998**). "Bilateral cochlear implants controlled by a single speech processor," Aust. Hosp. **19**, 758–761.

Licklider, J. C. R. (**1948**). "The influence of interaural phase relations upon the masking of speech by white noise," J. Acoust. Soc. Am. **20**, 150–159.

Lienert, G. A. (**1978**). *Verteilungsfreie Methoden in der Biostatistik (Non-Parametric Methods in Biostatistics)*, 2nd ed. (Anton Hain, Meisenheim am Glan).

McKay, C. M., and Carlyon, R. P. (**1999**). "Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains," J. Acoust. Soc. Am. **105**, 347–357.

Nelson, D. A., and Donaldson, G. S. (**2001**). "Psychophysical recovery from single-pulse forward masking in electric hearing," J. Acoust. Soc. Am. **109**, 2921–2933.

Nelson, D. A., and Donaldson, G. S. (**2002**). "Psychophysical recovery from pulse-train forward masking in electric hearing," J. Acoust. Soc. Am. **112**, 2932–2947.

R Development Core Team (**2004**). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Smith, Z. M., and Delgutte, B. (**2005**). "Binaural interactions in the auditory midbrains with bilateral cochlear implants," presented at the 28th Midwinter Research Meeting of the Association for Research in Otolaryngology, New Orleans.

van Hoesel, R. J. M., and Clark, G. M. (**1997**). "Psychophysical studies with two binaural cochlear implant subjects," J. Acoust. Soc. Am. **102**, 495–507.

van Hoesel, R., Ramsden, R., and O'Driscoll, M. (**2002**). "Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user," Ear Hear. **23**, 137–149.

van Hoesel, R. J. M., and Tyler, R. S. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.

Vandali, A. E. (**2000**). "Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system," Ear Hear. **21**, 608–624.

Wightman, F. L., and Kistler, D. L. (**1997**). "Factors affecting the relative salience of sound localization cues," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum Associates, Mahwah, NJ).

Wilson, B. S. (**2004**). "Engineering design of cochlear implants," in *Cochlear Implants*, edited by F. G. Zeng, A. N. Popper, and R. R. Fay (Springer, New York).

Zwislocki, J., and Feldman, R. S. (**1956**). "Just noticeable differences in dichotic phase," J. Acoust. Soc. Am. **28**, 860–864.

Fast head-related transfer function measurement via reciprocity

Dmitry N. Zotkin,[a] Ramani Duraiswami,[b] Elena Grassi,[c] and Nail A. Gumerov[d]
Perceptual Interfaces and Reality Laboratory, Institute for Advanced Computer Studies (UMIACS), University of Maryland at College Park, College Park, MD 20742

(Received 6 October 2004; revised 21 April 2006; accepted 3 May 2006)

An efficient method for head-related transfer function (HRTF) measurement is presented. By applying the acoustical principle of reciprocity, one can swap the speaker and the microphone positions in the traditional (direct) HRTF measurement setup, that is, insert a microspeaker into the subject's ear and position several microphones around the subject, enabling simultaneous HRTF acquisition at all microphone positions. The setup used for reciprocal HRTF measurement is described, and the obtained HRTFs are compared with the analytical solution for a sound-hard sphere and with KEMAR manikin HRTF obtained by the direct method. The reciprocally measured sphere HRTF agrees well with the analytical solution. The reciprocally measured and the directly measured KEMAR HRTFs are not exactly identical but agree well in spectrum shape and feature positions. To evaluate if the observed differences are significant, an auditory localization model based on work by J. C. Middlebrooks [J. Acoust. Soc. Am. **92**, 2607–2624 (1992)] was used to predict where a virtual sound source synthesized with the reciprocally measured HRTF would be localized if the directly measured HRTF were used for the localization. It was found that the predicted localization direction generally lies close to the measurement direction, indicating that the HRTFs obtained via the two methods are in good agreement.
 [DOI: 10.1121/1.2207578]

PACS number(s): 43.66.Qp, 43.66.Pn [AK] Pages: 2202–2215

I. INTRODUCTION

Spatial hearing is one of the fundamental sensory abilities encountered in the animal kingdom. Humans are very good at localizing acoustic sources in the environment, and some animals (such as the barn owl) are even better localizers. It is known (Hartmann, 1999) that various cues participate in spatial sound perception and that some of these cues are created by the process of scattering of the sound off the listeners themselves. Because of such scattering, the signals arriving at our ears are modified in a direction-dependent manner (Batteau, 1967; Wright *et al.*, 1974; Musicant and Butler, 1984).

The scattering process can be modeled as a linear filter applied to the sound emanating from the source. The transfer function, which when applied to the Fourier transform of the source sound converts it to the Fourier transform of the received sound, is called the "head-related transfer function (HRTF)." It captures the scattering behavior of the ear, head, and body of the listener for various locations at which acoustical sources may be. If the head is centered at a given point P and the sound source located at elevation θ, azimuth φ, and distance r in a head-centered spherical coordinate system, then the HRTF $H(r,\theta,\varphi,f)$ is the ratio of the Fourier transform of the signal at the eardrum $F_e(f)$ to the Fourier transform of the signal that would have been received at the point P in free-field $F_P(f)$, where f is the signal frequency. Thus the HRTF depends on four variables and must be measured as a function of these. For relatively distant sources, $r\rightarrow\infty$, the dependence on r is weak and is often ignored. In this case the HRTF may be characterized as a function of just three variables $H(\theta,\varphi,f)$. This assumption has shown to not hold for relatively nearby sources, and some researchers have measured the HRTF at different distances.

The head-related impulse response (HRIR) is the impulse response of the ear-head-body (i.e., the inverse Fourier transform of the HRTF). The HRIR (and therefore the HRTF) is traditionally measured by playing a broad-band test signal from various positions and recording the scattered wave field either at the entrance of the open or blocked ear canal or in the ear canal itself near the eardrum (Shaw and Teranishi, 1968). If the HRTF is treated as a four-dimensional function, then the source should be moved in three spatial dimensions, and the space of locations be sampled. If the dependence on range is neglected, then the source is placed at various locations on the surface of a sphere of radius a, and the HRTF is measured as a function of frequency and direction $H(\theta,\varphi,f)$. This method, characterized by the fact that the source is moved to various locations, is later referred to as the *direct HRTF measurement method*. In this paper, an alternative *reciprocal HRTF measurement method* is proposed, and comparison of the reciprocally measured HRTF with analytically derived HRTF for a sound-hard sphere and comparison between reciprocally and directly measured KEMAR manikin HRTFs are presented.

The paper is organized as follows. In Sec. II, the theoretical background for the method is discussed. In Sec. III, the proposed HRTF measurement method is described and its

[a]Electronic mail: dz@umiacs.umd.edu
[b]Electronic mail: ramani@umiacs.umd.edu Also at Department of Computer Science, University of Maryland, College Park, MD.
[c]Electronic mail: egrassi@umd.edu
[d]Electronic mail: gumerov@umiacs.umd.edu

advantages, limitations, and methods of overcoming these are discussed. In Sec. IV, the apparatus and the signal processing methods for the reciprocal HRTF measurement technique are described. In Sec. V, experimental results for a sphere (modeled using a bowling ball) and for the KEMAR manikin are presented. Section VI concludes the paper.

II. BACKGROUND

A. Existing HRTF measurement methods

Because of interpersonal differences in anatomy, the HRTF varies widely between individuals. Many studies have shown that with individualized HRTF virtual sound sources are indistinguishable from the real ones (Zahorik *et al.*, 1995; Hartmann and Wittenberg, 1996; Kulkarni and Colburn, 1998; Langendijk and Bronkhorst, 2000) and that the localization performance, particularly in the vertical dimension (in elevation), is degraded when a generic (nonindividualized) or distorted HRTF is used (Gardner and Gardner, 1973; Wenzel *et al.*, 1993).

Various approaches exist for HRTF personalization. Traditionally, the far-field ($r \to \infty$) HRTF for a given person is measured by placing miniature microphones or microphone probes into a person's ears and playing (from a relatively large distance) test sounds from each direction for which the HRTF is to be measured. Many papers are devoted to the topic of human HRTF measurement using the direct method (e.g., Shaw and Teranishi, 1968; Blauert, 1969; Mehrgardt and Mellert, 1977; Wightman and Kistler, 1989; Pralong and Carlile, 1994) or include HRTF measurement performed as a part of a larger experiment (Divenyi and Oliver, 1989; Bronkhorst, 1995). Among existing methods of HRTF personalization, the direct measurement procedure is the most accurate and is well developed. Various test signals, such as impulses, white noise, ML sequences (Schroeder, 1979), Golay codes (Zhou *et al.*, 1992), or in fact any broad-band signals with sufficient energy in the frequencies of interest (Grassi *et al.*, 2003), can be used for the measurement. Also, the measurement can be performed with either open or blocked (sealed) ear canals (Møller *et al.*, 1995; Carlile, 1996). In the first case, a miniature microphone or a probe microphone tip is placed inside the ear canal and the measurement includes the ear canal response; in the second case, the ear canal is sealed and a microphone is placed at the entrance to the ear canal, usually in the center of the sealing plug. With a blocked ear canal recording, also called "blocked-meatus" recording, the ear canal response is missing from the measured HRTF. However, in the most common scenario in which HRTFs are used—in a virtual auditory display where sound is delivered through headphones of sufficiently low acoustic impedance—the ear canal response is naturally reintroduced. Because of this reason and due to some dependence of the open-ear-canal measurement on precise microphone placement (Middlebrooks *et al.*, 1989), blocked ear canal measurement is usually performed. It has been shown to capture all location-dependent characteristics of the HRTF by Algazi *et al.* (1999).

In the experimental setup for the direct method numerous choices are involved regarding the type and the length of the signal (which determines the lowest reliably measured frequency and the required reverberant properties of the measurement space), the number of signal repetitions and the level of the room acoustic isolation [which determines the signal-to-noise ratio (SNR) and protects against outliers and incorrect HRTF measurement], the number of simultaneously mounted loudspeakers (which determines the amount of equipment reflections), the spatial HRTF sampling density (which obviously determines the measurement time period length), the equipment cost (which also influences SNR through the quality of the equipment and the availability of an acoustically treated space), and others. However, in any direct measurement setup it is necessary to move the sound emitter (usually, a loudspeaker) between measurement positions, and generally it takes from tens of minutes to 1–2 h to obtain the HRTF for about 300 to 1200 measurement directions. With conservative choices from the list above, HRTF measurement on a dense spatial grid in a regular room with 25 repetitions of the test pulse can take approximately 1.5 h for a single subject [as indicated by the authors of a carefully measured and validated extensive HRTF database (Algazi *et al.*, 2001c) in private communication].

When human listeners are tested in anechoic settings or in virtual presentation, the just noticeable differences (jnd's) in the azimuth vary from about 1° in the region in the front of the listener to 5°–10° to the extreme right and to the extreme left of the listener, whereas the jnd's in elevation are more signal dependent and can vary from 4° (white noise) to 17° (continuous speech by an unfamiliar person) (Blauert, 1997; Carlile *et al.*, 1997). These values and the fact that the precise locations of regions of higher sensitivity vary for subjects suggest that a fine HRTF sampling grid with a large number of points is necessary for applications where the location of sources must be presented with high fidelity. The use of multi-loudspeaker arrays, one per desired position, can introduce inter-reflections between the loudspeakers and contaminate the measurement signal. By using certain tradeoffs in the measurement process and by developing specialized equipment, different laboratories have made significant speed-ups in HRTF measurement. However, despite these advances, the direct measurement method has the fundamental limitation that measurements for different source positions must be made *sequentially* (i.e., the test sounds are emitted one after the other, with some waiting time in between, from the different positions where the HRTF sampling is to be performed).

In this paper, a novel HRTF measurement method based on the reciprocity principle is described and validated. It is based on a method first disclosed by Duraiswami and Gumerov (2003). The method retains most advantages of the direct method while having the potential to significantly decrease the HRTF acquisition time. Once the HRTF is measured with the reciprocal method, it can be used in the regular manner in place of the directly measured HRTF in any application (e.g., in virtual auditory displays).

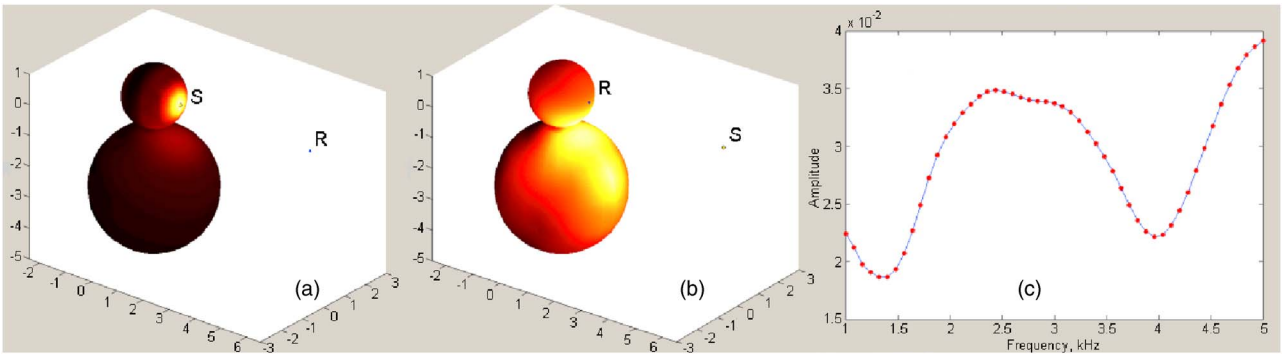


FIG. 1. (Color online) Reciprocity illustration (the acoustic fields are computed with numerical methods). (a) Case A: sound source S in the ear, receiver R at a distance of 4 m. (b) Case B: swapped source and receiver. (c) Dots: intensity versus frequency at the receiver for case (a). Solid line: intensity versus frequency at the receiver for case (b). While the fields in the domain in (a) and (b) are different, reciprocity holds between S and R, as shown in (c).

Other methods for the quick synthesis of a personalized HRTF include frequency scaling in accordance with positions of HRTF features, which were found to be highly correlated with the anthropometry of the subject (Middlebrooks, 1999a, b); use of a morphologically driven parametric HRTF model derived from a large HRTF database (Jin *et al.*, 2000); personalization based on the subjective experience of the user (Runkle *et al.*, 2000); or numerical modeling of acoustic wave propagation (Kahana and Nelson, 2000; Katz, 2001a, b) on a carefully measured mesh of the subject. These methods cannot yet yield the degree of personalization that is achievable when the HRTF is acquired via the direct measurement process.

B. The principle of reciprocity

The reciprocal HRTF measurement method is based on Helmholtz' principle of reciprocity, which states that in an arbitrary complex linear time-invariant acoustic scene "…the pressure at the measurement point **r**, caused by a source at $\mathbf{r}_0$, is equal to the pressure which would be measured at $\mathbf{r}_0$ if the source were placed at **r**" (Morse and Ingaard, 1968). The acoustic field is generally different elsewhere in the scene, but the signal picked up at the receiver is the same if the receiver and the source locations are interchanged. In Fig. 1, the reciprocity principle illustration computed by numerical methods is shown for a monopole (omnidirectional) source.

C. Analytical HRTF computation at low frequency

Accurate HRTF measurement at low frequencies remains a challenge for all measurement methods, mainly because (a) it is hard to produce significant low-frequency output using typical small loudspeakers, (b) a short signal must be used to window out wall and equipment reflections, and (c) the anechoic properties of the acoustic insulation degrade at low frequencies, leading to poor SNR and poorly attenuated wall reflections (Algazi *et al.*, 2001c). However, it is possible to compute the low-frequency HRTF analytically using a simplified model (Algazi *et al.*, 2002b).

The HRTF is a result of the scattering of the acoustic wave on the person's torso, head, and pinnae. These anatomical parts have different characteristic sizes and influence sound waves of wavelengths comparable to their dimensions. Accordingly, the HRTF can be roughly decomposed into parts influenced primarily by the head and torso and by the pinna (Algazi *et al.*, 2001a). Because of the small size of the pinna, the HRTF at low frequencies (longer wavelengths) is mainly a result of interaction of the sound wave with the torso and the head (Algazi *et al.*, 2001b), and this interaction can be well modeled analytically by approximating the head and torso by two spheres (HAT, head-and-torso, or snowman model). The model was first developed and verified by Algazi *et al.* (2002a) using numerical computational methods, and a simplified model, which is referred to as the HAT model, was proposed by Algazi *et al.* (2002b). By replacing the low-frequency part in the measured HRTF by the HAT model HRTF, compensation for the inaccuracies in HRTF measurement at low frequencies is performed.

The HAT model (Algazi *et al.*, 2002b) represents the low-frequency HRTF in terms of certain gross anatomical features of the listener using a "snowman" model. The model consists of two spheres of radii r_t and r_h, modeling the torso and the head, respectively, which are separated by a certain distance h_n, which models the neck height. The parameters r_t, r_h, and h_n are either measured on the subject directly, or are determined from a photograph of the subject that includes a scale, or are fitted to the absolute delays in the measured HRTF. Two pinnaless "ears" are located on the head of the "snowman" diametrically opposing each other on the interaural axis. Two different algorithms are used to model the sound propagation paths in the model and to compute the HAT model HRTF $H_h(f)$ depending on whether the source is located inside or outside of the torso shadow cone with respect to the given ear, as shown in Fig. 2. Sound from a source outside the torso shadow for the given ear arrives through the direct path and through the "shoulder bounce" path. Sound from a source in the torso shadow region is diffracted around the torso to reach the ear. In addition, if the arrival direction for the direct path, the shoulder bounce, or the around-the-torso path falls in the head shadow region, the head shadow is modeled. The HRTF synthesized by the HAT model is minimum phase.

Cross-fading is used on log-magnitudes of the HAT model HRTF $H_h(f)$ and the measured HRTF $H_m(f)$ to obtain the combined HRTF $H_c(f)$:

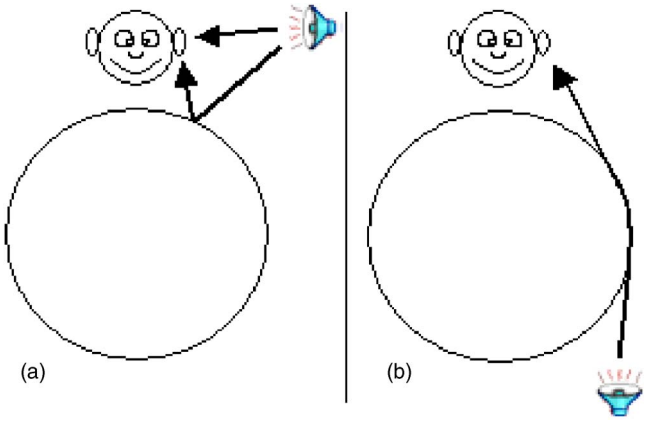


FIG. 2. (Color online) Propagation paths modeled by HAT model: (a) in the case of the source being out of the torso shadow region; (b) in the case of the source being in the torso shadow region.

$$A_c(f) = \begin{cases} A_h(f), & f < f_1, \\ A_h(f) + \dfrac{A_m(f) - A_h(f)}{f_2 - f_1}(f - f_1), & f_1 < f < f_2, \\ A_m(f), & f > f_2, \end{cases}$$

$$A_m(f) = \log|H_m(f)|, \quad A_h(f) = \log|H_h(f)|,$$

$$A_c(f) = \log|H_c(f)|. \tag{1}$$

Thus, the HAT model HRTF magnitude is used for frequencies below f_1; the measured HRTF magnitude is progressively blended in for frequencies from f_1 to f_2; and the measured HRTF magnitude is used for frequencies above f_2. Finally, the phase of the combined HRTF $H_c(f)$ is set to the phase of $H_h(f)$ at all frequencies. Such an approach to modeling HRTF phase is also justified by the results of Kulkarni *et al.* (1999), where it was shown that human listeners are sensitive almost exclusively to the HRTF magnitude.

III. PROPOSED APPROACH

A. Description and advantages

A new method for HRTF measurement is proposed here. It is referred to as the reciprocal method in this paper. The method is based on the reciprocity principle, which has already been used in numerical simulations of problems related to spatial hearing (Kahana, 2000). Application of reciprocity to the HRTF acquisition problem reveals an elegant modification to the HRTF measurement setup, which is to literally exchange the loudspeaker and the microphone, that is, to put a (miniature) loudspeaker in the person's ear and the microphones at the positions where HRTF is to be measured. The reciprocity principle implies that if the loudspeaker and microphone positions were exchanged exactly and no other changes were made in the setup, the measurements obtained would be identical to the ones obtained with the direct method. Inevitable slight microphone, loudspeaker, and subject positioning differences imply that a perfect match between recordings is unlikely to be achieved. Still, according to the reciprocity principle, the HRTF obtained by the reciprocal method should be in reasonable agreement with that measured with the direct method.

The reciprocal measurement method has several advantages over the direct method. First, because in practice microphones are much smaller than loudspeakers, it is possible to build an array of microphones around the person and have one microphone per HRTF measurement direction without artifacts arising due to interequipment reflections. (Arguably, if one uses microloudspeakers instead of regular ones, one can also mount as many of them as the number of the measurement directions without causing significant reflections. However, tiny speakers are expensive, and the measurement process is still sequential, as discussed below.) Second, HRTF sampling for many positions can be done *in parallel* by playing the test sound using an in-ear loudspeaker and recording the received sound simultaneously at all microphones. Thus, the HRTF for all recording positions is acquired at the same time (for one ear), unlike in the direct method. In essence, with an appropriate number of microphones, HRTF acquisition (for two ears) over the whole sphere of directions with the reciprocal method can be done in the same time that it takes for the direct method to perform HRTF acquisition at two positions. In principle, all technological and signal processing innovations that have been used to speed up direct measurements can also be employed in the reciprocal method. Finally, sampling of the HRTF at an additional set of directions can be quickly performed by rotating either the person or the measurement microphone array to a nonoverlapping configuration and performing another pair of measurements.

B. Issues that must be addressed in the reciprocal method

A possible disadvantage of the reciprocal HRTF measurement method is a narrower effective frequency band. The loudspeaker used for reciprocal measurement must obviously fit into the ear canal opening. Therefore, it must be physically small, leading to possibly poor low-frequency output. However, as mentioned above, measurement of the HRTF at low frequencies is always problematic, so this issue is not unique to the reciprocal method, though it is more significant for it. A solution to the problem is offered by augmentation of the measured HRTF by an analytical solution at low frequencies. Such an analytical solution may be obtained by using the HAT model discussed earlier. In fact, the HAT model is a simplified version of the low-frequency HRTF computation technique based on numerical methods, which was developed by Algazi *et al.* (2002a) and could be used as an alternative. No differences in localization performance were observed by Algazi *et al.* (2001b) with the bandlimited (up to 3 kHz) sound source signal when the measured HRTF was replaced by the analytically computed HRTF based only on the spherical-head shadow and torso reflection.

Another possible concern is the safety of the sound level of the in-ear speaker. It is necessary to keep the sound volume at a comfortable and safe level for the subject. An obvious solution is to provide acoustic insulation between the speaker and the eardrum. In the setup used in the experi-

ments, a plug made from a soft material (a silicone compound commonly used by swimmers to seal their ears) is used. The plug is made by fully wrapping the microspeaker in the silicone so that all surfaces of the microspeaker, except the frontal surface where the sound-emitting aperture is located, are covered. In this way, the back (eardrum-facing) side of the microspeaker, where two thin wires are attached, is isolated from the eardrum by the silicone layer, and the plug simultaneously performs the functions of acoustic insulation, blocking the ear canal to achieve a blocked-meatus configuration, and holding the speaker in place.

In human HRTF measurement, it would be also advisable for subject comfort to increase the test signal volume gradually during the first few seconds of the signal to allow gradual habituation of the middle ear mechanisms to the signal level (in particular, adaptation of the amplification controlled by contraction of the stapedius muscle). However, even with such gradual increase the sound volume can be raised no further than certain safety (and comfort) limit for the participant. Because of such a limit on the sound level, a lower SNR than the one achieved in the direct method may be expected. The contralateral side HRTF directions are particularly problematic, and more signal repetitions can be necessary to achieve results of good quality. To identify a safe sound level for the reciprocal method in human subjects, experiments to compare eardrum sound levels observed in the direct and in the reciprocal HRTF measurement methods were carried out using the KEMAR manikin. These results are reported in Sec. V F.

IV. EXPERIMENTAL SETUP

To verify the feasibility of the proposed reciprocal method, a set of experiments corresponding to those reported with the direct method in the literature was carried out, and the results are described in Sec. V. The new measurement method was first evaluated on a sound-hard sphere (bowling ball) and the results compared against the analytical solution obtained by Lord Rayleigh (Strutt, 1904), modified for finite-distance sources by Rabinowitz *et al.* (1993), and presented recently together with experimental verification by Duda and Martens (1998), who also used a bowling ball as a sound-hard sphere model. Then, the HRTF of the KEMAR manikin was measured using both the direct and the reciprocal HRTF measurement methods, and measurements were compared to each other with the help of a sound localization model (based on Middlebrooks, 1992). The experimental setup and experimental methods used to obtain these measurements are described below.

A. Reciprocal method apparatus

The setup used for validation consisted of a spherical mesh structure constructed with parts from the ZomeTool construction kit, which is sold as a toy (shown in Fig. 3). The mesh was made up of struts (sticks) and nodes (balls) as shown in the picture. The total number of nodes in the constructed spherical structure of radius 0.70 m is 131. Thirty-two microphones were mounted at the nodes of the structure in a symmetrical and roughly equispaced manner. (More microphones could have been attached at intermediate locations; the limiting factor in the experiment was the acquisition hardware available at the time the experiments were conducted, and in principle there is no limitation on the number of channels that can be acquired simultaneously.) The microphones were affixed with scotch tape to short struts, which in turn were attached to the nodes. The bottom part of the mesh was left open in order to accommodate a subject. The spherical mesh was suspended from the ceiling with several wires in a large office room ($4.7\times5.6\times2.7$ m^3) with acoustically untreated walls. The mesh was positioned at the center of the room to maximally delay the arrival of interfering wall reflections. The earliest reverberant reflections arrived well after the head response time and were windowed out in the data processing step. The measurements were done mostly in the evening to ensure that noise from neighboring offices was minimized. In addition, the computer used for measurements was placed outside the room during the recording with the microphone connecting cables running under the door, other computers in the room were turned off, and incandescent lighting was used instead of fluorescent lights.

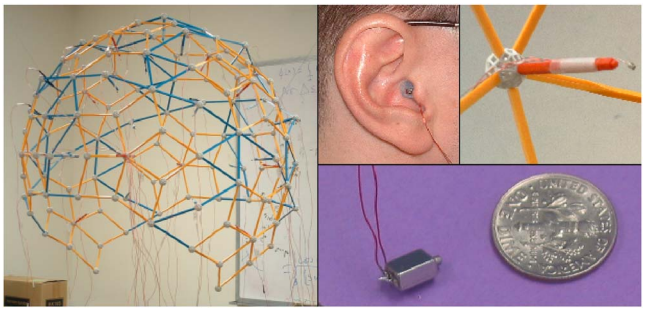


FIG. 3. (Color online) Left: The measurement mesh. Bottom right: The microspeaker. Top middle: The microspeaker inserted into the person's ear. Top right: An enlargement of the one node of the measurement mesh.

To minimize the load on the constructed spherical mesh, thin (gauge 30) enamel-coated wires were used to connect the microphones to two custom-made preamplification cards, each handling 16 channels. The outputs of the preamplifiers were connected to two National Instruments PCI-6071E data acquisition boards plugged into a Pentium Xeon 1.7 GHz Dell Dimension 8100 PC running Windows XP.

The microphone array was calibrated with an active tracking device (Polhemus 3D FASTRAK system). The tracking device has the ability to determine the three-dimensional coordinates of an associated receiver (a small sensor measuring about $12\times12\times6$ mm^3) with respect to a transmitter (a source box measuring about $51\times51\times51$ mm^3, with one corner marked as the origin). To perform the calibration, the transmitter was positioned so that its origin coincided with the center of the measurement sphere and leveled. Then, the coordinates of each of the 32 microphones were measured by putting the sensor next to the corresponding microphone and recording the observed sensor coordinates. The azimuth and the elevation of all microphones with respect to the center of the sphere were then computed from the obtained measurements. The coordinate system adopted for use with the reciprocal measurement pro-

cedure was a standard vertical-polar coordinate system with the azimuth varying from −180° to 180°, with 0° azimuth being in front of the subject and positive values going to the right, and with the elevation varying from −90° to 90°, with 0° elevation being in front of the subject and positive values going up.

B. Direct method apparatus (I)

Two methods were used for direct HRTF acquisition. The first setup was obtained by exchanging the microspeaker and the microphone positions in the reciprocal HRTF measurement setup described above, so that the microphone was inserted in the blocked ear canal of the subject (or in the hole of the sphere) as it is normally done in the direct measurement method, and the microspeaker was attached to the strut that was placed at the node of the spherical mesh structure. The advantage of this method is that the comparison of the direct and the reciprocal measurements performed at the same distance and with the same physical setup can be made. However, the microspeaker needed to be moved manually between measurement positions, slowing down the procedure significantly, and the number of measurement positions was limited in any case by the grid structure. Accordingly, the direct HRTF measurement in this setup was performed only at 32 positions (the same 32 positions at which the reciprocal HRTF was sampled). The HRTF measured in this setup is referred to as a sparse-grid direct HRTF. These measurements were performed at the same distance (0.70 m) as the reciprocal ones.

C. Direct method apparatus (II)

A direct (traditional) HRTF measurement apparatus was also used to obtain the directly measured HRTF with higher spatial resolution, which could not be done on the spherical mesh due to equipment and time constraints. As the focus of the current paper is on the validation of the reciprocal HRTF measurement method, the direct measurement setup and the corresponding measurement procedures are described only briefly here. A more detailed description may be found in Grassi *et al.* (2003). In the direct measurement, a set of loudspeakers (8-Ω Realistic 3 in. midrange tweeter, 700–20 000 Hz), mounted on a semicircular hoop rotating around the horizontal axis, emitted the acoustic signals used in the measurements. The radius of the hoop was 0.90 m, and the HRTF measurement distance (from the speaker membrane to the center of the hoop where the subject is placed) was 0.84 m. To avoid excessive interequipment reflections, only six loudspeakers were placed simultaneously on the hoop and recordings were taken in several sets to cover all desired azimuths. For each configuration of loudspeakers, the hoop stepped through all requested elevations, automatically controlled by the computer. Data collection was performed in a regular-sized office room ($3.2\times3.8\times2.4$ m^3). The room walls were coated with dispersive foam (4.5 cm egg crate foam) to dampen sound reflections. The HRTF was measured at 1149 directions covering the full sphere without the bottom part in approximately 5° steps, forming a dense-grid direct HRTF. The exact arrangement of directions in the direct grid can be found in Grassi *et al.* (2003). For most directions (elevation-azimuth pairs) in the reciprocity grid there was no exactly corresponding direction in the dense direct grid. However, the nearest direction was within 1.6° on average over all 32 reciprocal measurement directions and within 5° at most.

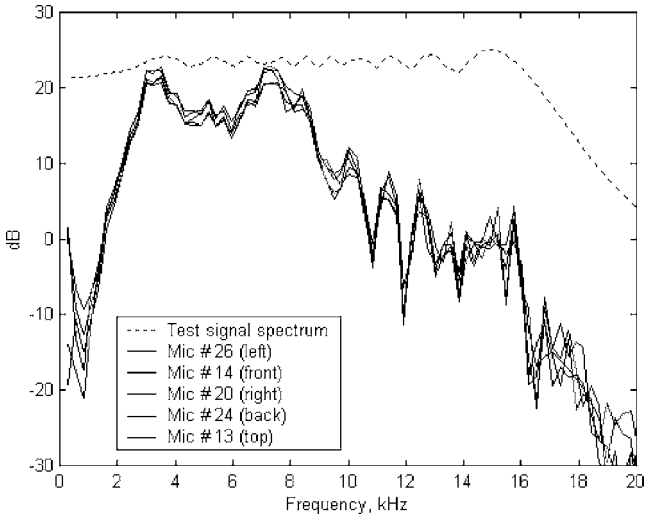


FIG. 4. Comparison between spectra of the test signal recorded at five microphones on the reciprocal measurement mesh.

D. Microphones and microspeaker

The microphones used for the measurements of the HRTF in the reciprocal setup were of type Knowles Electronic FG-3629 (omnidirectional microphone encased in a cylindrical case with a diameter of 2.57 mm and a height of 2.57 mm). The microspeaker used was of type Knowles Electronics ED-9689 (physical dimensions $6.33\times4.31\times2.97$ mm^3). The microspeaker is shown in the bottom right part of Fig. 3. An important question for the HRTF measurement is the microspeaker directivity pattern. An experiment was therefore performed to determine if the microspeaker behaves sufficiently closely to an omnidirectional point source. In the experiment, the same test signal that was later used for reciprocal HRTF measurement was employed. The signal was a frequency sweep with a roughly flat spectrum from 1 kHz up to 16 kHz. The upper frequency limit was chosen to be consistent with upper frequency HRTF measurement limit used by other researchers (Pralong and Carlile, 1994; Bronkhorst, 1995; Langendijk and Bronkhorst, 2000). The microspeaker was placed at the center of the reciprocity measurement apparatus with the speaker opening facing the "front" microphone (the microphone that was located in the front of the subject during HRTF measurement), the test signal was played through the speaker, and the recording was performed simultaneously at 32 microphones. In Fig. 4, the recorded signal spectra are shown for five microphones ("front," "left," "right," "back," and "top") along with the spectrum of the test signal. It can be seen that the microspeaker put out a smaller amount of energy at frequencies above 9 kHz, but the output was sufficient through the frequency band of interest. More importantly, the recordings were close to identical at all five microphones, showing the

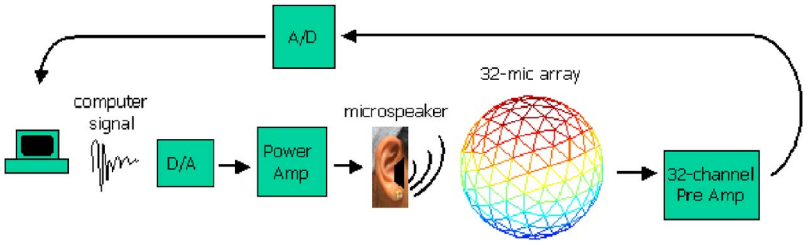


FIG. 5. (Color online) A schematic of the system for the reciprocal HRTF measurement.

omnidirectionality of the microspeaker. Therefore, no equalization to account for a speaker beampattern was necessary.

E. Signal acquisition

The schematic of the experiment and associated signal processing in the reciprocal HRTF measurement is depicted in Fig. 5. The excitation signal was generated in MATLAB, output through the D/A port of a National Instruments PCI-6071E card, power amplified, and played via the microphone inserted in the manikin ear or in a hole in the sound-hard sphere (bowling ball). The manikin/sphere was placed at the center of the spherical mesh. The received signal data were collected simultaneously from all 32 microphones in the spherical mesh. The signal recorded by the microphones was decoupled, amplified, and low-pass filtered with custom preamplifiers (fourth-order Bessel filter, cutoff frequency of 18 kHz, gain 100) to avoid aliasing. Signals were sampled at a rate of 39.0625 kHz and were acquired through the A/D port of the NI card. The D/A channel, which drives the microspeaker, and all A/D channels, which receive input from microphones, were triggered by a common hardware trigger and were running off a common clock source, ensuring that the synchrony necessary for time averaging is achieved.

F. Signal processing and HRTF computation

The test signal for the reciprocal HRTF measurement was a 96-sample-long (2.45 ms at 39.0625 kHz) linear frequency upsweep pulse with a maximum frequency of 16 kHz. The pulse duration was chosen to prevent an overlap of reflections from the experimental equipment and surrounding walls. One test session consisted of an upsweep signal repeated 48 times with a 420-ms pause between repetitions (dictated by an experimentally determined time for reverberant echoes to die out) for a total recording length of about 20 s. At the end of each test session, the manikin/sphere was removed from the spherical mesh and the microspeaker was placed at the center of the mesh on a thin stick in order to measure the reference (calibration) signal for each microphone. In this way any mismatch between microphones or preamplifier channels was captured in the reference signal and was removed during normalization.

The signal processing for the reciprocal HRTF measurement was performed as follows. The raw recorded pulses were averaged coherently in time to decrease the random noise first (the noise magnitude decreases proportionally to the number of pulses). Then the start of the pulse was detected using thresholding and a time window of 144 subsequent samples (3.7 ms) was kept. The measurement window was thus about 1.25 ms longer than the length of the pulse so that additional time was allocated after the end of the pulse to accommodate the system impulse response. The window length was chosen with knowledge of the room geometry and the relative placement of the measurement apparatus to ensure that all wall reflections arrive after the end of the measurement window and thus are discarded.

The HRTF magnitude for a particular direction (θ,φ) was then computed by taking the ratio of the Fourier transform of the signal acquired from the given microphone to the Fourier transform of the reference signal for the same microphone, i.e., $H(\theta,\varphi,f)=F(y(\theta,\varphi,t))/F(u(\theta,\varphi,t))$, where F is a discrete-time Fourier transform operator (i.e., a 144-point FFT), $y(\theta,\varphi,t)$ is the time-averaged pulse recorded at the microphone with the elevation θ and azimuth φ, and $u(\theta,\varphi,t)$ is the time-averaged reference pulse for the same microphone. A resolution of about 270 Hz was thus obtained in the resultant HRTF. Additional frequency smoothing was then performed on the magnitude of the obtained $H(\theta,\varphi,f)$ using a second-order Butterworth filter with cutoff $0.5\times$ normalized frequency.

Sparse-grid direct HRTF acquisition was performed with exactly the same test signal and signal processing procedures as described here for the reciprocal HRTF acquisition. For the dense-grid direct HRTF acquisition, a similar test signal was used and a similar experimental procedure was followed, with the pulse length being 150 samples (1.8 ms at 83.333 kHz) and the window length being 225 samples (2.7 ms). Smoothing was also performed similarly, but with a different normalized cutoff frequency so that the same frequency resolution was achieved in the final computed HRTF.

G. Sphere

A bowling ball was used as a physical model of a sound-hard sphere. The ball (polyester Ebonite Mirage) had a diameter of 21.84 cm, weighed 5.45 kg, and had a hard-coated surface with no finger holes. A 9.5-mm-diameter hole was drilled in the ball to accommodate the microspeaker for the reciprocal measurement experiment. No more holes were drilled. For the reciprocal HRTF measurement experiment, the microspeaker was inserted into the hole together with the silicone plug holding it in place so that the microspeaker was centered in the hole with the sound-emitting aperture facing outwards. The silicone plug surface and the microspeaker in the plug were both flush with the sphere surface.

The sphere was placed at the center of the reciprocal HRTF measurement mesh on a photographic tripod and was centered within the mesh using the mesh structure itself as a visual guide, as described below. To minimize tripod reflec-

tions, the tripod legs were covered with a soft fabric, and all tripod handles located just below the sphere were removed.

H. KEMAR manikin

An object traditionally used as a reference for HRTF measurement is the Knowles Electronics Manikin for Acoustic Research (KEMAR) (Burkhard and Sachs, 1975). Measurements of the HRTF for KEMAR have been performed by several researchers (e.g., Gardner and Martin, 1995; Algazi *et al.*, 2001c) and are widely available to the research community on the Internet. The KEMAR model used in the experiments described in this paper was a DB-4004, configured with two neck rings and a torso. Pinnae used were of type DB-060 and DB-061 (left and right "small" pinna, respectively). They were mounted on the DB-050 ear canal extensions with DB-100 occluded ear simulators (Zwislocki couplers). In both ears, a microphone (Knowles FG-3629) was placed into the opening of the Zwislocki coupler and sealed with the silicone. The signals recorded at these microphones were used to measure "eardrum" sound pressure levels during the reciprocal HRTF measurement.

The direct measurement of the KEMAR HRTF was done in the blocked-meatus setup. The microphone (Knowles FG-3629) was wrapped in a silicone plug and was inserted into the ear canal so that the ear canal was sealed and the microphone was located in the center of the seal, flush with the seal surface. The microspeaker insertion for the reciprocal measurement was done in the same manner. The KEMAR was mounted on a long metal pipe that was screwed into the heavy metal stand at the bottom end and into the mounting aperture on the KEMAR bottom plate at the top end so that any interference created by the mounting hardware was negligible.

I. KEMAR alignment procedures

In the direct measurement apparatus, the following procedure was used to center the KEMAR manikin in the measurement hoop. Two calibrated lasers were placed at the pivotal points of the hoop so that the beam of each laser was pointing towards the opposite pivotal point within a few millimeters of the other laser's aperture. A third laser was placed on the wall in front of the setup and was calibrated so that it was pointing to the 0° azimuth mark of the measurement hoop both when the hoop was at 0° elevation (in front of the subject) and at 180° elevation (at the back of the subject). The fourth laser, placed on the ceiling, projected a beam vertically downwards and was calibrated to point to the premarked center of the setup (the location midway between the hoop pivotal points) on the ground and to the 0° azimuth mark on the hoop when the hoop was at 90° elevation (up). The KEMAR was placed in the setup so that the beams of lasers located at pivotal points produced spots at the ear canal openings and was further aligned so that the spot from the frontal laser was located on the tip of the KEMAR's nose and the spot from the overhead laser was projected on the center of the cross etched on the head of the KEMAR.

In the reciprocal setup, a natural coordinate system for alignment was provided by the regular structure of the mesh itself. Four microphones in the mesh were located directly in the front, in the back, on the left, and on the right of the subject. They were mounted at the centers of four crosses so that it was possible to align the manikin in the center of the mesh in the same manner as for the direct measurement by visual inspection. For example, when viewing the mesh from the front (with the KEMAR inside), the front microphone vertical line of mounting, the KEMAR nose, and the back microphone vertical line of mounting should all lie in the same plane. This condition was easy to check visually. In the same manner the left and right microphones were aligned with the KEMAR ear canals and the height of the mounting was adjusted so that the tip of the manikin's nose was in the equatorial plane of the measurement mesh. The alignment achieved was verified for each measurement set to ensure no deviations from centered position.

J. Augmentation by analytical solution

The validity of measured HRTF is limited approximately to the range from 1.5 to 16 kHz, determined by the frequency limitations of the microspeaker used in the experiments. Therefore, it is necessary to extrapolate the HRTF to reasonable values outside this range (especially at low frequencies) to make them suitable for sound rendering. The following approximations are applied:

(i) The high-frequency end is tapered to zero using a half Blackman window, starting at 16 kHz and reaching zero at 22.05 kHz.
(ii) At low frequencies the HRTF is approximated by the analytical solution obtained for the head-and-torso (HAT) model described in Sec. II C [Eq. (1)] with cutoff frequencies f_1 and f_2 being 1 and 3 kHz, respectively.

Augmentation of the measured HRTF with a HAT model compensates for the weak low-frequency response of the microspeaker and is necessary for accurate reproduction of virtual auditory scenery as most of the signals to be reproduced (e.g., speech or music) contain significant energy at lower frequencies.

V. RESULTS AND DISCUSSION

A. SNR estimation

The reciprocal experiments were conducted in the large office room described above. The SNR was estimated for the reference signal obtained by putting the microspeaker at the center of the reciprocity measurement mesh and was found to be 23.7 dB. After averaging with 48 pulses, the SNR increased to 37.4 dB. As a reference, the SNR achieved in the direct measurement setup was 32.5 dB and was improved to 41.5 dB by averaging.

B. Sphere

The plots presented in Fig. 6 show the results of the reciprocal HRTF measurement of the sphere together with the analytical solution for the sphere (Rabinowitz *et al.*,

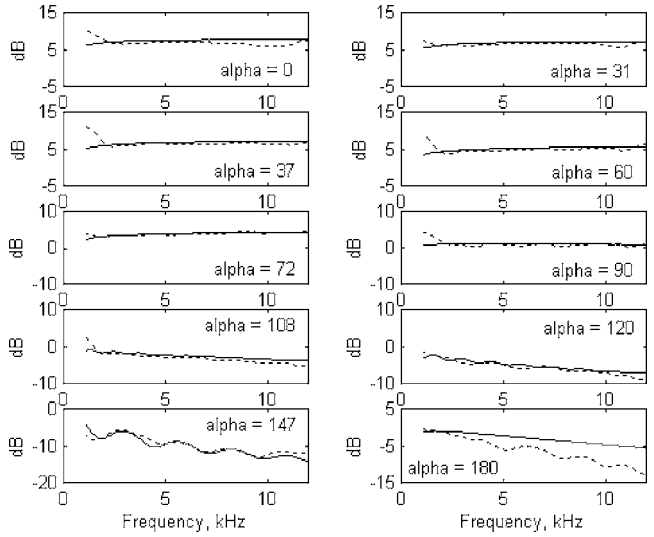


FIG. 6. Comparison between the analytically computed HRFT (solid line) and reciprocally measured HRFT (dotted line) for the sphere. Alpha is the incidence angle.

1993) for ten incident angles (indicated in the plots). The solid line represents the analytical HRTF solution, and the dotted line shows the measured HRTF.

It can be deduced from the plots that the measured HRTF matches the analytical solution with good precision above approximately 1.5 kHz, as can be expected based on the microspeaker effective frequency band. The 6-dB gain at normal incidence is well reproduced, suggesting that the microspeaker is sufficiently small to avoid disturbing the sound field and that the silicone plug surface acts as a sound-hard material (the gain is as expected a little more than 6 dB because the source distance is 0.70 m instead of infinity), and even for the weakest signal presented in the plots (at 147°), the characteristic ridges in the analytical solution are followed very well by the experimental HRTF. There are also no systematic differences between the analytically computed HRTF and the measured HRTF at most positions. However, a significant difference is observed at high frequencies for the 180° incidence angle. The ridge pattern exhibited by the measured HRTF suggests a small angular error in the positioning of the sphere, which brings the measurement position out of the bright spot. All other discrepancies are limited by approximately 2 dB and are similar to the experimental errors observed by Duda and Martens (1998) in the comparison between the sphere HRTF (measured with the direct method) and the analytical HRTF expression for the source at various distances.

C. KEMAR manikin

The second group of plots shown in Fig. 7 includes 32 pairs of direct and reciprocal HRTF measurements of the KEMAR manikin (left ear). The reciprocal measurement was done using the reciprocal method apparatus described above, and all available data (i.e., 32 measurement positions) from one measurement set are presented. The direct measurement was performed in the sparse-grid direct setup described above so that the results presented in this figure were obtained literally by application of the reciprocity principle without changing anything else in the acoustic scene, in the test signal, and in the signal processing. No additional HRTF smoothing was performed beyond the light smoothing described in Sec. IV F. Annotations in each plot show the (elevation, azimuth) pair of the measurement position.

The direct and the reciprocal measurements presented in the plots were done on two different days, and the KEMAR manikin was moved to the different room and then brought back and recentered in the spherical mesh for the second experiment. Nevertheless, the agreement between measurements is quite good within the effective frequency band of the reciprocal HRTF measurement. Positions of HRTF features (such as peak and notches) are matched well in all plots, although at some positions the notch depth differs between the direct and the reciprocal measurement. Also, a level disagreement is observed at the contralateral (0, 90) ("bright spot") position. As in the sphere case, the reason for

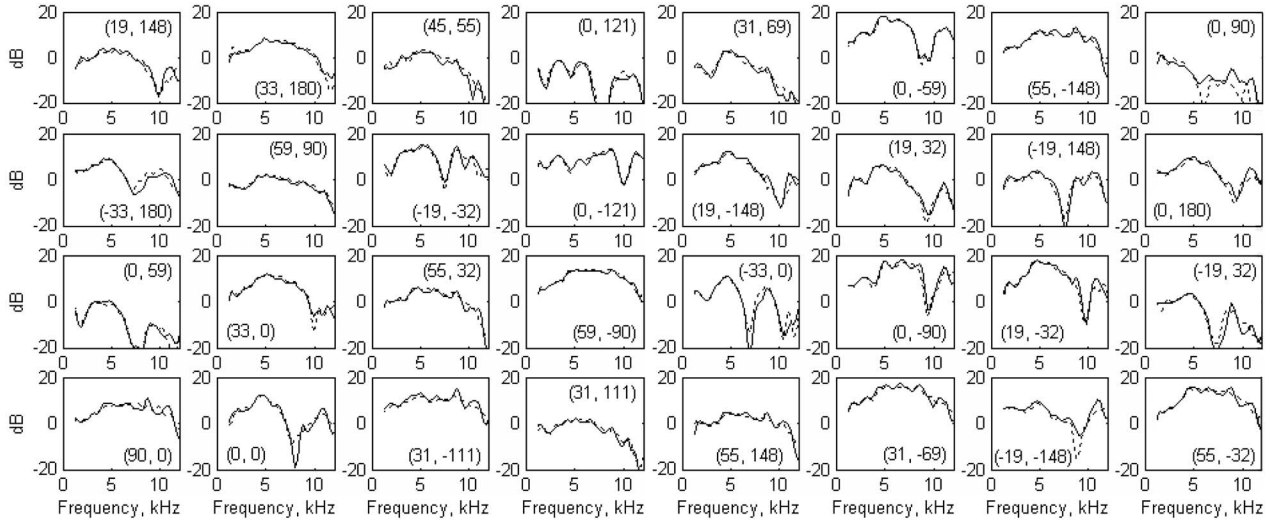


FIG. 7. Comparison between KEMAR left ear HRTF measurements using the direct method at 0.70 m (solid line) and the reciprocal method at 0.70 m (dotted line). The measurement directions are annotated on the plots as (elevation, azimuth) pairs.

the disagreement is likely to be a small positioning error, which moves the measurement position out of the bright spot.

D. Sound localization model

To further validate the reciprocal measurement technique, a simulated sound localization experiment was performed with the help of a localization model that uses both temporal and spectral cues and is based on the narrow-band sound localization model proposed by Middlebrooks (1992). It was chosen to modify Middlebrooks (1992) model somewhat by using the rms difference between HRTF magnitude spectra as the spectral similarity metric, which was used more recently by other researchers to quantify differences between HRTFs (Blommer and Wakefield, 1997; Kulkarni and Colburn, 2004), and by omitting the ILD-based similarity term (due to the unavailability of right ear reciprocal HRTF data). In the experiment, the HRTF obtained with the direct method was taken as the baseline HRTF. Then, for the HRTF measured reciprocally at the direction p, the degree of similarity to the baseline HRTF for each direction q in the baseline HRTF was computed according to the localization model, and the similarity-maximizing direction q' was chosen, corresponding to the sound being "localized" at the direction q'. The validation goal was to have q' at or sufficiently close to p for all p comprising the reciprocally measured HRTF.

Ignoring the range dependence of the HRTF, let the reciprocal HRTF for the direction p be $H_p=H(\theta_p,\varphi_p,f)$ and the direct HRTF for the direction q be $\hat{H}_q=\hat{H}(\theta_q,\varphi_q,f)$. The spectral similarity coefficient $b_{pq}^{(s)}$ was computed for the pair $(H_p,\hat{H}_q)$ as

$$b_{pq}^{(s)}(H_p,\hat{H}_q)=\left[\frac{1}{2\pi(f_2-f_1)}\int_{2\pi f_1}^{2\pi f_2}(20\log_{10}|H(\theta_p,\varphi_p,f)| - 20\log_{10}|\hat{H}(\theta_q,\varphi_q,f)|)^2\,df\right]^{1/2} \quad (2)$$

in the continuous frequency case, or the same with the integral replaced by a summation in the discrete frequency case. For a given reciprocal measurement direction p, the coefficients $b_{pq}^{(s)}$ for all direct measurement directions q were computed. Then, they were normalized to have a zero mean and unit variance, as suggested by Middlebrooks (1992).

It was also desirable to include temporal information in the HRTF comparisons, at least to compensate for the missing ILD term. The natural choice for the temporal cue would be the interaural time difference (ITD). However, ITD information was not immediately available in the reported measurements because the reciprocal KEMAR HRTF measurement was performed only for the left ear. Therefore, an estimation of ITD for reciprocal measurement was constructed by utilizing the symmetry of the setup. The ITD for a given direction $p=(\theta,\varphi)$ is usually computed as the difference in time of arrival (TOA) of the signal to the left and to the right ears. If the subject and the setup are sufficiently symmetric, then the TOA for the right ear for the direction p should be the same as the TOA for the left ear for the direction $p^*=(\theta,-\varphi)$. The ITD τ_p for the reciprocal measurement was therefore constructed as $\tau_p=t_{p^*}-t_p$, where t_p and t_{p^*} were the TOAs (for the left ear) of the signal for the directions p and p^*, respectively. The actual values of the TOAs were determined from the experimental data via thresholding.

Because of the symmetric arrangement of the measurement directions, the measurement grid included the direction p^* for 31 of 32 directions p [e.g., for $(-19,-32)$ the symmetric direction was $(-19,32)$, and some directions such as (33, 180) were symmetric pairs to themselves so the ITD there was forced to be zero]. Only one direction (45, 55) did not have a symmetric counterpart; therefore, t_{p^*} for it was computed as average TOA for two directions closest to $(45, -55)$, which were $(31,-69)$ and $(55,-32)$.

The ITDs $\hat{\tau}_q$ for all positions q in the direct setup were computed using data from both ears (as both sparse-grid and dense-grid direct HRTF measurements were performed on both ears). Temporal similarity coefficients $b_{pq}^{(t)}$ were then computed to be simply

$$b_{pq}^{(t)}=|\tau_p-\hat{\tau}_q|. \quad (3)$$

The coefficients $b_{pq}^{(t)}$ were computed for all directions q and then were normalized to have zero mean and unit variance. After extracting the pure time delay, the remaining phase information in the HRTF was not used for matching. This decision was based on the observation by Kulkarni *et al.* (1999) that humans are largely insensitive to the HRTF phase. Then, finally, the direction $q'=\arg\min_q(b_{pq}^{(s)}+b_{pq}^{(t)})$ was taken to be the localization direction predicted by the model. If q' is at or close to p, then it can be concluded that the reciprocal measurement method provides sufficiently feasible HRTF.

E. Analysis of KEMAR manikin results using the auditory model

The sound localization model was used first with the sparse-grid direct left ear HRTF as the baseline HRTF and with $f_1=1.5$ kHz, $f_2=16$ kHz, which was the validity range of reciprocal HRTF measurement based on the microspeaker characteristics and on the test signal used (see Sec. IV D). For each direction p in the reciprocally measured left ear HRTF, the model was applied to predict the localization direction q' for the sound if it were filtered with the reciprocal HRTF for the direction p. It was found that the matching obtained was, in fact, perfect, and that the model output q' was equal to p for all 32 directions p. For reference, in Table I the values of the spectral similarity coefficient (prenormalization) between reciprocal and direct measurements taken at the same direction p (in other words, rms difference between magnitude spectra of HRTF taken by direct and by reciprocal method) are shown for all p. This table essentially quantifies the data shown in Fig. 7.

However, as the baseline HRTF in this case was sampled only at 32 directions, the perfect matching was perhaps not surprising. For further testing, the sound localization model was applied using the dense-grid direct left ear HRTF as the baseline. For each direction p of the reciprocal HRTF mea-

TABLE I. Root-mean-square difference (RMSD) between the sparse-grid direct and the reciprocal HRTF spectra magnitude shown in Fig. 7 for each measurement direction p. Direction are listed as (elevation, azimuth) pairs in a vertical-polar coordinate system.

p	RMSD (dB)	p	RMSD (dB)	p	RMSD (dB)	p	RMSD (dB)
(19, 148)	2.5	(0, 59)	3.0	(31, 69)	2.8	(−33,0)	3.3
(33, 180)	1.8	(33, 0)	1.9	(0,−59)	1.1	(0,−90)	2.1
(45, 55)	4.4	(55, 32)	2.8	(55,−148)	2.0	(19,−32)	2.1
(0, 121)	3.3	(59,−90)	1.2	(0, 90)	5.0	(−19,32)	2.9
(−33,180)	1.8	(90, 0)	1.9	(19,−148)	2.2	(55, 148)	1.9
(59, 90)	1.5	(0, 0)	3.0	(19, 32)	1.9	(31,−69)	1.7
(−19,−32)	2.4	(31,−111)	2.1	(−19,148)	2.4	(−19,−148)	3.2
(0,−121)	1.0	(31, 111)	2.3	(0, 180)	2.0	(55,−32)	2.2

surement, the predicted localization direction q' from the dense-grid direct HRTF is shown in Table II.

It should be noted here that most directions in the reciprocal grid were not present in the dense direct grid, and for such p the closest possible q' was several degrees off. Also, the localization model was basically "monaural" (operating on left ear data only) so its performance can be expected to be worse than the performance obtained with normal, binaural localization. Finally, the dense-grid direct HRTF was measured here at 0.84 m and the reciprocal HRTF was measured at 0.70 m, which may create slight discrepancies in HRTF. It was shown, among others, by Brungart and Rabinowitz (1999) and by Shinn-Cunningham *et al.* (2000) that the difference in the measurement distances creates changes in the measured HRTF that can not be represented by a simple intensity difference. A first reason for this is the fact that the ear may lie in the head shadow for a given source direction and distance but be out of the shadow for a source in the same direction but at the larger distance. Further, the pinna-related HRTF features may shift due to the parallax effect as the source distance decreases and the difference in source directions with respect to the head center and with respect to the pinna becomes more pronounced.

The results shown in Table II demonstrate that in most cases the predicted localization direction q' fell very close to the actual measurement direction p. The localization error, measured as an angle between vectors p and q', was within 5° for 22 (69%) reciprocal measurement directions, within 10° for 29 (91%) reciprocal measurement directions, and the maximum observed error was 19.5°. The average value of error over 32 measurement directions was 4.7°. Two measurement directions showing the largest errors were located in the overhead-back region of space in the contralateral hemisphere. It is known that in this region the HRTF is relatively featureless and does not vary much with the source direction, which explains the relatively large errors observed there. In fact, a study of errors made by humans in sound localization experiments (Carlile *et al.*, 1997) also suggested larger natural localization errors in the area over the listener and to the back. Because of the head shadowing, the recorded microphone signal for these two directions was also weak and noise-prone. However, even for these two directions the model-predicted localization direction lay on the correct cone of confusion (i.e., p and q' had very close azimuth values in the interaural-polar coordinate system; this is

TABLE II. Comparison between reciprocal HRTF measurement direction p and model-predicted localization direction q' (selected from the dense-grid direct HRTF set) for all 32 p. Angle between p and q' is shown as ang(p,q'). Directions are listed as (elevation, azimuth) pairs in a vertical-polar coordinate system.

No.	p	q'	ang(p,q')	No.	p	q'	ang(p,q')
1	(19, 148)	(20, 148)	1.0	17	(31, 69)	(35, 69)	4.0
2	(33, 180)	(35, 180)	2.0	18	(0,−59)	(0,−55)	4.0
3	(45, 55)	(50, 51)	5.4	19	(55,−148)	(60,−137)	7.7
4	(0, 121)	(5, 120)	5.1	20	(0, 90)	(5, 81)	10.3
5	(−33,180)	(−30,180)	3.0	21	(19,−148)	(20,−153)	4.8
6	(59, 90)	(65, 90)	6.0	22	(19, 32)	(20, 32)	1.0
7	(−19,−32)	(−20,−27)	4.8	23	(−19,148)	(−15,149)	4.1
8	(0,−121)	(0,−120)	1.0	24	(0, 180)	(0, 180)	0.0
9	(0, 59)	(0, 55)	4.0	25	(−33,0)	(−35,0)	2.0
10	(33, 0)	(30, 0)	3.0	26	(0,−90)	(0,−85)	5.0
11	(55, 32)	(60, 31)	5.0	27	(19,−32)	(20,−32)	1.0
12	(59,−90)	(50,−90)	9.0	28	(−19,32)	(−15,31)	4.1
13	(90, 0)	(90, 0)	0.0	29	(55, 148)	(70, 130)	17.0
14	(0, 0)	(0, 0)	0.0	30	(31,−69)	(30,−71)	2.0
15	(31,−111)	(30,−109)	2.0	31	(−19,−148)	(−15,−154)	7.0
16	(31, 111)	(25, 90)	19.5	32	(55,−32)	(50,−32)	5.0

not immediately obvious from Table II as angles in Table II are listed in vertical-polar coordinate system).

The fact that the HRTF measurements obtained at two ranges (0.70 m for the reciprocal setup and 0.84 m for the direct setup) compared well could have been expected from the work of Brungart and Rabinowitz (1999), where the effects of source distance on measured HRTF were studied. Only subtle differences were noticed when the HRTF measurements at 1.0 m and at 0.50 m were compared. The difference between the far-field HRTF and the HRTF taken at 0.70 m can be expected to be even smaller.

The simulated localization experiment was also performed using the sound localization model operating only with the spectral similarity coefficient given by Eq. (2) [i.e., ignoring the temporal similarity coefficient given by Eq. (3) altogether]. This was found to cause only a small change in the results. This is consistent with the observations made by Middlebrooks (1992) that subjects localized sounds consistently in the direction for which the directional transfer function magnitude spectrum most closely resembled the stimulus spectrum. This further confirmed the validity of the model used in the current paper for prediction of the localization direction. Furthermore, even with the model operating only on left ear data and with lack of exactly matching direct measurement directions for most of the reciprocal measurement directions, the observed predicted localization errors were small (except for the region of comparatively large natural human localization errors). The error could be expected to decrease further if HRTF for both ears were used in the model.

F. Eardrum SPL measurement

The sound intensity level produced at the eardrum by the in-ear microspeaker embedded within a silicone plug was then evaluated to ensure safety of the technique when used with a human subject. The KEMAR used in the experiment was equipped with a microphone located inside the head in the opening of the occluded ear simulator at the termination of the ear canal (i.e., at the simulated eardrum). Four signal recordings were performed at this location.

In the first recording, the KEMAR was placed in the direct measurement setup, and the ear canal was left open. The direct setup loudspeaker broadcast the test pulses used in the direct HRTF measurement technique. This recording thus provided an estimation of the sound level that would be observed at the eardrum during the direct HRTF measurement if the ear canal were not blocked. In the second recording, the setup was unchanged, but the ear canal was sealed with a silicone plug with an embedded microphone. In this way, the sound level in the direct measurement procedure with a blocked ear canal was evaluated. In the third recording, the sound level at the simulated eardrum was measured during the course of a reciprocal measurement of KEMAR HRTF. Finally, the fourth recording consisted of the voice of a person talking in a normal voice approximately 1 m in front of the KEMAR. Table III shows the sound pressure levels evaluated over different time windows for all four signals. The first column is the dB level corresponding to the peak amplitude of the signal, the second column is the rms dB SPL over the window that begins at the start of the recorded pulse and is twice as long as the pulse (i.e., 4.9 ms for the reciprocal HRTF measurement method and 3.6 ms for the direct HRTF measurement method), and the third column is the rms dB SPL taken over the whole signal. (All dB levels are relative to 20 μPa sound pressure.) It can be seen that the peak SPL for the reciprocal measurement method was about 14 dB higher than the peak speech SPL; however, it was well below tolerance threshold shown in OSHA (1974). Also, the average SPL for the reciprocal measurement method was only 11 dB higher than the average speech SPL due to the long interpulse interval.

TABLE III. Comparison of the SPL observed in the KEMAR ear canal for the direct method, for the reciprocal method, and for a speech signal.

	Peak SPL (dB)	rms SPL, windowed (dB)	rms SPL, whole signal (dB)
Direct method	80.1	67.1	50.0
Direct w/blocked ear canal	57.6	44.8	28.5
Reciprocal method	89.4	78.9	63.3
Speech recording	75.8	N/A	52.2

G. Discussion

The main issue to be decided is whether the directly measured HRTF and the reciprocally measured HRTF are perceptually identical despite the small differences observed (see Fig. 7). The perceptual identity question is impossible to answer without doing an actual perceptual experiment with a set of human subjects. In this paper, a sound localization model [an extension of the model presented in Middlebrooks (1992)] was used instead. The results of this comparison showed that the average localization error predicted by the model is less than the spacing between HRTF measurement directions typically used now in the state-of-the-art HRTF measurement facilities. In fact, this result is all the more remarkable because the comparison was performed monaurally and because the reciprocal and the direct HRTF measurements were obtained under somewhat different conditions.

Based on the results presented, it can be reasonably expected that the reciprocally measured HRTF can be used interchangeably with the directly measured HRTF in virtual audio synthesis and that the errors introduced by such exchange would lie within the errors caused by the discreteness of the measurement grid, by the natural errors in sound localization by humans, and by the natural variability in HRTF between repetitive experimental HRTF measurements due to subject positioning variability, microphone/speaker positioning variability, and noise.

VI. CONCLUSIONS

A new experimental method of measuring the HRTF rapidly is presented. The method is based on the reciprocity principle and exchanges the positions of the speaker and the microphone in the direct HRTF measurement setup. The

method is validated in two setups (a sphere and a manikin head), and it is shown that the transfer functions measured with the proposed method and with the direct measurement method are in good agreement.

ACKNOWLEDGMENTS

Partial support of NSF Award Nos. 0086075 and 0205271 is gratefully acknowledged. We thank Dr. Kenneth W. Grant of the Army Audiology and Speech Center at Walter Reed Army Medical Center, Washington DC, for allowing us to use the center's KEMAR manikin in these experiments. We would also like to acknowledge helpful discussions with Dr. R. O. Duda, Dr. V. R. Algazi, and Dr. B. G. Shinn-Cunningham regarding KEMAR HRTF measurement procedure and to thank Emkay electronics for providing us with the microspeakers used in the experiments. Finally, we would like to thank Dr. Armin Kohlrausch, associate editor, and three anonymous reviewers who provided a very careful review that helped us to significantly improve the manuscript.

Algazi, V. R., Avendano, C., and Thompson, D. (**1999**). "Dependence of subject and measurement position in binaural signal acquisition," J. Audio Eng. Soc. **47**, 937–947.

Algazi, V. R., Duda, R. O., Morrison, R. P., and Thompson, D. M. (**2001a**). "Structural composition and decomposition of HRTFs," Proc. IEEE WASPAA 2001, New Paltz, NY, pp. 103–106.

Algazi, V. R., Avendano, C., and Duda, R. O. (**2001b**). "Elevation localization and head-related transfer function analysis at low frequencies," J. Acoust. Soc. Am. **109**, 1110–1122.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001c**). "The CIPIC HRTF database," Proc. IEEE WASPAA 2001, New Paltz, NY, pp. 99–102.

Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., and Tang, Z. (**2002a**). "Approximating the head-related transfer function using simple geometric models of the head and torso," J. Acoust. Soc. Am. **112**, 2053–2064.

Algazi, V. R., Duda, R. O., and Thompson, D. M. (**2002b**). "The use of head-and-torso models for improved spatial sound synthesis," Proc. AES 113th Convention, Los Angeles, CA, preprint 5712.

Batteau, D. W. (**1967**). "The role of the pinna in human localization," Proc. R. Soc. London, Ser. B **168**, 158–180.

Blauert, J. (**1969**). "Sound localization in the median plane," Acustica **22**, 205–213.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Blommer, M. A. and Wakefield, G. H. (**1997**). "Pole-zero approximation for head-related transfer function using a logarithmic error criterion," IEEE Trans. Speech Audio Process. **5**, 278–287.

Bronkhorst, A. W., (**1995**). "Localization of real and virtual sound sources," J. Acoust. Soc. Am. **95**, 2542–2553.

Brungart, D. S. and Rabinowitz, W. M. (**1999**). "Auditory localization of nearby sources. Head-related transfer functions," J. Acoust. Soc. Am. **106**, 1465–1479.

Burkhard, M. D. and Sachs, R. M. (**1975**). "Anthropometric manikin for acoustic research," J. Acoust. Soc. Am. **58**, 214–222.

Carlile, S., (ed.) (**1996**). *Virtual Auditory Space: Generation and Applications* (Landes, Austin).

Carlile, S., Leong, P., and Hyams, S. (**1997**). "The nature and distribution of errors in the localization of sounds by humans," Hear. Res. **114**, 179–196.

Divenyi, P. L. and Oliver, S. K. (**1989**). "Resolution of steady-state sounds in simulated auditory space," J. Acoust. Soc. Am. **85**, 2042–2052.

Duda, R. O. and Martens, W. L. (**1998**). "Range dependence of the response of a spherical head model," J. Acoust. Soc. Am. **104**, 3048–3058.

Duraiswami, R. and Gumerov, N. A. (**2003**). "Method for measurement of head related transfer functions," United States Patent Application 20040091119, Serial No. 10/702465, filed 7 November 2003.

Gardner, M. B. and Gardner, R. S. (**1973**). "Problem of localization in the median plane: Effect of pinna cavity occlusion," J. Acoust. Soc. Am. **53**, 400–408.

Gardner, W. G. and Martin, K. D. (**1995**). "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908.

Grassi, E., Tulsi, J., and Shamma, S. A. (**2003**). "Measurement of head-related transfer functions based on the empirical transfer function estimate," Proc. 2003 Intl. Conf. on Auditory Displays (ICAD 2003), Boston, MA, pp. 119–121.

Hartmann, W. M. (**1999**). "How we localize sound," Phys. Today **1999**, 24–29.

Hartmann, W. M. and Wittenberg, A. (**1996**). "On the externalization of sound images," J. Acoust. Soc. Am. **99**, 3678–3688.

Jin, C., Leong, P., Leung, J., Corderoy, A., and Carlile, S. (**2000**). "Enabling individualized virtual auditory space using morphological measurements," Proc. of the 1st IEEE Pacific-Rim Conf. on Multimedia (2000 Intl. Symposium on Multimedia Information Processing), Sydney, Australia, pp. 235–238.

Kahana, Y. (**2000**). "Numerical modeling of the head-related transfer function," Ph.D. thesis, ISVR, University of Southampton, UK.

Kahana, Y. and Nelson, P. A. (**2000**). "Spatial acoustic mode shapes of the human pinna," Proc. AES 109th Convention, Los Angeles, CA, preprint 5218.

Katz, B. F. G. (**2001a**). "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," J. Acoust. Soc. Am. **110**, 2440–2448.

Katz, B. F. G. (**2001b**). "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements," J. Acoust. Soc. Am. **110**, 2449–2455.

Kulkarni, A. and Colburn, H. S. (**1998**). "Role of spectral detail in sound-source localization," Nature (London) **396**, 747–749.

Kulkarni, A. and Colburn, H. S. (**2004**). "Infinite-impulse-response models of the head-related transfer function," J. Acoust. Soc. Am. **115**, 1714–1728.

Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (**1999**). "Sensitivity of human subjects to head-related transfer-function phase spectra," J. Acoust. Soc. Am. **105**, 2821–2840.

Langendijk, E. H. A. and Bronkhorst, A. W. (**2000**). "Fidelity of three-dimensional-sound re-production using a virtual auditory display," J. Acoust. Soc. Am. **107**, 528–537.

Mehrgardt, S. and Mellert, V. (**1977**). "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. **61**, 1567–1576.

Middlebrooks, J. C. (**1992**). "Narrow-band sound localization related to external ear acoustics," J. Acoust. Soc. Am. **92**, 2607–2624.

Middlebrooks, J. C. (**1999a**). "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. Am. **106**, 1480–1492.

Middlebrooks, J. C. (**1999b**). "Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency," J. Acoust. Soc. Am. **106**, 1493–1510.

Middlebrooks, J. C., Makous, J. C., and Green, D. M. (**1989**). "Directional sensitivity of sound-pressure levels in the human ear canal," J. Acoust. Soc. Am. **86**, 89–108.

Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (**1995**). "Head-related transfer functions of human subjects," J. Audio Eng. Soc. **43**, 300–321.

Morse, P. M. and Ingard, K. U. (**1968**). *Theoretical Acoustics* (Princeton U.P., Princeton, NJ).

Musicant, A. D. and Butler, R. A. (**1984**). "The influence of pinnae-based spectral cues on sound localization," J. Acoust. Soc. Am. **75**, 1195–1200.

Occupational Safety and Health Administration, U. S. Dept. of Labor (**1974**). "Occupational Noise Exposure," Regulation 1910.95, Standards 29 CFR.

Pralong, D. and Carlile, S. (**1994**). "Measuring the human head-related transfer functions: A novel method for the construction and calibration of a miniature in-ear recording system," J. Acoust. Soc. Am. **95**, 3435–3444.

Rabinowitz, W. M., Maxwell, J., Shao, Y., and Wei, M. (**1993**). "Sound localization cues for a magnified head: Implications from sound diffraction about a rigid sphere," Presence **2**, 125–129.

Runkle, P., Yendiki, A., and Wakefield, G. H. (**2000**). "Active sensory tuning for immersive spatialized audio," Proc. 2000 Intl. Conf. on Auditory Displays (ICAD 2000), Atlanta, GA, pp. 141–144.

Schroeder, M. R. (**1979**). "Integrated-impulse method measuring sound decay without using impulses," J. Acoust. Soc. Am. **66**, 497–500.

Shaw, E. A. G. and Teranishi, R. (**1968**). "Sound pressure generated in an

external-ear replica and real human ears by a nearby point source," J. Acoust. Soc. Am. **44**, 240–249.
Shinn-Cunningham, B. G., Santarelli, S. G., and Kopco, N. (**2000**). "Tori of confusion: Binaural localization cues for sources within reach of a listener," J. Acoust. Soc. Am. **107**, 1627–1636.
Strutt, J. W., (Lord Rayleigh) (**1904**). "On the acoustic shadow of a sphere," Philos. Trans. R. Soc. London, Ser. A **203**, 87–89.
Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123.
Wightman, F. L. and Kistler, D. J. (**1989**). "Headphone simulation of free-field listening. I. Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.
Wright, D., Hebrank, J. H., and Wilson, B. (**1974**). "Pinna reflections as cues for localization," J. Acoust. Soc. Am. **56**, 957–962.
Zahorik, P. A., Wightman, F. L., and Kistler, D. J. (**1995**). "On the discriminability of virtual and real sound sources," Proc. IEEE WASPAA 1995, New Paltz, NY, pp. 76–79.
Zhou, B., Green, D. M., and Middlebrooks, J. C. (**1992**). "Characterization of external ear impulse responses using Golay codes," J. Acoust. Soc. Am. **92**, 1169–1171.

Effects of directional microphone and adaptive multichannel noise reduction algorithm on cochlear implant performance

King Chung[a)]
Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana 47907

Fan-Gang Zeng
Department of Otolaryngology/Head & Neck Surgery, University of California, Irvine, CA 92697

Kyle N. Acker
Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana 47907

(Received 12 March 2006; revised 3 July 2006; accepted 5 July 2006)

Although cochlear implant (CI) users have enjoyed good speech recognition in quiet, they still have difficulties understanding speech in noise. We conducted three experiments to determine whether a directional microphone and an adaptive multichannel noise reduction algorithm could enhance CI performance in noise and whether Speech Transmission Index (STI) can be used to predict CI performance in various acoustic and signal processing conditions. In Experiment I, CI users listened to speech in noise processed by 4 hearing aid settings: omni-directional microphone, omni-directional microphone plus noise reduction, directional microphone, and directional microphone plus noise reduction. The directional microphone significantly improved speech recognition in noise. Both directional microphone and noise reduction algorithm improved overall preference. In Experiment II, normal hearing individuals listened to the recorded speech produced by 4- or 8-channel CI simulations. The 8-channel simulation yielded similar speech recognition results as in Experiment I, whereas the 4-channel simulation produced no significant difference among the 4 settings. In Experiment III, we examined the relationship between STIs and speech recognition. The results suggested that STI could predict actual and simulated CI speech intelligibility with acoustic degradation and the directional microphone, but not the noise reduction algorithm. Implications for intelligibility enhancement are discussed. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2258500]

PACS number(s): 43.66.Ts, 43.60.Qv, 43.71.Ky, 43.71.An, 43.60.Fg [BLM] Pages: 2216–2227

I. INTRODUCTION

Although the performance of cochlear implant (CI) users has been increasing constantly with recent improvements in CIs (Loizou, 1998; Zeng, 2004), understanding speech in noise still remains a great challenge. Multiple attempts have been made to increase the signal-to-noise ratio (SNR) of sounds before the signal reaches the CI preprocessor but with limited success. For example, the Audallion BEAMformer that summed the inputs from two directional microphones worn in the implanted ear and in the contralateral ear was reported to have limited benefit on localizing the source of sounds (Figueiredo *et al.*, 2001), limited directional effects (Goldsworthy, 2005), and limited acceptability among CI users. In addition, several groups of researchers have reported positive findings using spectral subtraction noise reduction algorithms to enhance the speech recognition of CI users in background noise (Hochberg *et al.*, 1992; Weiss, 1993; Goldsworthy, 2005). The computational demand, however, prevents such algorithms from being implemented in wearable CIs. As technologies advance, new generations of front-end processors (e.g., directional microphones and adaptive directional microphones) are implemented in some recent CI models to combat noise but few studies have reported the effectiveness of these strategies. In this study, we will explore the effects of two noise reduction strategies commonly used in hearing aids, directional microphones and adaptive multichannel (AMC) noise reduction algorithms, on cochlear implant performance.

Directional microphones have higher sensitivity to sounds coming from the front than from the sides or the back. They are applied to hearing devices to take advantage of spatial separation between speech and noise. Most directional microphones implemented in high performance hearing devices utilize two omni-directional microphones, which are equally sensitive to sounds from all directions. When the directional microphones are activated, the electrical signal generated by the back microphone is subtracted from that of the front microphone. Depending on the ratio of the distance between the two omni-directional microphones and the electronic delay added to the back microphone, the polar pattern can vary from bipolar, to hypercardioid, supercardioid or cardioid. This means that the least sensitive location(s) of the microphone (i.e., the null) can change from 90° and 270° in bipolar patterns to 180° in cardioid pattern. The overall improvement provided by directional microphones is roughly 3–5 dB in real-world environments with low reverberation compared to omni-directional microphones for listeners with

[a)]Electronic mail: kingchung@purdue.edu

 0001-4966/2006/120(4)/2216/12/$22.50

acoustic hearing (Amlani, 2001; Chung, 2004; Ricketts, 2001; Valente *et al.*, 2000; Wouters *et al.*, 1999; Bentler, 2005).

Speech is a highly modulated signal (Plomp, 1983; Rosen, 1992). Speech sounds are characterized by temporal modulations between 2 and 50 Hz with a peak modulation rate of 3–8 Hz depending on speaking rate (Powers *et al.*, 1999). Noise, however, is often displayed as having a steadier temporal envelope or having a modulation rate outside the modulation range of speech. When speech and noise coexist in a signal, the modulation depth of speech is reduced because noise fills the silence between phonemes or sentences and masks the low-level components of speech.

Most noise reduction algorithms implemented in digital hearing devices utilize an adaptive gain reduction mechanism to take advantage of spectral separation between speech and noise across multiple frequency channels (i.e., AMC noise reduction algorithms). They use a modulation detector to monitor the modulation characteristics of the signal within each frequency channel and apply gain reduction to reduce noise interference. If a high modulation depth with center modulation rates similar to those of speech is detected in a channel, the noise reduction algorithm infers that speech exists in the channel and the SNR is high so the gain of that frequency channel is not reduced. Otherwise, the algorithms infer that the SNR is low or noise dominates the frequency channel and the gain of the channel is reduced. Usually the amount of gain reduction increases as the SNR decreases. The exact amount of gain reduction depends on the decision rules of the particular noise reduction algorithm which, in turn, is determined by the modulation depth, the estimated SNR in the frequency channel, overall level of the signal, the level detected in the channel, the frequency-importance of the channel for speech recognition, and so on (Bentler, 2005; Chung, 2004; Ricketts and Hornsby, 2005). Although few noise reduction algorithms are reported to be effective in modulated background noise (e.g., single-talker babble), many are reported to reduce noise interference and increase sound quality, listening comfort, or overall preference in noise with limited temporal variations (e.g., speech spectrum noise) (Edwards *et al.*, 1988; Walden *et al.*, 2000; Johns *et al.*, 2003; Kuk *et al.*, 2002; Mueller *et al.*, 2003; Powers and Hamacher, 2002; Ricketts and Hornsby, 2005).

The speech transmission index (STI) was first proposed to predict speech intelligibility in rooms with different acoustic properties (Houtgast and Steeneken, 1973). It is calculated by comparing the changes in modulation depths of the different frequency regions between a probe (a modulated noise) and a transmitted signal after the probe is transmitted across an acoustic medium. A number of researchers have also proposed several speech-based transmission index calculation methods to utilize speech as the probe signal, instead of the modulated noise (Payton *et al.*, 2002; Drullman *et al.*, 1994; Ludvigsen *et al.*, 1990; Koch, 1992; Holube and Kollmeier, 1996). The advantage of speech-based calculation methods is that they allow the estimation of speech intelligibility in different acoustic environments and under different speech signal processing algorithms. In general, the higher the modulation depth in the transmitted or processed signal, the higher the speech transmission index and the higher the predicted speech intelligibility. The disadvantage is that STI has been reported to fail to predict the speech intelligibility of nonlinearly processed speech for people using acoustic hearing (e.g., for spectral subtraction noise reduction algorithm, see Ludvigsen *et al.*, 1993, Goldsworthy and Greenberg, 2004; for envelope clipping, see Drullman, 1995; for compression, see Hohmann and Kollmeier, 1995; for envelope thresholding, see Goldsworthy, 2005). It appeared that factors other than the modulation depth affected the performance of people using acoustic hearing.

Nevertheless, there are fundamental differences between how speech is encoded in acoustic and electric hearing. Previous studies have shown that listeners with normal hearing and hearing aid users can use both spectral and temporal information for speech understanding (Van Tasell *et al.*, 1987, 1992). Yet, CI users are forced to rely on the temporal envelope cues to understand speech because the spectral fine structure information is only coarsely presented in 6 to 22 channels (Loizou, 1998). Our previous recordings showed that both directional microphones and AMC noise reduction algorithms increase temporal modulation depths of speech in background noise (Chung *et al.*, 2004b). It is possible that both of these strategies make speech envelopes more salient and can help the CI speech processor determine which speech peaks to present to the electrodes and thus improve speech recognition for CI users. Additionally, as the STI calculations are also based on temporal modulations in nonoverlapping filter bands, similar to CIs, it is possible that STI could be used to predict speech intelligibility of both noise reduction strategies for CI users.

In a previous study, Chung *et al.* (2004a, b) conducted a series of preliminary studies to investigate whether directional microphones and noise reduction algorithms could enhance CI performance. They recorded speech in noise testing materials processed by a 9-channel and a 6-channel digital hearing aid when the hearing aids were set to omnidirectional microphone, directional microphone, and directional microphone plus AMC noise reduction. The testing materials were then presented to CI users via direct audio input. The results showed that the two conditions with the directional microphone yielded significantly better speech recognition and higher sound quality rankings than the omnidirectional microphone condition for both hearing aids (Chung, 2004a, b). Significantly better speech recognition scores were also observed for the directional microphone with noise reduction condition compared with the directional microphone alone condition for the 9-channel digital hearing aid. These studies, however, were conducted with relatively small sample sizes (i.e., CI users $N=4$ and 8). It is unknown if these positive results can be generalized to a larger CI population. Additionally, the effect of noise reduction algorithms alone was not investigated and the CI users showed near floor performance in some conditions because the speech testing materials were recorded at a low SNR (i.e., +3 dB).

The purposes of this series of experiments were to investigate whether (1) directional microphones and AMC noise reduction algorithms used as preprocessors could en-

hance CI performance in noisy environments; and (2) a speech-based STI could be applied to predict the CI performance using these two noise reduction strategies. These research questions were addressed in 3 experiments: in Experiment I, the speech recognition ability of CI users was tested when listening to speech in noise testing materials processed by a digital hearing aid with directional microphone and noise reduction algorithm. The CI users also rated their overall preferences of the test conditions at three SNRs in a paired-comparison paradigm. In Experiment II, the speech recognition ability of two groups of listeners with normal hearing was tested when they listened to the same testing materials with 4- or 8-channel CI simulations. In Experiment III, the relationship between the speech recognition scores of CI users and normal hearing individuals and STIs calculated using the speech-based STI method proposed by Payton *et al.* (2002) was explored. In all the experiments involving research participants, repeated measure designs were used to reduce variability due to subject variability and all participants were blinded to the testing conditions to eliminate systematic errors caused by subject bias.

II. EXPERIMENT I

A custom in-the-ear 9-channel digital hearing aid was made for Knowles Electronic Manikin for Acoustic Research (KEMAR)'s right ear. Speech testing materials were recorded when the hearing aid was programmed to different settings then presented to CI users via direct audio input. This procedure was used to simulate the condition in which a hearing aid signal processor preceded a CI speech processor.

As CI users have a wide range of speech understanding ability, we recorded the speech testing materials at five SNRs to avoid floor and ceiling effects, and to bracket the SNR for 50% correct speech recognition (SNR_50). All the recordings were made in an anechoic chamber to minimize the effects of reverberation on STI calculations in Experiment III. As one of the goals of amplification devices is to improve perceived sound quality, CI users also rated their overall preference of the processed speech recorded at three SNRs. The following are detailed descriptions of the experimental procedures used in this study:

A. Methods

1. Subjects

Seven male and 13 female CI users (mean age=58.2 years old) were recruited for this experiment. Their demographic information and information on their CIs are summarized in Table I. All listeners participated in the speech recognition in noise test and 13 listeners rated overall preferences of the experimental conditions. The tests were conducted at Purdue University in West Lafayette, IN and at University of California in Irvine, CA.

2. Characteristics of the digital hearing aid

The digital hearing aid used in this study had a first-order directional microphone with a fixed hypercardioid pattern. Only one hearing aid was used because we wanted to control the amount of directional effects in the directional condition across subjects. The test hearing aid had a −3 dB/octave low-frequency roll-off in the directional setting compared to the omni-directional setting in order to compensate for half of the low-frequency roll-off of the first-order directional microphone.

TABLE I. The demographic and cochlear implant information of listeners.

Subject	Gender	Age	Test ear	Number of years of CI use	Speech processor	Coding strategy
1	M	79	L	2;3	ESPrit 3G	CIS
2	M	47	R	11;10	Spectra 22	SPEAK
3	M	73	R	1;11	Auria	HiRes
4	F	41	L	4;9	Combi40+	CIS
5	F	60	R	6;7	Clarion	CIS
6	F	72	R	4;10	ESPrit 3G	ACE
7	F	43	R	19;6	Ineraid	CIS
8	F	62	R	1;0	ESPrit 3G	ACE
9	M	78	R	1;10	ESPrit 3G	ACE
10	M	63	L	14;10	Spectra 22	SPEAK
11	F	67	R	4;3	Clarion	MPS
12	F	72	L	11;0	S-Series	CIS
13	M	27	L	0;6	ESPrit 3G	ACE
14	F	44	L	5;11	ESPrit 3G	SPEAK
15	F	51	L	6;7	S-Series	SAS
16	M	57	R	3;7	ESPrit 3G	ACE
17	F	57	R	6;0	ESPrit 3G	SPEAK
18	F	45	L	1;9	Combi40+	CIS
19	F	70	R	7;7	ESPrit 3G	ACE
20	F	55	L	3;0	ESPrit 3G	ACE

The noise reduction algorithm implemented in this hearing aid was an AMC noise reduction algorithm with nine signal processing channels. The amount of gain reduction in each channel was inversely proportional to the estimated SNR. No gain reduction was executed if the SNR was estimated to be at or higher than 24 dB. A maximum of 12 dB gain reduction was exercised if the SNR was estimated to be at 0 dB at a frequency channel. The noise reduction algorithm reduced the noise to within 3 dB of the steady noise level in 8 s and to within 1 dB in 14 s.

3. Preparation of speech recognition testing materials

Prior to making the recordings of the speech testing materials, the hearing aid was programmed to have linear signal processing and flat frequency response when it was worn in KEMAR's ear (i.e., flat *in situ* frequency response). The expansion algorithm at very low input level was turned off in all testing conditions. Thus the omni-directional microphone setting of the hearing aid provided little frequency or amplitude alterations to the incoming sounds. It acted as a reference condition for other hearing aid processed conditions as if no hearing aid preprocessor were added to the CI speech processor.

The speech recognition testing materials were recorded using the equipment setup in Fig. 1. In the calibration process, Computer 1 (2.39 GHz Pentium 4 with 1 Gbyte of RAM) presented the speech spectrum calibration noise from the Hearing In Noise Test (HINT, Nilsson *et al.*, 1994) to Speaker 1, which was a Mackie HR824 powered amplifier

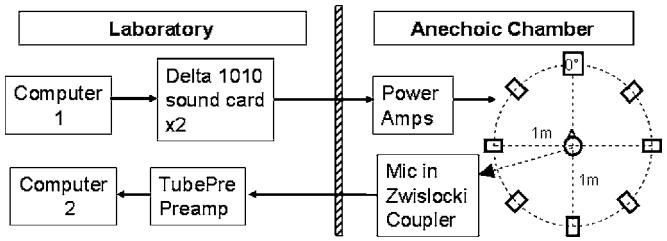


FIG. 1. The equipment setup for the recording of hearing aid processed testing materials.

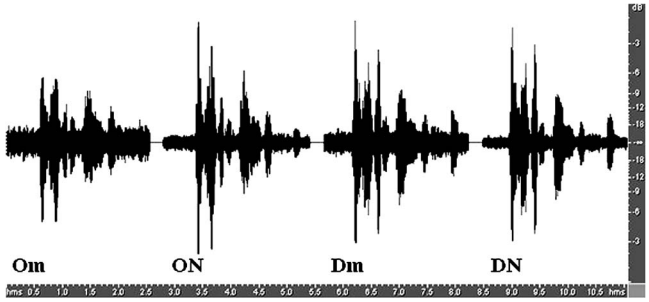


FIG. 2. The temporal envelopes of the sentence "The house had nine bedrooms" processed under four hearing aid settings: omni-directional microphone (Om), omni-directional microphone plus noise reduction algorithms (ON), directional microphones (Dm), and directional microphones plus noise reduction algorithms (DN) at SNR of +8 dB.

with a ±1.5 dB frequency response from 39 to 20 kHz. The level of the calibration noise was adjusted to be 68, 70.5, 73, 75.5, and 78 dB SPL measured by a sound level meter placed in the center of the speaker array in the absence of KEMAR.

An uncorrelated uniform noise field was generated by a total of 8 speakers at 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315° (Speakers 2–8 were Hafler M5 speakers). The level of the speech spectrum calibration noise was adjusted to be 56 dB SPL from individual speakers which resulted in a uniform noise field with an overall level of 65 dB SPL. An uncorrelated noise field was used to avoid the comodulation masking release effect which is a result of correlated noise field and could yield an increase in speech understanding scores compared to an uncorrelated noise field (Cox and Bisset, 1984; Grose and Hall, 1992; Kwon, 2002; Moore, 1990). A continuous noise field, instead of the gated noise used in the original HINT test, was utilized to ensure that the noise reduction algorithm was always engaged.

In the recording process, KEMAR was placed in the center of the speaker array. The output of the hearing aid was recorded by an ER11 $\frac{1}{2}$ in. microphone (Etymotic Research) placed in the medial opening of a Zwislocki coupler attached to KEMAR's ear canal, and then fed to Computer 2 (1.8 GHz Intel Pentium M processor with 512 Mbytes RAM).

Four lists of HINT sentences were recorded at each SNR (i.e., +3, +5.5, +8, +10.5, and +13 dB) when the hearing aid was set to omni-directional microphone (Om), omni-directional microphone with noise reduction (ON), directional microphone (Dm), and directional microphone with noise reduction (DN). The sentences were presented approximately 10 s after the presentation of noise to ensure that they were recorded after the actions of the noise reduction algorithm had stabilized. Each sentence was separated by approximately 5–6 s of noise, as in the original HINT test.

After the testing materials were recorded, the rms levels of speech were equalized across experimental conditions in order to minimize the possibility that CI speech processors with narrow input-dynamic range peak-clipping signals with higher levels. The temporal envelope plots of the sentence "The house has nine bedrooms" processed by the four hearing aid settings is shown in Fig. 2.

Three sentences were arbitrarily chosen at each of the +3, +8, +13 dB SNR to create paired-comparison tokens for overall preference ratings. Each sentence formed 12 tokens (i.e., 6 combinations: Om-ON, Om-Dm, Om-DN, ON-Dm, ON-DN, Dm-DN; and 6 reversals: ON-Om, Dm-Om, DN-Om, Dm-ON, DN-ON, DN-Dm). A 500 ms silence was inserted between the sentences. All wave-editing tasks were carried out using Adobe AUDITION 1.0.

4. Speech recognition in noise test

Prior to the administration of speech recognition tests, the key words in the HINT sentence lists were analyzed and the comfortable listening levels were determined for individual listeners. Any auxiliary words for continuous tense were not counted as key words, but the same words used as verb were counted. For example, the word "is" was not counted as a key word in "(A/The) *car* (*is/was*) *going too fast*," but it was counted as a key word in "(A/The) *fire* (*is/was*) *very hot*." Articles were not counted as key words. Any variations allowed in the original HINT test were also allowed in this study. For example, in the sentence "(A/The) *fire* (*is/was*) *very hot*," both "is" and "was" were counted as correct. Using this scoring method, the number of key words for the sentence lists ranged from 38 to 45 words. In addition, each listener's comfortable listening level for the recorded speech via direct audio input was determined using the loudness scale and procedures developed by the Independent Hearing Aid Fitting Forum (IHAFF, 1994, i.e., "1" – very soft, "2" – soft, "3" – comfortable but slightly soft, "4" – comfortable, "5" – comfortable but slightly loud; "6" – loud but ok, and "7" – too loud).

During speech recognition tests, the sentences recorded at +8 dB SNR were administered to the listeners first. If a listener obtained a speech recognition score close to 0% or 100% at SNR of +8 dB for most hearing aid settings, HINT lists recorded at +3 and +13 dB were presented to see if the listener reached the floor or ceiling of performance. If so, no further tests were administered and the listener was discharged. Otherwise, HINT lists from higher and lower SNRs were presented to bracket the SNR_50 for the individual CI user. These procedures were adopted because a three-parameter sigmoidal function was used to fit the data points and SNR_50s were estimated using the parameters generated from the sigmoidal function. Thus, if a listener was tested at their floor or ceiling performance levels, the sigmoidal function would generate erroneous results. The speech testing materials were presented to listeners at their comfortable listening levels (i.e., "4" in the IHAFF scale). Listener 13 lis-

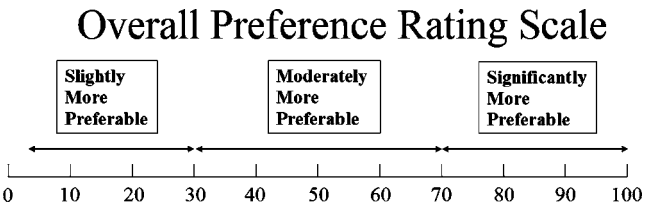


FIG. 3. The overall preference rating scale used in the paired comparison categorical rating paradigm.

tened at a "comfortable but slightly soft" level (i.e., "3") because she reported distortions of speech at the "4" level during a routine check.

The percentage correct speech recognition scores were calculated by dividing the number of key words the listener repeated correctly by the total number of key words in the particular sentence list. The percent correct scores for each hearing aid signal processing setting were then converted to SNR_50s using a three parameter sigmoidal function:

$$\text{Percent Correct} = A/(1 + \exp(-(\text{SNR} - X0)/B)), \quad (1)$$

$$\text{SNR}\_50 = X0 - B\ln(A/0.5 - 1), \quad (2)$$

where A is the asymptotic performance, $X0$ is the SNR at which the percent correct performance is 50% of A, and B is a parameter related to the slope.

5. Overall preference ratings

Thirteen CI users rated their overall preferences of sentences in a combined paired comparison and categorical rating paradigm. After listening to each paired-comparison token, they reported their preference of Condition 1 or Condition 2 to the examiner and then rated the magnitude of their preferences using the scale shown in Fig. 3. A total of 18 tokens (i.e., 6 combinations/reversals $\times 3$ sentences) were administered at each SNR to each listener.

In the scoring process, the 6 combinations and their reversals were grouped into 6 pairs (e.g., Om-ON tokens were grouped with ON-Om tokens). If a listener indicated that they preferred Condition 1 (say ON) over Condition 2 (say Om) by 30 points, a score of 30 was entered for ON and a score of 0 was entered for Om. The average of Om or ON equaled the sum of scores in the Om-ON combination divided by the total number of trials.

B. Results

1. Speech recognition in noise

The raw speech recognition scores of all subjects are depicted in Fig. 4. The SNR_50 for each hearing aid setting was calculated for all the listeners except listeners 1, 6, 7, 15, and 18 whose SNR_50s could not be calculated in at least one hearing aid setting because of poor performance. Subsequently, the SNR_50 of 16 listeners was analyzed using repeated measure ANOVA on hearing aid settings to determine whether the directional microphone and noise reduction algorithm could improve listeners' speech recognition in background noise. The results showed significant main effects of hearing aid settings ($F[15,3]=17.9, p<0.0001$). The average SNR_50 for the four hearing aid conditions were 6.0, 4.4, 2.5, and 1.9 dB for the Om, ON, Dm, and DN conditions.

Scheffe pairwise comparisons were carried out to determine hearing aid settings that yielded significant differences. The difference was significant between Om and Dm, Om and DN, ON and Dm, and ON and DN ($p<0.0083$, adjusted to account for multiple test conditions). The critical difference for significance was 1.9 dB. The absence of significant difference between Om and ON or between Dm and DN indicated that the directional microphone improved CI users' speech understanding in noise but the noise reduction algorithm did not.

The CI users exhibited a wide range of speech coding and electrical stimulation strategies. While it was not the intention of this study to test the applicability of the directional microphone and the noise reduction algorithm to a particular group of CI users, the results of the largest group of participants with the same speech processor (i.e., the seven EsPrit 3G listeners) were analyzed as a group to investigate whether the speech processor played a role in the process. The repeated measure ANOVA indicated a significant hearing aid setting main effect ($F[6,3]=6.13, p<0.01$). Scheffe pairwise comparison tests showed significant difference between Om and Dm, and between Om and DN. The critical difference for significance was 3.1 dB. The general results were similar to those of the whole group, that directional microphone enhanced speech understanding while the noise reduction algorithm did not.

2. Overall preference ratings

Six paired t-tests were performed to determine significant differences among the comparison pairs at each SNR. The p level for 0.05 significance level was adjusted to be 0.003 to account for multiple t-tests [i.e., 0.05/(3 SNR $\times 6$ tests)]. The average overall preference ratings of the comparison pairs are summarized in Fig. 5 and the significantly different pairs are indicated by asterisks (*). Significant results were obtained between Om-ON, Om-Dm, Om-DN, ON-DN, and Dm-DN at SNRs of +3 and +8 dB ($p<0.0025$). The magnitude of preferences ranged from 23% (slightly more preferable) to 57% (moderately preferable). Overall, DN was ranked the most preferable and Om the least preferable. No significant differences were reported between ON and Dm at these SNRs. These results indicate that, at low SNRs, CI users preferred the conditions with the noise reduction algorithm and/or the directional microphone over Om, and their preferences were similar for the conditions with noise reduction alone and with directional microphone alone.

At a SNR of +13 dB, significant differences were obtained between Om-DN and between ON-DN only. The magnitudes of the preferences were 45% (moderately preferable) for both comparison pairs. The differences between all other pairs did not reach statistical significance. These results suggest CI users only preferred a combination of directional microphone and noise reduction algorithm over the unprocessed or noise reduction conditions at a high SNR. This may be because speech was already clear to them and enhance-

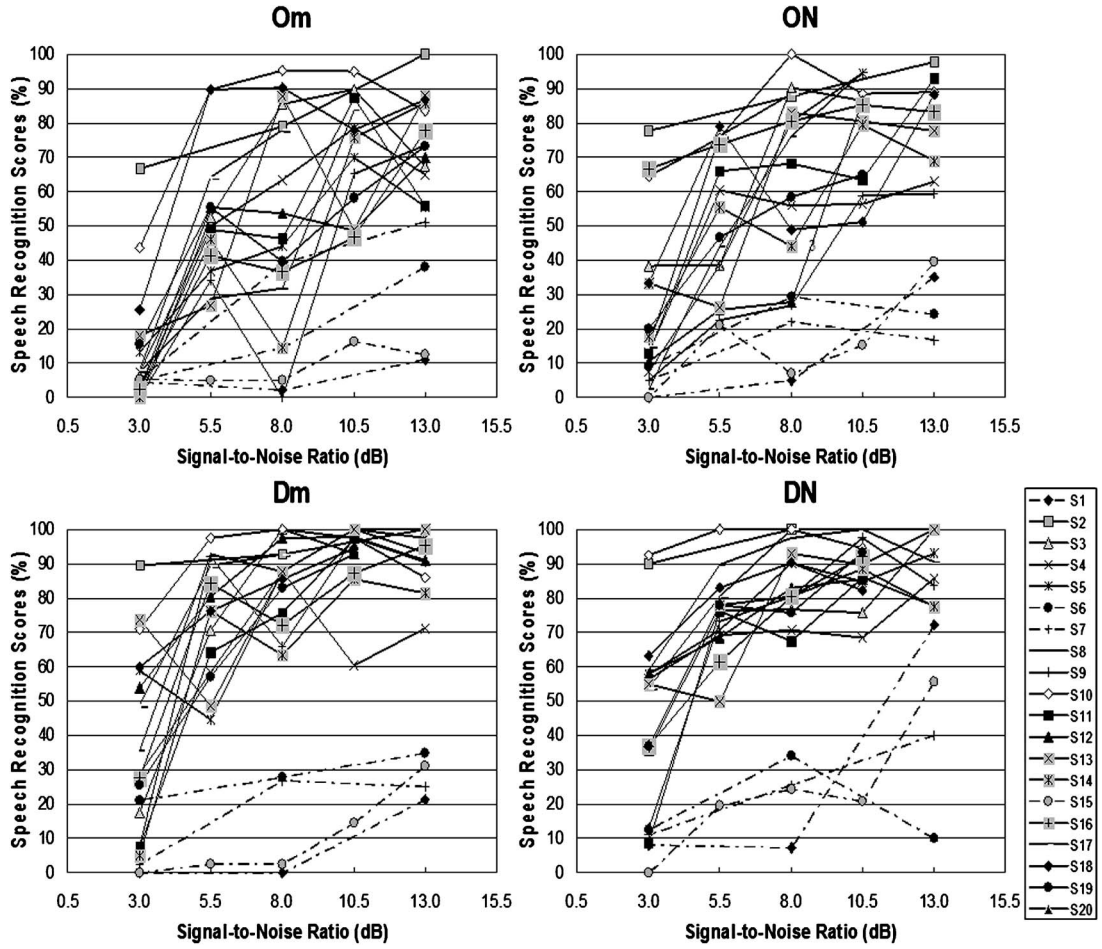


FIG. 4. The raw scores obtained by all the cochlear implant listeners when they listened to speech in noise materials processed by a digital hearing aid set to omni-directional microphone (Om), omni-directional microphone plus noise reduction algorithms (ON), directional microphones (Dm), and directional microphones plus noise reduction algorithms (DN) at 5 SNRs. Om represents a condition in which no hearing aid signal processing was added to the cochlear implant speech processor.

ments by directional microphone or noise reduction algorithm alone did not make a significant impact on their overall preferences.

C. Discussion

CI users obtained significantly better speech recognition scores when they listened to speech processed by directional microphones (i.e., Dm and DN) compared to the conditions without directional microphones (i.e., Om and ON, respectively). The directional microphone provided an average of 3.5 and 3.6 dB improvement in SNR (Om vs Dm) for all the CI users and the EsPrit 3G listeners, respectively. The amount of improvement is consistent with that reported in studies with simulated real-world environments (Amlani,

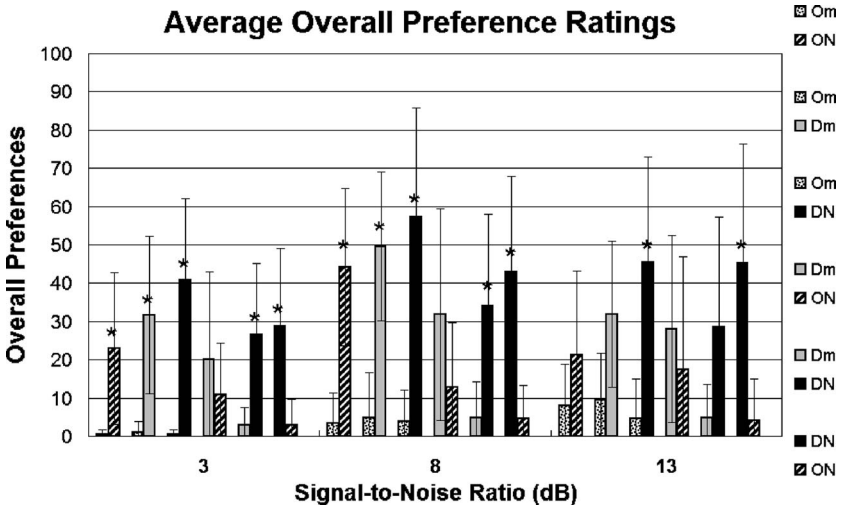


FIG. 5. The average overall preference ratings of cochlear implant listeners when they listened to speech processed at omni-directional microphone (Om), omni-directional microphone plus noise reduction algorithms (ON), directional microphones (Dm), and directional microphones plus noise reduction algorithms (DN) at 3 SNRs.

2001; Chung, 2004; Ricketts, 2001; Valente *et al.*, 2000; Wouters *et al.*, 1999; Bentler, 2005). Note that noise was not only presented from the sides and the back but also from 0° azimuth in the present study. This is rarely the case in other studies. The improvement from the directional microphone could be higher if noise was only presented from the sides and the back.

On the other hand, the SNR_50 obtained in conditions with AMC noise reduction algorithm (i.e., ON or DN) did not reach statistical significance when compared to that in the conditions without the noise reduction algorithm (i.e., Om and Dm, respectively). The difference between Om and ON for all CI users and the EsPrit listeners was 1.6 and 2.4 dB, respectively. It seems that the noise reduction algorithm used in this study somewhat increased the speech recognition of CI users. Future investigations are needed to explore if there are interaction effects between the noise reduction algorithm and speech processors, and if a different amount of noise reduction or noise reduction algorithms from other digital hearing aids would have a better noise reduction effect in helping CI users understand speech in background noise.

In Chung *et al.* (2004b), the improvement provided by the DN condition reached statistical significance when the CI users were tested at a SNR of +3 dB. In this study, the DN condition did not provide any significant improvement compared to Dm for either the CI or the EsPrit 3G group. This suggested that improvement provided by DN was minimal when the noise reduction algorithm was applied across a wider range of SNRs.

Subjectively, CI users preferred conditions with directional microphone and/or noise reduction algorithm. This result is consistent with other reports in hearing aid literature on the same hearing aid for listeners with acoustic hearing (Bray and Nilsson, 2001; Johns *et al.*, 2003). The amount of preference ranged from slightly more preferable to moderately preferable. According to Gabrielsson and Sjogren (1979), sound quality is a multidimensional phenomenon, namely, clarity, sharpness, brightness, fullness, spaciousness, nearness, noisiness, and loudness. The weightings of these dimensions are different for individual listeners and different tasks. Although it is unclear which dimension(s) determined subjective preferences in this study, it is possible that the reduced overall noise contributed to the higher overall preferences of the conditions with noise reduction algorithms (see Fig. 2). The implications are that AMC noise reduction algorithm can be applied in CIs to enhance perceived sound quality and to reduce the overall background noise.

Another interesting finding is that the average preference ratings of this group of CI users suggest that the preference ratings obtained using categorical rating paradigm were not transitive. In other words, if the listeners rated condition *A* *x* amount more preferable than condition B, and condition *B* *y* amount more preferable than condition *C*, the amount of preference for condition C compared to condition A did not equal to $x+y$. For example, at SNR of +3 dB, the difference between Om and DN rated by the listeners was 40.3. If transitivity held, the sum of differences (in absolute values) for Om_ON and Om_Dm (i.e., improvement from noise reduction (Om_ON)+ improvement from directional microphone (Om_Dm)=sum of improvement (Om_DN)) should be 40.3. However, the sum of differences was 53.0. Similar lack of transitivity was also observed in SNR of +8 and +13 dB. The implication is that, if we desire to know the preference ratings between multiple experimental conditions using paired comparison categorical rating paradigm, the ratings should be performed but not inferred or calculated.

III. EXPERIMENT II

Previous studies have shown that simulating listening to the CI for normal hearing individuals is a viable tool for providing insight into the effect of various signal processing strategies on CI users. At the same time, it eliminates sources of variability such as differences in electrical stimulation strategies, survival of cell bodies of first-order auditory neurons, location of electrodes relative to surviving neurons, etc. (Dorman and Loizou, 1998; Fu *et al.*, 1998, 2004; Stickney and Zeng, 2004; Nelson and Jin, 2004). In this study, normal hearing individuals listened to the speech materials recorded in Experiment I which were then processed to simulate CI processing with 4 or 8 channels of temporal envelope cues.

A. Materials and Methods

1. Subjects

Two groups of listeners with normal hearing were recruited ($N=27$) to participate in the study. Their hearing thresholds from 250 to 8000 Hz were tested in octave intervals prior to admission in the study. The hearing sensitivity of the listeners was within 20 dB HL at all test frequencies and they had normal middle ear functions.

Group I [NH(Mod4)] consisted of 15 listeners who listened to 4-channel CI simulation. The data for 3 listeners were excluded in the final analysis because their scores in one or more hearing aid settings (mainly Om and ON) were so poor that SNR_50 could not be estimated. Subsequently, 12 listeners with normal hearing were recruited in Group II [NH(Mod8)] to listen to the 8-channel CI simulated speech. The mean ages for the final Group I and Group II listeners were 21.3 and 20.7 years old, respectively.

2. Speech recognition in noise for CI simulations

The speech testing materials recorded in Experiment I were processed by a MATLAB program based on the algorithms used in the experiments conducted by Shannon *et al.* (1995). This program extracted and preserved the temporal envelope cues of the speech sentences to 4 or 8 channels and, at the same time, eliminated spectral fine structures within the channels. Briefly, the MATLAB program divided the stimuli into four or eight spectral filter bands by using bandpass filters. The cut-off frequencies of these bandpass filters were calculated from the Greenwood map (1990), which was intended to divide the tonotopically arranged basilar membrane into equal distances and map the corresponding physical frequency range accordingly. The final wave form was derived from the sum of the temporal envelopes after each channel of filtered stimuli was full-wave rectified, low-pass

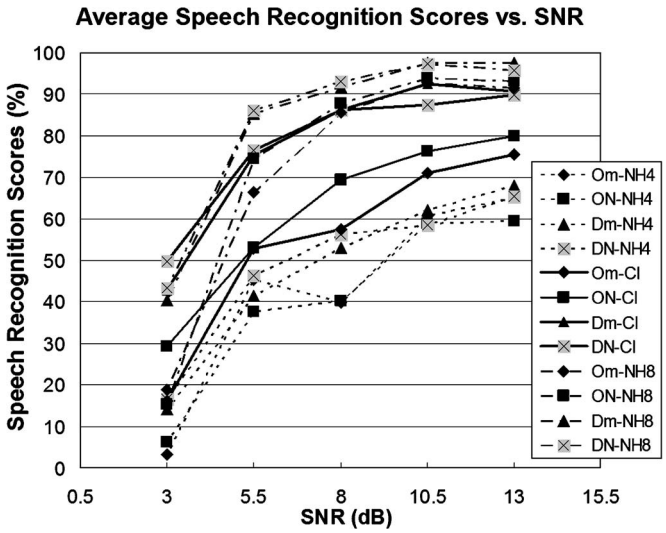


FIG. 6. The average speech recognition scores of cochlear implant users and normal hearing individuals listening to 4 and 8 channels of cochlear simulations in omni-directional microphone (Om), omni-directional microphone plus noise reduction algorithms (ON), directional microphones (Dm), and directional microphones plus noise reduction algorithms (DN) conditions at 5 SNRs.

filtered at 400 Hz, multiplied by a white noise, and then filtered by its corresponding filter again.

During testing, the CI simulated speech materials were presented from a computer to a GSI 61 Clinical Audiometer (Grason-Stadler, Inc). Listeners with normal hearing listened to the stimuli from a pair of ER-3A insert earphones at 65 dB HL. Ten CI simulated sentences in quiet were given for practice prior to testing. A verbal feedback of the correct response was given after they listened to the practice sentences but not during actual testing.

B. Results

The average scores of the two groups of normal hearing individuals and the 16 CI users are plotted in Fig. 6 and the standard deviations are tabulated in Table II. The average scores of CI users fell between those of the NH(Mod4) and NH(Mod8) groups and the standard deviations of the CI users were relatively higher than those of the normal hearing users. It is also noteworthy that the speech recognition scores obtained in Dm and DN for CI users were equal to or exceeded the scores obtained in Om and ON for the NH(Mod8) group.

The percent correct scores of normal hearing individuals obtained at different SNRs were converted to SNR_50s. A repeated measure ANOVA was also performed on hearing aid settings for each listener group. No significant main effect was found for the NH(Mod4) listeners ($p > 0.05$). However, a significant main effect of hearing aid settings was found for the NH(Mod8) listeners ($F[15,3]=18.3, p < 0.0001$).

The average SNR_50s for the NH(Mod8) were 4.5, 4.4, 3.4, and 3.3 dB for Om, ON, Dm, and DN. Scheffe pairwise comparisons were carried out to examine significant differences among the hearing aid settings for the NH(Mod8) group. Significant differences were found between Om and Dm, Om and DN, ON and Dm, ON and DN ($p < 0.0083$).

TABLE II. The standard deviations of speech recognition scores of CI users (CI) and normal listeners listening to 8 (Mod8) and 4 (Mod4) channels of cochlear implant simulation.

Group	SNR				
	13.0	10.5	8.0	5.5	3.0
			Om		
Mod(8)	9.2%	6.2%	8.2%	13.4%	11.8%
CI	20.1%	21.6%	30.8%	22.2%	14.3%
Mod(4)	12.4%	11.6%	17.5%	16.1%	3.8%
			ON		
Mod(8)	5.7%	3.9%	7.2%	13.2%	11.3%
CI	17.4%	21.8%	27.2%	20.7%	21.3%
Mod(4)	12.1%	9.2%	8.3%	18.0%	4.2%
			Dm		
Mod(8)	2.6%	4.1%	6.9%	11.3%	9.9%
CI	20.4%	23.2%	23.8%	24.8%	26.2%
Mod(4)	14.2%	19.2%	10.9%	10.9%	6.8%
			DN		
Mod(8)	5.9%	3.8%	6.5%	10.4%	16.3%
CI	13.8%	19.7%	18.2%	18.6%	25.1%
Mod(4)	14.2%	16.4%	19.3%	10.5%	7.6%

The critical difference was 0.8 dB. No significant difference was found between Om and ON or between Dm and DN.

C. Discussion

The results of this experiment indicate that the directional microphone enhanced the speech recognition ability of normal hearing individuals in noise when they listened to 8-channel CI simulation, but the noise reduction algorithm did not. These results were consistent with those obtained for the CI users in Experiment I. The speech recognition scores of this study were also consistent with previous research studies that the performance of CI users fell roughly between the performances of normal hearing individuals listening to 4 and 8 channels of CI simulated speech in noise (Friesen *et al.*, 2001; Stickney and Zeng, 2004; Zeng *et al.*, 2005). One exciting finding is that when CI users listened to speech processed by Dm and DN, their speech recognition scores exceeded or equaled those of the NH(Mod8) listeners in the Om and ON conditions, especially at low SNRs.

The results of NH(Mod4) showed no significant difference between any hearing aid setting for normal hearing individuals when they listened to 4-channel CI simulation. Previous studies showed that the performance of normal hearing individuals listening to 4- and 6-channel CI simulated speech was similar to the performance of CI users with an equal number of speech processing channels (Fu *et al.*, 1998; Dorman *et al.*, 1997; Dorman and Loizou, 1998). This suggests that the extra spectral information provided by the 8-channel simulation made it a more sensitive tool for detecting changes in signal processing strategies for this mixed group of multichannel CI users whose speech processors have more than 4 signal processing channels.

IV. EXPERIMENT III

In this experiment, the applicability of a speech-based STI program to predict the speech intelligibility of speech

processed by a directional microphone and an AMC noise reduction algorithm was explored. The calculated STIs were also used as indications of the amount of improvement in temporal envelope modulations in the processed signal to shed light on the factors that determine speech intelligibility.

A. Materials and Methods

1. Speech-based STI calculation method

There are at least four originally proposed speech-based STI programs, namely the normalized covariance method by Koch (1992) and Holube and Kollmeier (1996), envelope regression method by Ludvigsen *et al.* (1990), real cross-power spectrum method by Drullman *et al.* (1994), and magnitude cross-power spectrum method by Payton *et al.* (1994, 1999, 2002). These methods differ in how the transmission indexes, or the modulation depth of each frequency band, are estimated.

The speech-based STI calculation method proposed by Payton *et al.* (2002) was used in this study because Goldsworthy and Greenberg (2004) reported that this method produced STI values that are closest to those calculated by the original non-speech-based STI by Hougast and Steeneken (1985) if the transmitted signal was degraded in acoustic environments. The STI was calculated from the transmission indexes of speech at seven filter bands centered at 125, 250, 500, 1000, 2000, 4000, and 8000 Hz multiplied by the speech weighting of the corresponding frequency band. In this study, the probe was a sound file with 30 concatenated HINT sentences in quiet (i.e., the reference sound file) and the transmitted signals were 20 speech-in-noise sound files processed by Om, ON, Dm, and DN at SNR of +3, +5.5, +8, +10.5, and +13 dB (i.e., the processed sound files).

2. Recording of processed sound files

New recordings of the processed signals with concatenated sentences were made because the recordings from Experiment I had noise between sentences. Before the recording of the processed signals, a 300 ms 1000 Hz tone was placed 700 ms before and after the first and last sentences in the reference sound file. These tones served as markers to mark the beginning and the end of the concatenated sentence stream. The same calibration and recording procedures used in Experiment I were carried out to record the processed sound files. After the recordings were made, the marker tones and the silence were removed from the reference sound file. In the processed sound file, the noise before the first marker, the first marker tone, and the 700 ms noise were removed from the beginning of the sentence stream, and the 700 ms and the second marker tone were removed from the end of the sentence stream. These procedures generated a reference and processed signals with exact length. The STIs were then calculated by comparing the modulation depth in the reference and the processed files.

B. Results

Three-parameter sigmoidal functions were used to fit the STIs and speech recognition scores obtained in the Om and Dm conditions for the three groups of listeners (Fig. 7). The

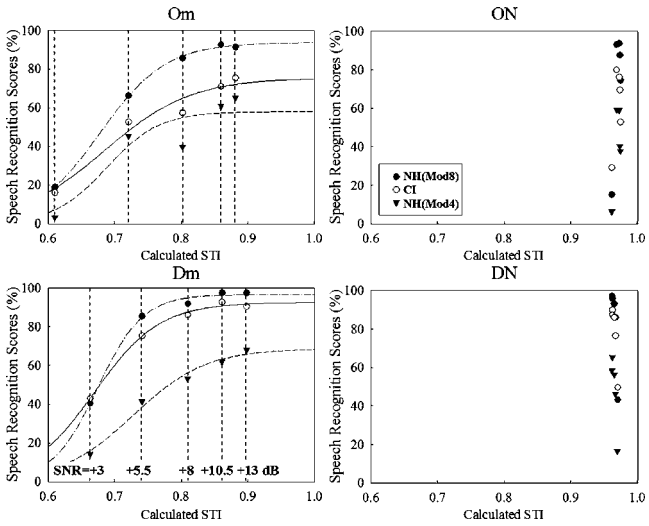


FIG. 7. The relationship between the average speech recognition scores and the calculated speech transmission indexes (STIs) at the four hearing aid settings for the three groups of listeners at 5 SNRs.

speech recognition scores increased monotonously with STIs, indicating that STI could be applied to predict the speech recognition scores of CI users and normal hearing individuals in these conditions.

No attempt was made to fit the scores obtained in the ON and DN conditions because the STI values for these conditions only differ by 0.0129. Fitting the data based on such a small range on one variable may bear little practical relevance and lead to erroneous conclusions.

C. Discussion

The monotonous increase in STI from low to high SNR conditions in the Om and Dm condition indicated that the modulation depth of the temporal envelopes of the processed sentences increased as the SNR increased (see the dashed lines in Fig. 7). The rapid increase in STIs at low SNRs and slower increase at high SNRs indicates that the STI predicts a greater improvement in speech recognition scores for a reduction in noise at a low SNR than the same reduction in noise at a high SNR. In this study, the speech recognition scores were in general agreement with the STI prediction in the Om and Dm conditions. Therefore, we concluded that STI was a good predictor of acoustic degradations and directional microphones for CI users. These results were also consistent with findings of studies involving listeners with acoustic hearing (Steeneken and Houtgast, 1982; Payton *et al.*, 1994; Goldworthy, 2005; Ricketts and Hornsby, 2003).

In contrast, the ranges of calculated STIs were 0.0129 for ON and 0.0082 for DN across the SNRs while the ranges of the speech recognition scores varied between 40.1% and 77.8% for the three groups of listeners. These results in electrical hearing parallel the findings of previous studies in acoustic hearing that STI is a poor predictor of speech recognition scores for nonlinearly processed speech (Ludvigsen *et al.*, 1993; Drullman, 1995; Hohmann and Kollmeier, 1995; Goldsworthy, 2005). The extremely narrow range of STIs suggests that the modulation depth of the speech temporal envelope is almost identical across the SNRs. This finding is

also consistent with the hearing aid manufacturer's descriptions of the noise reduction algorithm, which increased gain reduction as the level of noise increased. In other words, higher gain reductions were applied to the signal as the SNR decreased at frequency channels with noise dominance and thus the hearing aid output had similar temporal envelope modulations at all SNRs.

The relationships between STIs and speech recognition scores obtained in different acoustic and signal processing conditions suggests that one of the determining factors for speech understanding prediction is the within-channel SNR. Although both directional microphone and AMC noise reduction algorithm increased the modulation of temporal envelope of speech in background noise, only the directional microphone significantly enhanced the speech intelligibility of CI users and normal hearing individuals listening to CI simulated speech. As mentioned in Sec. I, directional microphones are more sensitive to sounds from the front than the sides or the back. When the testing materials were recorded with speech presented from the front and noise presented from all around, the SNR across frequency regions and within frequency channels was improved in the Dm condition compared to the Om condition.

The noise reduction algorithm, on the other hand, does not improve within-channel SNR because any gain reduction applied to the frequency channels would have affected both speech and noise. The fact that the speech recognition scores of ON and DN increased as SNR increased, instead of reaching a plateau as predicted by the calculated STIs, suggests that one of the determining factor for speech recognition is the within-channel SNR but not the overall SNR of the signal as assumed by STI. Our results in the Om conditions (i.e., speech understanding decreased with acoustic degradation) also support this notion.

Further, the above-presented conclusion appeared to be indirectly supported by studies investigating noise reduction algorithms using spectral subtraction. Spectral subtraction is a noise reduction strategy in which a speech in noise signal is transformed to the frequency domain, the estimated noise spectrum is subtracted from the speech in noise signal, and the speech with reduced noise signal is then converted back to the time domain. If the noise reduction algorithm could accurately estimate the noise spectrum, the within-channel SNR is improved in the processed signal. Several research groups have implemented the spectral subtraction noise reduction algorithm and reported improved speech recognition scores for CI users (Weiss, 1993; Hochberg *et al.*, 1992; Goldsworthy, 2005).

V. SUMMARY AND CONCLUSIONS

Directional microphones and AMC noise reduction algorithms are two noise reduction strategies commonly used to reduce noise interferences in hearing aids. Both of these strategies increased the temporal modulation of speech envelope in background noise. The goals of this study were to determine whether directional microphones and AMC noise reduction algorithm could enhance speech recognition and/or sound quality in background noise, and to examine if a speech-based speech transmitted index could predict the speech intelligibility of CI users and normal hearing individuals listening to CI simulated speech. The long-term goal of this study was to investigate suitable signal processing strategies for enhancing CI performance.

The results of this study are encouraging: the directional microphone significantly enhanced speech recognition ability and overall preferences of CI users in background noise. Although the AMC noise reduction algorithm did not significantly improve speech recognition, it significantly improved cochlear implant users' sound quality ratings. Taken together, the rankings of speech recognition are Om=ON<Dm=DN and the rankings of overall preferences are Om<ON=Dm<DN. Overall, DN is the most desirable and Om is the least.

The positive findings in this study suggest that other advanced hearing aid technologies may also be implemented as preprocessors to CI speech processors to enhance CI performance and user convenience. Some examples include the microphone matching algorithm that can maintain directional effects of directional microphones over time, the switchless telecoil that can sense the magnetic field emitted by telephone headsets and automatically switch to telecoil input, before switching back to microphone input if the telephone headset is removed from the ear.

Previous studies reported that directional microphone performance may be reduced in real-life environments with reverberations (Hawkins and Yacullo, 1984; Ricketts and Hornsby, 2003). Some hearing aid studies reported that some directional hearing aid users obtained significant improvement in laboratory testing environments but they did not notice significant benefit in everyday life environments (Cord *et al.*, 2002; Mueller *et al.*, 1983; Ricketts and Hornsby, 2003; Surr *et al.*, 2002; Walden *et al.*, 2000). As the recordings of the present study were made in an anechoic chamber (i.e., little reverberation), field trials should be conducted to further evaluate the efficacy of directional microphones for CI users. In addition, as noise reduction algorithms rely on the differences in physical characteristics of speech and noise to separate speech and noise, the AMC noise reduction algorithm may not be effective if the background noise is speech. Further improvements are still needed to optimize the noise reduction algorithms to operate effectively in more acoustically complex environments.

Although STI used a mechanism similar to CI speech coding strategies to predict speech understanding, it successfully predicted the effects of noise and directional microphones but failed to predict the speech recognition scores of CI users for the AMC noise reduction algorithm used in this study. The results of this study also suggested that the within-channel SNR, instead of the overall level of noise, is the determining factor for speech understanding for CI users.

A caution in interpreting the results of this study is that the within-channel SNR in the CI speech processor depends on the number of signal processing channels in the AMC noise reduction algorithm and the CI speech processor. If an AMC noise reduction algorithm divided and processed signals in, say, 6 channels and then the processed signal is sent to a CI speech processor with 6 channels, the within-channel SNR in frequency channels with speech components may not

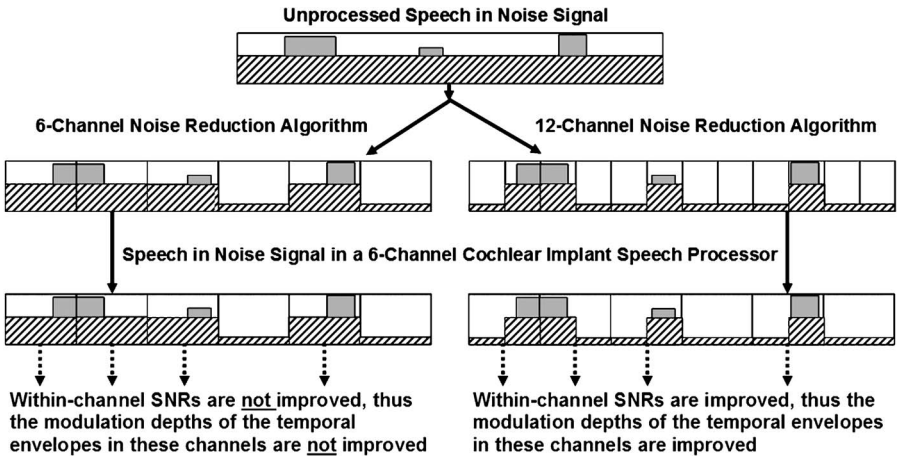


FIG. 8. The effects of the relative number of channels in the noise reduction algorithm and cochlear implant speech processor. Assume that this particular noise reduction algorithm does not reduce the gain of a frequency channel if speech, regardless of SNR, is detected in the channel. If the noise reduction algorithm has the same or lower number of channels than the cochlear implant speech processor (the example on the left side), the overall modulation of the temporal envelope is increased but the within-channel SNRs are not. On the other hand, if the noise reduction algorithm has higher number of channels than the cochlear implant speech processor (the example on the right side), both overall modulation of the temporal envelope and within-channel SNRs are increased.

be improved for the CI user because there are more channels in the speech processor than the noise reduction algorithm (Fig. 8). The reverse (i.e., the AMC noise reduction algorithm has 12 channels and the speech processor has 6 channels), however, could provide an increase in within-channel SNR because the noise reduction algorithm has a finer processing scale (Fig. 8). By the same logic, the within-channel SNR in the CI speech processor also depends on the relative spectrum of speech and noise as well as the cut-off frequencies of the frequency channels in the noise reduction algorithm and the speech processor. Therefore, the results of this study do not rule out the possibility that an AMC noise reduction algorithm with narrower frequency bands or more channels may enhance speech understanding of CI users with speech processor with fewer processing channels.

Further, if the above-mentioned assumptions hold, they support the implementation of noise reduction algorithms with more signal processing channels—as many channels as the signal processing power of the digital platform permits without significant overall processing delay—than CI speech processors. As implemented in the present form, it is possible that a noise reduction algorithm with higher number of channels can improve the within-channel SNR of a CI speech processor. Future studies are needed to determine if the effectiveness of the AMC noise reduction algorithm varies for different CI speech coding/electrical stimulation strategies and if noise reduction algorithms with a different number of signal processing channels would be more effective for CI users.

ACKNOWLEDGMENTS

We would like to thank Rachael Fischer for data collection. We also want to thank Sheng Liu and Tiffany E. Chua for providing software support on the speech-based STI programs and Ray Goldsworthy for sharing his version of speech-based STI program and helpful discussion on the noise reduction algorithms using spectral subtraction. This work is supported in part by NIH (2R01 DC002267).

Amlani, A. M. (**2001**). "Efficacy of directional microphone hearing aids: A meta-analytic perspective," J. Am. Acad. Audiol **12**, 202–214.

Bentler, R. A. (**2005**). "Effectiveness of directional microphones and noise reduction schemes in hearing aids: A systematic review of the evidence," J. Am. Acad. Audiol **16**, 473–484.

Bray, V., and Nilsson, M. (**2001**). "Additive SNR benefits of signal processing features in a directional DSP aid," Hear. Rev. **8**, 48–51, 62.

Chung, K., Zeng, F.-G., and Waltzman, S. (**2004a**). "Using hearing aid directional microphones and noise reduction algorithms to enhance cochlear implant performance," ARLO **5**, 56–61.

Chung, K., Zeng, F.-G., and Waltzman, S. (**2004b**). "Utilizing hearing aid directional microphones and noise reduction algorithms to improve speech understanding and listening preferences of cochlear implant users," Intl. Congress Series **1273**, 89–92.

Chung, K. (**2004**). "Challenges and recent developments in hearing aids. I. Speech understanding in noise, microphone technologies and noise reduction algorithms," Trends Amp **8**, 83–124.

Cord, M. T., Surr, R. K., Walden, B. E., and Olsen, L. (**2002**). "Performance of directional microphone hearing aids in everyday life," J. Acoust. Soc. Am. **13**, 295–307.

Cox, R. M., and Bisset, J. D. (**1984**). "Relationship between two measures of aided binaural advantage," J. Speech Hear Disord. **49**, 399–408.

Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (**1997**). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels," J. Acoust. Soc. Am. **104**, 3583–3585.

Dorman, M. F., and Loizou, P. C. (**1998**). "The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels," Ear Hear. **19**, 162–166.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of reducing slow temporal modulations on speech recognition," J. Acoust. Soc. Am. **95**, 2670–2680.

Drullman, R. (**1995**). "Temporal envelope and fine structure cues for speech intelligibility," J. Acoust. Soc. Am. **97**, 585–592.

Edwards, B. W., Struck, C. J., Dharan, P., and Hou, Z. (**1998**). "New digital processor for hearing loss compensation based on the auditory system," Hear. J. **51**, 38–49.

Figueiredo, J. C., Abel, S. M., and Papsin, B. C. (**2001**). "The effect of the audallion BEAMformer noise reduction preprocessor on sound localization for cochlear implant users," Ear Hear. **22**, 539–547.

Friesen, L. M., Shannon, R. V., Baskent, D., and Want, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels:

Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.
Fu, Q. J., Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.
Fu, Q. J., Chinchilla, S., and Galvin, J. J. (**2004**). "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users," J. Assoc. Res. Otolaryngol. **5**, 253–260.
Gabrielsson, A., Schenkman, B., and Hagerman, B. (**1988**). "The effects of different frequency responses on sound quality judgements and speech intelligibility," J. Speech Hear. Res. **31**, 166–177.
Goldsworthy, R. L., and Greenberg, J. E. (**2004**). "Analysis of speech-based Speech Transmission Index methods with implications for nonlinear operations," J. Acoust. Soc. Am. **116**, 3679–3689.
Goldsworthy, R. L. (**2005**). *Noise Reduction Algorithms and Performance Metrics for Improving Speech Reception in Noise by Cochlear-Implant Users* (MIT, Cambridge).
Grose, J. H., and Hall, J. W., III (**1992**). "Comodulation masking release for speech stimuli," J. Acoust. Soc. Am. **91**, 1042–1050.
Hawkins, D. B., and Yacullo, W. S. (**1984**). "Signal-to-noise ratio advantage of binaural hearing aids and directional microphones under different levels of reverberation," J. Speech Hear Disord. **49**, 278–286.
Hochberg, I., Boothroyd, A., Weiss, M., and Hellman, S. (**1992**). "Effects of noise and noise suppression on speech perception by cochlear implant users," Ear Hear. **13**, 163–271.
Hohmann, V., and Kollmeier, B. (**1995**). "The effect of multichannel dynamic compression on speech intelligibility," J. Acoust. Soc. Am. **97**, 1191–1195.
Holube, I., and Kollmeier, K. (**1996**). "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated speech perception model," J. Acoust. Soc. Am. **100**, 1703–1715.
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.
Independent Hearing Aid Fitting Forum (**1994**). "*Comprehensive hearing aid fitting protocol for the 21st century*," Jackson Hole, WY.
Johns, M., Bray, V., and Nilsson, M. (2003). "Effective noise reduction," Audiol. Online 1-3-2003.
Kwon B. J. (**2002**). "Comodulation masking release in consonant recognition," J. Acoust. Soc. Am. **112**(2), 634–641
Koch, R. (**1992**). "Gehörgerechte schallanalyse zur vorhersage und der sprachverständliichkeit (Auditory sound analysis for the prediction and improvement of speech intelligibility)," Universität Göttingen.
Kuk, F., Ludvigsen, C., and Paludan-Muller, C. (**2002**). "Improving hearing aid performance in noise: Challenges and strategies," Hear. J. **55**, 34–46.
Loizou, P. (**1998**). "Mimicking the human ear: An overview of signal processing techniques used for cochlear implants," Ear Hear. **15**, 101–130.
Ludvigsen, C., Elberling, C., Keidser, G., and Poulsen, T. (**1990**). "Prediction of intelligibility on non-linearly processed speech," Acta Oto-Laryngol., Suppl. **469**, 190–195.
Ludvigsen, C., Elberling, C., and Keidser, G. (**1993**). "Evaluation of a noise reduction method - Comparison of measured scores and socres predicted from STI," Scand. Audiol. Suppl. **38**, 50–55.
Moore, B. C. (**1990**). "Co-modulation masking release: Spectro-temporal pattern analysis in hearing," Br. J. Audiol. **24**, 131–137.
Mueller, H. G., Grimes, A. M., and Erdman, S. A. (**1983**). "Directional microphone," Hear. Instrum. **34**, 14–16, 47–48.
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.
Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of a hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.
Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95**, 1581–1592.
Payton, K. L., and Braida, L. D. (**1999**). "A method to determine the speech transmission index from speech waveforms," J. Acoust. Soc. Am. **106**, 3637–3648.
Payton, K. L., Braida, L. D., Chen, S., Rosengard, P., and Goldsworthy, R. (**2002**). "Computing the STI using speech as a probe stimulus," in TNO Human Factors (ed.) *Past, Present and Future of the Speech Transmission Index* (Soesterberg, The Netherlands), pp. 125–138.
Plomp, R. (**1983**). "Perception of speec s a modulated signal," in *Proceedings of the Tenth International Congress of Phonetic Sciences*, Utrecht, The Netherlands, 1–6 August, The Congress, pp. 29–40.
Powers, T. A., Holube, I., and Wesselkamp, M. (**1999**). "The use of digital filters to combat background noise," in*High Performance Hearing Solutions*, edited by S. Kochkin and K. E. Strom [Hear Rev, **3**(suppl), 36–39].
Powers, T. A., and Hamacher, V. (**2002**). "Three-microphone instrument is designed to extend benefits of directionality," Hear. J. **55**, 38–45.
Ricketts, T. A. (**2001**). "Directional hearing aids," Trends Amp. **5**, 139–176.
Ricketts, T. A., and Hornsby, B. W. (**2003**). "Distance and reverberation effects on directional benefit," Ear Hear. **24**, 472–484.
Ricketts, T. A., and Hornsby, B. W. Y. (**2005**). "Sound quality measures for speech in noise through a commercial hearing aid implementing "Digital Noise Reduction"," J. Am. Acad. Audiol **16**, 270–277.
Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Steeneken, H. J. M., and Houtgast, T. (**1982**). "Some applications of the speech transmission index (STI) in auditoria," Acustica **51**, 229–234.
Stickney, G. S., and Zeng, F.-G. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.
Surr, R. K., Walden, B. E., Cord, M. T., and Olsen, L. (**2002**). "Influence of environmental factors on hearing aid microphone preference," J. Am. Acad. Audiol **13**, 308–322.
Valente, M., Schuchman, G., Potts, L. G., and Beck, L. B. (**2000**). "Performance of dual-microphone in-the-ear hearing aids," J. Am. Acad. Audiol **11**, 181–189.
Van Tasell, D. J., Soli, S., Kirby, V., and Widin, G. (**1987**). "Speech waveform envelop cues for consonant recognition," J. Acoust. Soc. Am. **82**, 1152–1161.
Van Tasell, D. J., Greenfield, D. G., Logemann, J. J., and Nelson, D. A. (**1992**). "Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants," J. Acoust. Soc. Am. **92**, 1247–1257.
Walden, B. E., Surr, R. K., Cord, M. T., Edwards, B., and Olson, L. (**2000**). "Comparison of benefits provided by different hearing aid technologies," J. Am. Acad. Audiol **11**, 540–560.
Weiss, M. R. (**1993**). "Effects of noise and noise reduction processing on the operation of the Nucleus-22 cochlear implant processor," J. Rehabil. R. D **30**, 117–128.
Wouters, J., Litere, L., and Van Wieringen, A. (**1999**). "Speech intelligibility in noisy environments with one and two microphone hearing aids," Audiology **38**, 91–98.
Zeng, F.-G. (**2004**). "Trends in Cochlear Implants," Trends Amp. **8**, 1–34.
Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. U.S.A. **102**, 2293–2298.

Acoustic roles of the laryngeal cavity in vocal tract resonance

Hironori Takemoto, Seiji Adachi, Tatsuya Kitamura, Parham Mokhtari, and Kiyoshi Honda
ATR Human Information Science Laboratories, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan

(Received 23 August 2005; revised 9 July 2006; accepted 10 July 2006)

The acoustic effects of the laryngeal cavity on the vocal tract resonance were investigated by using vocal tract area functions for the five Japanese vowels obtained from an adult male speaker. Transfer functions were examined with the laryngeal cavity eliminated from the whole vocal tract, volume velocity distribution patterns were calculated, and susceptance matching analysis was performed between the laryngeal cavity and the vocal tract excluding the laryngeal cavity (vocal tract proper). It was revealed that the laryngeal cavity generates one of the formants of the vocal tract, which is the fourth in the present study. At this formant, the resonance of the laryngeal cavity (the 1/4 wavelength resonance) induces the open-tube resonance of the vocal tract proper (the 3/2 wavelength resonance). At the other formants, on the other hand, the vocal tract proper acts as a closed tube, because the laryngeal cavity has only a small contribution to generating these formants and the effective closed end of the whole vocal tract is the junction between the laryngeal cavity and the vocal tract proper. [DOI: 10.1121/1.2261270]

PACS number(s): 43.70.Bk, 43.70.Aj [BHS] Pages: 2228–2238

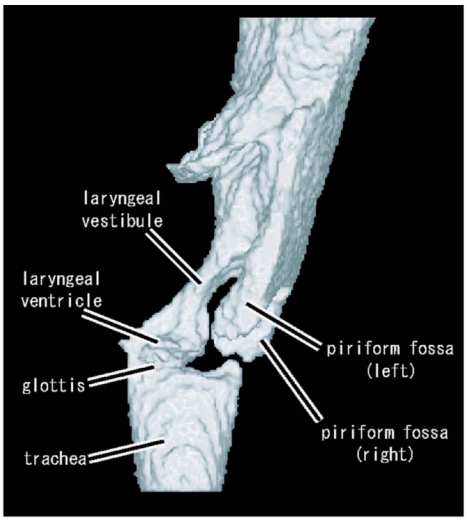
I. INTRODUCTION

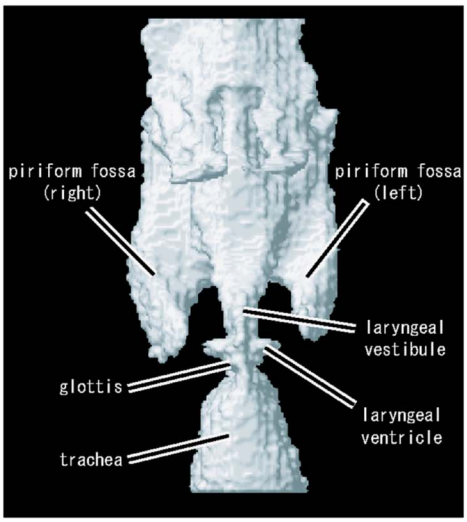
The hypopharyngeal cavities are located at the bottom of the vocal tract, and consist of the bilateral cavities of piriform fossae and the laryngeal cavity [Fig. 1(a)]. The piriform fossae are two small pockets behind the larynx and are located just above the entrance of the esophagus. Their acoustic properties were investigated in detail by Dang and Honda (1997): they each act as a side branch of the vocal tract to generate a spectral minimum in the frequency range from 4 to 5 kHz. The laryngeal cavity anatomically divides into three parts [Fig. 1(b)] (Williams *et al.*, 1995). The upper (vestibular) and middle (ventricular) regions are simply referred to as the laryngeal cavity in this study, because the lower (subglottal) region is excluded from the vocal tract. Although previous studies have reported that the fourth (F4) or fifth (F5) formant is sensitive to the laryngeal cavity shape (e.g., Fant, 1960), it is still unclear how the laryngeal cavity generates such a formant. In the present study, we focus on the laryngeal cavity to reveal its acoustic characteristics.

The acoustic characteristics of the laryngeal cavity have been studied mainly with respect to the sung voice. Bartholomew (1934) reported that a "high formant" could be observed around 2800 to 2900 Hz in sung vowels and hypothesized that the formant could be produced in the laryngeal cavity. Lewis (1936) observed that five vowels had a common, fourth formant close to 3200 Hz. He suspected that the formant was generated by some fixed cavity such as the laryngeal cavity, because the formant of a sung vowel ("AH") was dependent on voice pitch. Chiba and Kajiyama (1942), nonetheless, regarded the formant at 3200 Hz observed by Lewis (1936) as the resonance frequency of the laryngeal cavity, because the frequency matched with the resonance frequency computed from an effective laryngeal cavity of a tube length of 2.8 cm.

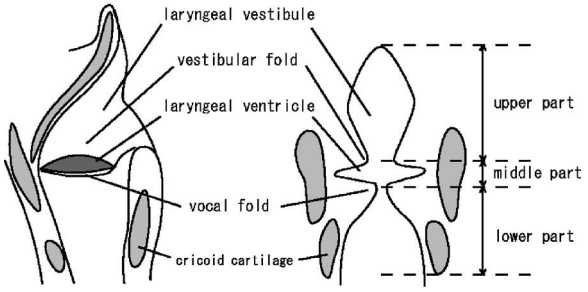
The formant reported by Bartholomew (1934) is now called the "singer's formant," which is a high spectrum envelope peak around 2.8 kHz, typically observed in western operatic singing by male singers. Sundberg has investigated the acoustical and physiological mechanism of generating the singer's formant in association with the laryngeal cavity (Sundberg, 1974, 1987, 1995, 2003). Sundberg (1974) explained that the singer's formant is a cluster of the third (F3), fourth (F4), and fifth (F5) formants. To generate the singer's formant, the vocal tract must satisfy two conditions: the cross-sectional area in the pharynx must be at least six times wider than that of the exit of the laryngeal cavity, and the laryngeal ventricle must be widened. The former condition enables the laryngeal cavity to act as a separate resonator (Sundberg, 1974, after Ingard, 1953), and the latter condition tunes the formant location. Based on small-perturbation analyses of area functions, Sundberg (1995) reported that F4, in particular, but also F5 of the sung vowels /a/ and /i/, is highly sensitive to the area function of the laryngeal cavity, although other regions of the area function also affect the locations of these formants.

Researchers have also investigated the acoustic characteristics of the laryngeal cavity outside the context of the singer's formant. Fant (1960) performed perturbation analyses of the area functions of the six Russian vowels and concluded that the laryngeal cavity has a marked influence on the F4 of /a/, /u/, and /i/, and on F5 of all the vowels except for /i/. Fant and Pauli (1974) calculated acoustic energy distributions within the vocal tract and noted that the laryngeal cavity was the source of F4. More recently, Fant and Båvegård (1997) reported that a shortening of the laryngeal cavity by 0.5 cm increases F5 greatly and F4 to a lesser extent. Our preliminary study (Takemoto *et al.*, 2003) revealed that the expansion of the laryngeal ventricle or constriction of the vestibule reduced F4 frequency, while constriction of the ventricle or expansion of the vestibule increased the F4 frequency.





(a) hypopharyngeal cavities



(b) laryngeal cavity

FIG. 1. Geometry of the laryngeal cavity and its anatomical relation. (a) Lateral and front view of the hypopharyngeal cavities. (b) Anatomical regions of the laryngeal cavity.

The previous studies described above suggest that the laryngeal cavity is responsible for one or two formants of the vocal tract and that these formants are exclusively sensitive to changes in laryngeal cavity shape, both in sung and spoken vowels. Although these acoustic properties of the laryngeal cavity have been reported, questions remain as to why

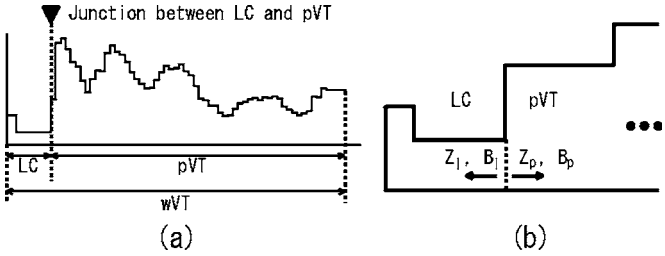


FIG. 2. Configuration of the vocal tract. (a) Vocal tract configuration. LC: laryngeal cavity, pVT: vocal tract proper, and wVT: whole vocal tract. (b) Positive direction of input impedance and susceptance for LC and pVT.

the changes in laryngeal cavity shape exclusively influence the formants and how the laryngeal cavity works in generating the formants. To address these problems, we conduct acoustic simulations while gradually altering the area function of the laryngeal cavity or completely removing it from the whole vocal tract, and assess the effects on the vocal tract transfer function. We then analyze the distribution pattern of the volume velocity in the vocal tract to reveal the contribution of the laryngeal cavity to the entire vocal tract resonance pattern.

II. MATERIALS AND METHODS

Using vocal tract area functions of the five Japanese vowels, three simulations described in Sec. II B are performed to reveal the acoustic effects of the laryngeal cavity on the vocal tract transfer functions. Three analyses shown in Sec. II C are also conducted to discuss the simulation results. Note that the piriform fossae are not implemented in these simulations and analyses to avoid spectral complexity and show clearly the acoustic effects of the laryngeal cavity. However, the acoustic effects on the spectrum are discussed at the end of the discussion. Figure 2(a) depicts the vocal tract configuration in terms of its area function. The laryngeal cavity is the region from the glottis to the junction at which its narrow exit is connected with the entrance of the wide pharyngeal cavity. The vocal tract excluding the laryngeal cavity is referred to as the vocal tract proper. In the present study, hereafter, the whole vocal tract, laryngeal cavity, and vocal tract proper are denoted as wVT, LC, and pVT, respectively. Additionally, the nth formant is denoted by Fn ($n=1,2,\ldots$), and where it is necessary to indicate which of wVT or pVT generates the formant, it is represented by Fnw or Fnp. In the simulations and analyses designed to investigate the acoustic interaction between LC and pVT, LC is modeled as a two-tube resonator whose geometry is shared among the five area functions.

A. Vocal tract area functions and laryngeal cavity model

Figure 3 shows the vocal tract area functions of the five Japanese vowels /a/, /i/, /u/, /e/, and /o/, and Table I represents numerical vocal tract area functions of the right and left piriform fossae. These area functions were extracted from three dimensional cinematographic magnetic resonance imaging (3D cine-MRI) data obtained from an adult native Japanese male (Takemoto *et al.*, 2006). Note that the vowel

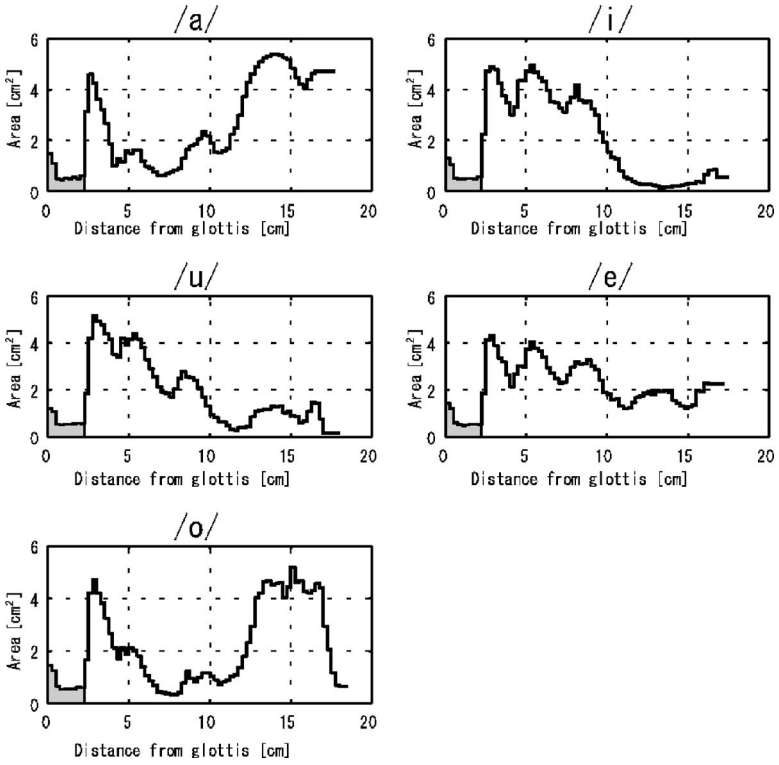


FIG. 3. Vocal tract area functions for the five Japanese vowels /a/, /i/, /u/, /e/, and /o/, measured in an adult male Japanese. The gray portion represents the laryngeal cavity.

/u/ was produced as a mid vowel /ɯ/, as the subject speaks Tokyo dialect. For all the vowels, the first and second sections in the area functions represent the ventricular region, while the vestibular region is represented from the third to the ninth sections. The overall mean and standard deviation of the ventricular area are 1.23 and 0.15 cm^2, respectively, and those of the vestibular area are 0.51 and 0.04 cm^2, respectively. Because of the small standard deviations, LC was reasonably modeled as a two-tube resonator; the length and area of the ventricular region are 0.5 cm and 1.2 cm^2, respectively, while those of the vestibular region are 1.75 cm and 0.5 cm^2, respectively.

B. Simulation

In the following simulations, the vocal tract transfer function is calculated using a transmission line model based on a model proposed by Adachi and Yamada (1999). The model includes viscothermal losses and yielding wall effects, and an infinite glottal impedance is assumed. Although the acoustical effects of the piriform fossae are not considered in the following simulations, they are discussed only in Sec. III B 5 by following the method of Dang and Honda (1997). The radiation impedance is approximated by the method proposed by Caussé *et al.* (1984). In the present study, the sound velocity c is set at 346.37 m/s, and air density ρ at 1.17 kg/m^3.

Calculation of the transfer function is based on the assumption of plane wave propagation in a vocal tract made up of a concatenation of short cylindrical sections. This is valid up to a certain frequency, at which the first transverse mode (its half wavelength resonance) appears. This upper-limit frequency is approximated by $c/2d$, where d is the maximum diameter among the cylindrical tubes. According to Fig. 3, because areas are less than 6 cm^2, d is 2.76 cm, from which

TABLE I. Equal interval (0.25 cm) area functions of the piriform fossae for the five vowels. Both right and left piriform fossae are connected with the vocal tract at the 11th section.

	/a/		/i/		/u/		/e/		/o/	
Sec No.	lt. PF	rt. PF	lt. PF	rt. PF	lt. PF	rt. PF	lt. PF	rt. PF	lt. PF	rt. PF
1	0.84	0.75	0.90	0.77	0.80	0.84	0.90	0.74	0.99	0.82
2	0.72	0.69	0.84	0.68	0.75	0.73	0.72	0.66	0.83	0.74
3	0.67	0.62	0.63	0.70	0.70	0.64	0.68	0.57	0.67	0.69
4	0.55	0.52	0.68	0.45	0.80	0.46	0.59	0.64	0.71	0.68
5	0.26	0.35	0.57	0.25	0.52	0.39	0.37	0.40	0.44	0.52
6	0.10	0.21	0.37	0.15	0.31	0.14	0.12	0.21	0.21	0.30
7					0.05					0.22
Total length	1.50	1.50	1.50	1.50	1.75	1.50	1.50	1.50	1.50	1.75

the upper-limit frequency may be determined as $34637/(2 \times 2.76) \cong 6275$ Hz. Therefore, we calculate transfer functions up to 6000 Hz.

1. Removing laryngeal cavity

The acoustic effect of eliminating LC from wVT was examined by comparing transfer functions of wVT and pVT. In a calculation of wVT's transfer function, the areas of LC were not modified. The first section of each tract is the input end with an infinite terminating impedance.

2. Changing ventricular area

The effect of changing the ventricular area was examined with the scale factors 1.5, 1.25, 1.0, 0.75, and 0.5 for each vowel, while the vestibular area was kept constant.

3. Changing vestibular area

The effect of changing the vestibular area was examined with the same scale factors for each vowel with the ventricular area being constant.

C. Analysis

1. Resonance mode at formant frequency

The distribution pattern of volume velocity along the vocal tract, referred to as "resonance mode," was computed for each formant to examine the locations of nodes and antinodes. If the volume velocity and pressure are given at the input end of a transmission line model, pressure $P_n(i)$ and volume velocity $U_n(i)$ are obtained at each section, where i is the section number and n is the formant number. In this simulation, the input volume velocity is set at 1.0, and the scale factor of LC is 1.0. The input pressure is equal to the input impedance Z_{in} and $U_n(i)$ is calculated relative to the input volume velocity.

2. Susceptance matching between LC and pVT

To explore the process by which LC gives rise to a formant, susceptance matching analysis was conducted for LC and pVT. Susceptance is defined as the imaginary part of admittance, i.e., the imaginary part of the reciprocal of input impedance. The normalized susceptance of pVT, B_p, and that of LC, B_l, are represented by

$$B_p = \text{Im}\left(\frac{Z_c}{Z_p}\right), \tag{1}$$

$$B_l = \text{Im}\left(\frac{Z_c}{Z_l}\right), \tag{2}$$

where Z_p is the input impedance of pVT viewed forward at the entrance of pVT, Z_l is the input impedance of LC viewed backward from the exit of LC [see Fig. 2(b)], and Z_c is the characteristic impedance of LC. Note that Z_l is calculated when the exit of LC is considered as the input end and the glottis as the terminal end where the radiation impedance is infinite. The positive direction of B_p and B_l is shown in Fig. 2(b). Z_c is calculated as $Z_c = \rho c / A_l$ where A_l is the area of the exit of LC ($0.5\ \text{cm}^2$).

Zero crossings of B_p indicate the formants of pVT, and intersections of B_p and $-B_l$ represent the formants of wVT (Stevens, 1998). At a zero of the input impedance, the susceptance curve has an asymptote. Although all the losses in the model can be assumed to be small, the susceptance curve may continue across the asymptote. The continuous part is removed in the figures to show the asymptotic line clearly. If a curve of B_p has no zero crossing because of losses, an imaginary value is used as the frequency of the point on B_p whose absolute value is minimum. If there is no intersection of B_p and $-B_l$, an imaginary one is used as the point on $-B_l$ whose vertical distance to B_p is minimum.

3. Energy concentration into laryngeal cavity at formant frequency

The acoustic energy concentration into LC, i.e., the ratio of the kinetic and potential energies in LC to those in wVT, was calculated to quantify LC's contribution to generating each formant. If pressure and volume velocity at a section are given, the kinetic energy KE and potential energy PE are obtained (Mrayati *et al.*, 1988; Story *et al.*, 2001 after Fant and Pauli, 1974). They are calculated by

$$\text{KE}_n(i) = \frac{1}{2}\frac{\rho l(i)}{a(i)} U_n(i)^2, \tag{3}$$

$$\text{PE}_n(i) = \frac{1}{2}\frac{a(i)l(i)}{\rho c^2} P_n(i)^2, \tag{4}$$

where $a(i)$ and $l(i)$ are the area and length of section i within a given area function, respectively. At the formant number n, the ratio Er_n of energy in LC to that in wVT is

$$\text{Er}_n = \frac{\sum_{i=1}^{Nlc} [\text{KE}_n(i) + \text{PE}_i(i)]}{\sum_{i=1}^{Nwvt} [\text{KE}_n(i) + \text{PE}_n(i)]} 100, \tag{5}$$

where Nlc is the number of sections in LC and $Nwvt$ is the total number of sections in wVT. Er_n quantifies the degree of energy concentration in LC. For each of the five vowels, Er_n is calculated at each of the first six formants in order to quantify the relative strength of the affiliation of each formant with LC.

4. Small perturbation of area function at formant frequency

To examine the acoustic influence of pVT on the formant provided by LC, a perturbation analysis was performed. From the kinetic and potential energies, the sensitivity function $S_n(i)$ in a vocal tract can be obtained by

$$S_n(i) = \frac{\text{KE}_n(i) + \text{PE}_n(i)}{\sum_{i=1}^{Nwvt} [\text{KE}_n(i) + \text{PE}_n(i)]}. \tag{6}$$

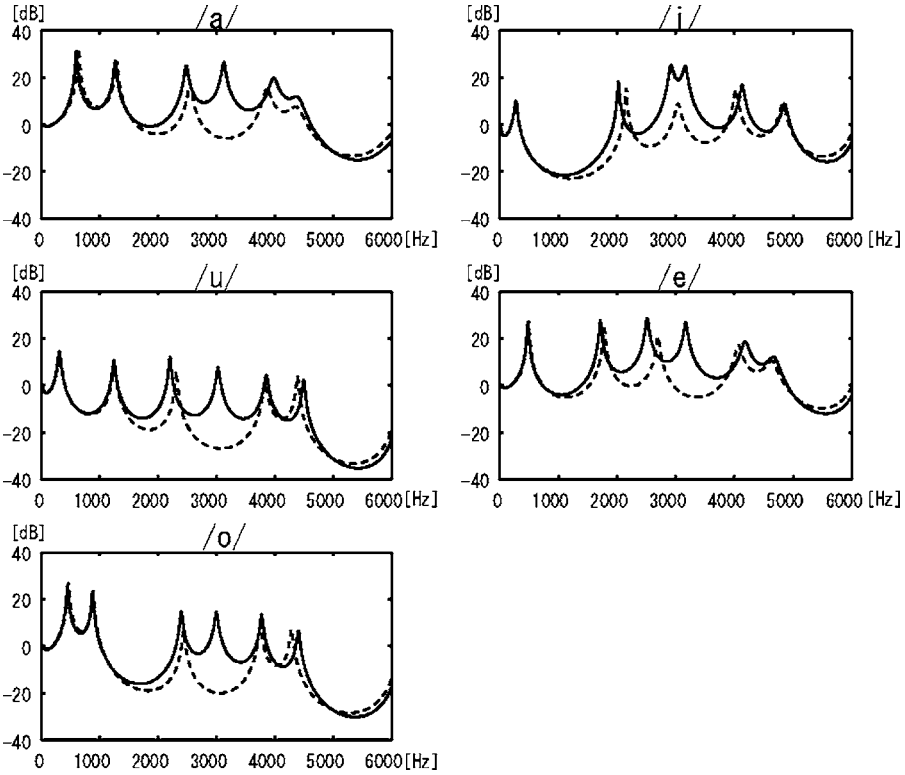


FIG. 4. Transfer functions of wVT (solid lines) and pVT (dashed lines).

Using the sensitivity function, the change in a particular formant frequency F_n due to a small perturbation to the area function $\Delta a(i)$ can be computed with the following relation:

$$\frac{\Delta F_n}{F_n} = \sum_{i=1}^{Nwvt} S_n(t)\frac{\Delta a(i)}{a(i)}, \tag{7}$$

where in the present study $\Delta a(i)$ is set at 5% of $a(i)$.

III. RESULTS AND DISCUSSION

A. Simulation results

In this section, we describe the results from three simulations with some discussion. The vocal tract transfer functions of the five vowels are computed when the modeled LC is modified by following the condition of each simulation.

1. Result of removing laryngeal cavity

Figure 4 gives the transfer functions of wVT and pVT to show acoustic effects caused by removing LC from wVT. In all the vowels shown in the figure, the elimination of LC also eliminates F4w, while retaining the other formants. Conversely, this indicates that LC supplies F4w while pVT is responsible for the other formants and overall spectral envelope. This is supported by the fact that when the radiation impedance is set at zero, the resonance frequency of LC is 3041 Hz, which is close to F4w listed in Table II. In the vowel /i/, although it is ambiguous which formant disappears at a glance, it is indeed F4w that disappears, because the remaining F3p (3035 Hz) is slightly closer to F3w (2918 Hz) than to F4w (3158 Hz), as shown in Tables II and III. With the disappearance of F4w, the lower formants F1w, F2w, and F3w move to higher frequencies, the higher formants F5w and F6w move to lower frequencies, and the level of the transfer function decreases in the frequency range from 2 to 4 kHz.

TABLE II. The first six formant frequencies (Hz) of wVT for the five vowels.

	/a/	/i/	/u/	/e/	/o/
F1w	598	275	311	480	445
F2w	1266	2021	1242	1717	879
F3w	2479	2918	2203	2514	2396
F4w	3123	3158	3023	3170	3000
F5w	3979	4131	3855	4178	3773
F6w	4365	4852	4488	4658	4406

2. Result of changing ventricular area

Figure 5 shows the vocal tract transfer functions for the five vowels while changing the scale factor of the ventricular area. In all the vowels, constriction of the ventricular area increases F4, while other formants are almost stable. This fact indicates that changes in LC shape selectively affect F4w.

TABLE III. The first five formant frequencies (Hz) of pVT for the five vowels.

	/a/	/i/	/u/	/e/	/o/
F1p	639	275	311	492	469
F2p	1289	2159	1248	1781	891
F3p	2543	3035	2314	2695	2438
F4p	3861	4020	3838	4066	3750
F5p	4354	4822	4389	4629	4283

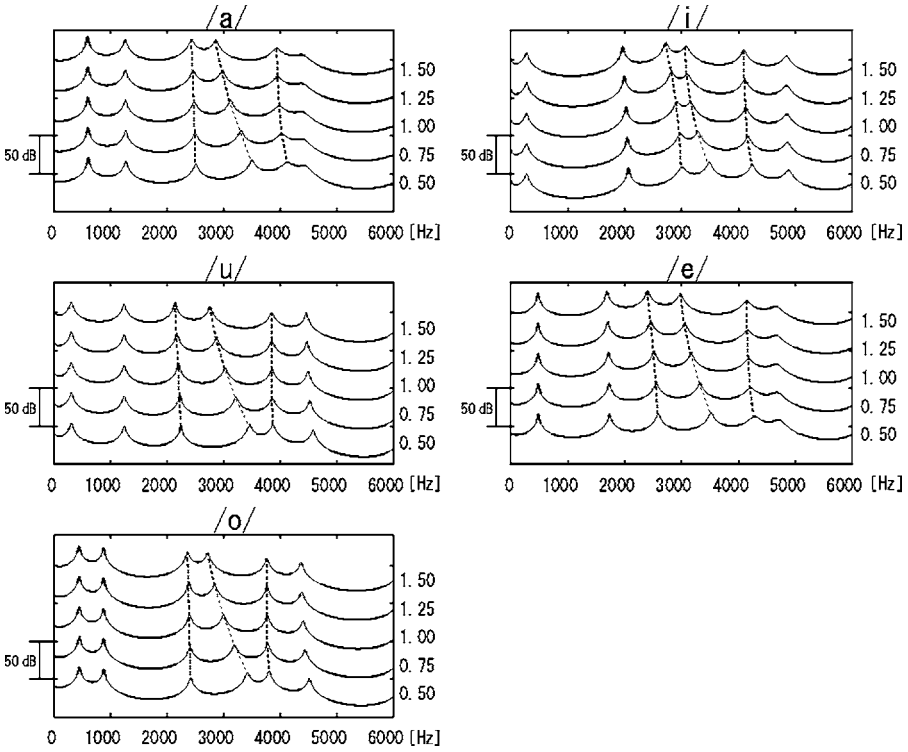


FIG. 5. Vocal tract transfer functions for five scale factors of the ventricular area. Peaks of F3, F4, and F5 across the five conditions are connected with dashed lines.

In the vowel /i/, however, F3 rises together with F4. Table IV lists the frequency, bandwidth, and amplitude of F3 and F4 of the vowel /i/ for each scale factor. In this table, the bandwidth is smaller in F3 than in F4 with a wide ventricle, and the amplitude is larger in F3 than in F4, while the relations are opposite with a narrow ventricle. According to Badin *et al.* (1990), a resonance affiliated with a vocal tract cavity can cross over an adjacent resonance due to changes in the cavity shape resulting in a formant exchange; nonetheless, the relative order of size of bandwidths and amplitudes is retained among the resonances. Table IV supports this principle and indicates that a formant exchange takes place between F3 and F4: in the vowel /i/, as the ventricular area is gradually constricted, LC resonance appears as F3 in the early stage and as F4 in the later stage. Thus, in all the vowels, constricting the ventricle area moves the formant affiliated with LC upward.

3. Result of changing vestibular area

Figure 6 shows the changes in transfer functions for the five vowels due to the changes of the vestibular area. As the area decreases, F4 also decreases, while other formants are almost stable. In the vowel /i/, both F3 and F4 move downward. The changes in bandwidths and amplitudes of F3 and F4 in the vowel /i/ (Table V) indicate that the formant exchange also occurs between F3 and F4: when the vestibular area is wide, the formant affiliated with LC appears as F4; when the vestibular area is narrow, the formant becomes F3. These results indicate that constricting the vestibular area moves the formant provided by LC downward in all the vowels.

TABLE IV. The bandwidths (B3 and B4, in Hz) and amplitudes (A3 and A4, in dB) of the third and fourth formants of the vowel /i/, when the laryngeal ventricular area is changed according to the scale factor.

Scale factor	1.50	1.25	1.00	0.75	0.50
B3	48.5	54.9	69.7	82.7	84.9
B4	87.7	86.2	76.3	62.4	56.4
A3	23.2	24.9	25.2	22.6	19.1
A4	18.2	21.6	25.0	25.8	25.3

B. Further analyses

The results of the first simulation indicate that LC generates F4, while pVT produces the other formants and shapes the spectral envelope. The results of the second and third simulations indicate that changes in LC shape influence almost exclusively F4 and that the acoustic effects of LC on F4 resemble those of a Helmholtz resonator: the ventricular area corresponds to the resonant cavity and the vestibular area to the neck. These three sets of results also imply that LC has a high degree of acoustic independence from pVT. In the following section, we conduct several analyses of vocal tract acoustics based on computations of perturbation, resonance mode, susceptance matching, and energy distribution, with the scale factor of LC fixed at 1.0.

1. Small perturbation of area function

Figure 7 shows for each vowel, the changes in F4 frequency when the area at each vocal tract section is increased by 5% independently of the other sections. The figure indicates that F4 is mainly sensitive to LC and insensitive to pVT. Focusing on LC, this figure supports the above results: an increase of the ventricular area (the first two sections) decreases F4 and that of the vestibular area (from the third to ninth sections) increases F4. These facts also indicate that the acoustic effects of LC on F4 can be compared to a Helmholtz

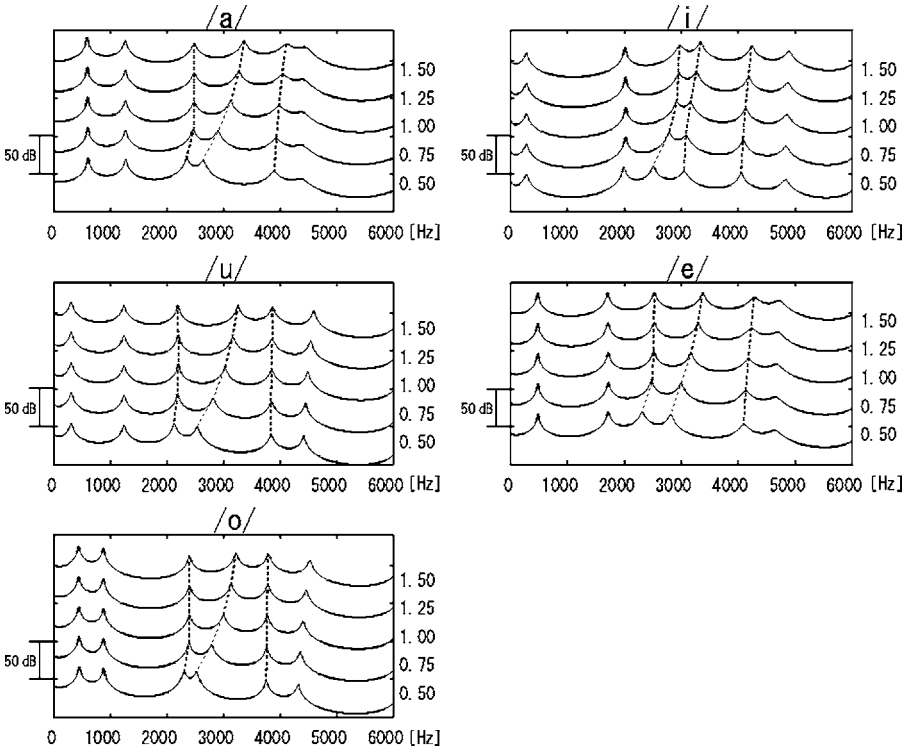


FIG. 6. Vocal tract transfer functions for five scale factors of the vestibular area. Peaks of F3, F4, and F5 across the five conditions are connected with dashed lines.

resonator. However, the resonance frequency of LC estimated with the Helmholtz equation was 3256 Hz with open-end correction or 3804 Hz without open-end correction, which is more than 200 Hz higher than the estimate from the transmission line model (3041 Hz).

2. Resonance mode at each formant

Figures 8–12 show the transfer function (TF), area function (AF), and the resonance mode diagram at each formant for the five vowels. The resonance mode is calculated at the six formants of wVT (F1w-F6w) and at the five formants of pVT (F1p-F5p). In the figures, the resonance modes at F1w, F2w, F3w, F5w, and F6w are drawn together with those at F1p, F2p, F3p, F4p, and F5p. Only at F4w is the resonance mode plotted separately.

As a general tendency, the volume velocity of wVT (solid lines) at the junction of LC and pVT gradually increases from F1w to F3w, reaches a maximum at F4w, then falls at F5w and F6w. At F4w, antinodes exist at both ends of pVT, which means that pVT resonates as an open tube at the formant close to the LC resonance. At the other formants, pVT can be regarded as a closed tube, because the volume velocity is relatively small at the LC-pVT junction, and the resonance mode of wVT agrees well with that of pVT (dashed line). These observations indicate that at the formants except F4w the junction can be regarded as an effective input end of wVT and that wVT resonance can be approximated by pVT resonance. This accounts for the fact that the removal of LC from wVT does not affect the other formants. The vowel /i/ is an exception to the general tendency. In this vowel, pVT is found to resonate as an open tube, not only at F4w, but also at F3w. This is because both F3w and F4w are close to the resonance frequency of LC, and a remarkable increase of the volume velocity occurs at the LC-pVT junction.

TABLE V. The bandwidths (B3 and B4, in Hz) and amplitudes (A3 and A4, in dB) of the third and fourth formants of the vowel /i/, when the laryngeal vestibular area is changed according to the scale factor.

Scale factor	1.50	1.25	1.00	0.75	0.50
B3	79.2	77.2	69.7	57.0	53.9
B4	55.2	61.3	76.3	90.4	88.2
A3	20.4	22.6	25.2	24.5	20.2
A4	24.8	25.2	25.0	20.9	13.6

The resonance mode diagrams also explain the shift of the formants when LC is removed from wVT. Figure 13 shows that the lower formants (F1w, F2w, and F3w) and the higher formants (F5w and F6w) move toward the missing formant of LC resonance. In the lower formants, the volume velocity at the LC-pVT junction increases toward the LC resonance frequency. Especially at F3w, the junction can no longer be regarded as a closed end, and the effective closed end moves toward the glottal side [Fig. 13(a)]. At the higher formants (F5w and F6w), the volume velocity at the LC-pVT junction decreases, and the node of the volume velocity of wVT emerges slightly on the labial side of the junction [Fig. 13(b)]. Consequently, the higher formant frequencies of wVT are slightly higher than those of the pVT, and the effective length of the vocal tract is observed to be shorter. Taken together, the two formant groups shift farther apart when LC is reinstated to form wVT because LC is capacitive for pVT below its own resonance frequency and inductive above that frequency.

3. Susceptance matching between the laryngeal cavity and vocal tract proper

The above resonance mode analysis indicates that F4w is generated by the acoustic coupling between the resonance

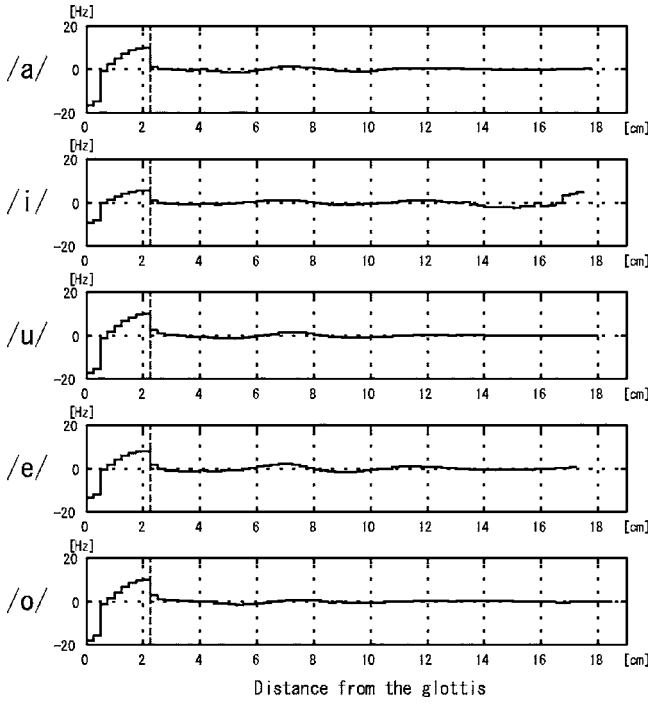


FIG. 7. The influence of small perturbations of the area functions on F4. The vertical axis indicates F4 frequency change at each section when the area is increased by 5% independently of the other sections. The horizontal axis represents the distance from the glottis. A dashed line at 2.25 cm from the glottis indicates the LC-pVT junction.

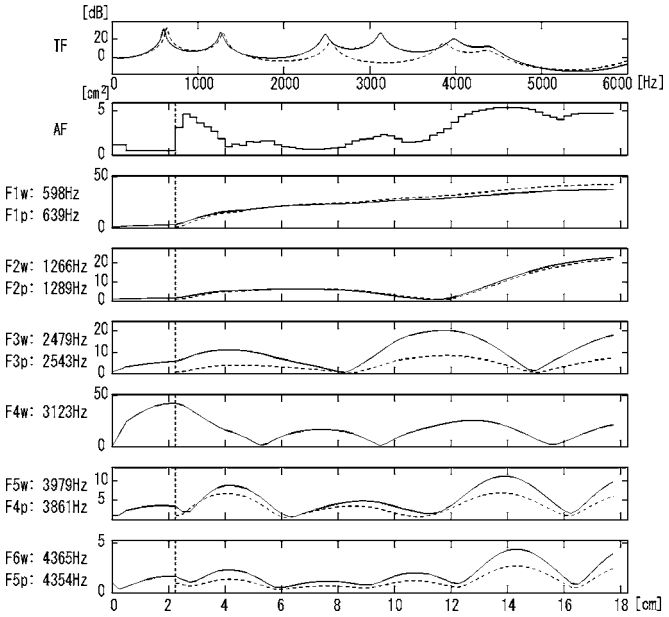


FIG. 8. Transfer function (TF), area function (AF), and resonance mode diagram at each formant for the vowel /a/. Solid lines represent the transfer function or the resonance modes of wVT, and dashed lines indicate those of pVT. A vertical dashed line situated at 2.25 cm from the glottis indicates the LC-pVT junction. Note that the Y axis has no units because the amplitude of volume velocity is shown relative to the input volume velocity (1.0), and that the scale of the vertical axis of the resonance mode diagram differs across diagrams, because the maximum volume velocity has a large deviation.

of LC and the open-tube resonance of pVT. Therefore, we can assume that F4w is located between the resonance frequency of LC and the third zero frequency of the input impedance of pVT Z_p, at which pVT resonates as an open tube. This assumption is examined by using susceptance matching analysis between LC and pVT, when LC is connected to pVT.

Figure 14 shows the susceptance of pVT; B_p, and the negative susceptance of LC, $-B_l$. Zero crossings of B_p represent the formants of pVT, and asymptotes of B_p correspond to zeros of Z_p. Because the susceptance of LC is calculated in the direction from the LC-pVT junction to the glottis, the resonance of LC appears as an asymptote of $-B_l$. Intersections of B_p with $-B_l$, $B_p+B_l=0$ indicate the formants of wVT. Some intersections do not clearly appear because of losses.

Below the resonance frequency of LC, the zero crossings of pVT descend along the upward-sloping curves of B_p, to reach the intersections of B_p with $-B_l$. Therefore, F1w, F2w, and F3w are lower than F1p, F2p, and F3p, respectively. On the other hand, above the resonance frequency of LC, the zero crossings of pVT ascend along the curves of B_p, and thus, the formants of wVT are higher than the corresponding ones of pVT. These results support the previous observations that LC acts as a capacitance against pVT below the resonance frequency of LC and as an inductance above the frequency.

Only F4w, however, has no corresponding formant of pVT, and this fact clearly indicates that LC adds F4w to the resonance pattern. F4w for all the vowels are located between the asymptote of $-B_l$ and the third asymptote of B_p. These facts indicate that F4w appears between the resonance frequency of LC and the third open-tube resonance (the 3/2 wavelength resonance) of pVT.

4. Energy concentration in laryngeal cavity

Table VI lists the acoustic energy concentration in LC: the ratio of the kinetic and potential energies in LC to those in wVT at each formant of the five vowels. In all the vowels,

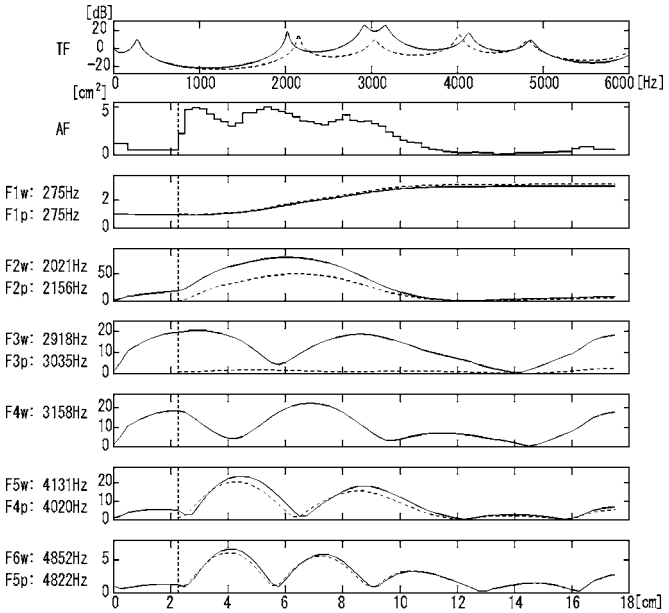


FIG. 9. Transfer function (TF), area function (AF), and resonance modes at each formant for the vowel /i/.

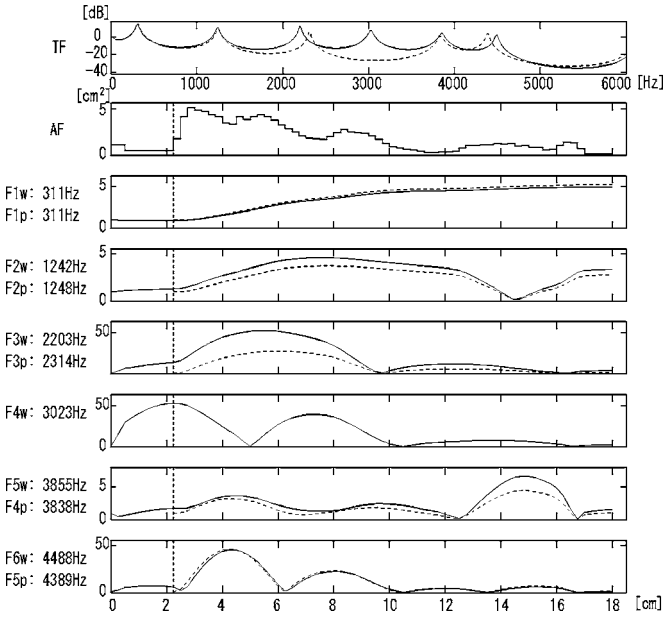


FIG. 10. Transfer function (TF), area function (AF) and resonance modes at each formant for the vowel /u/.

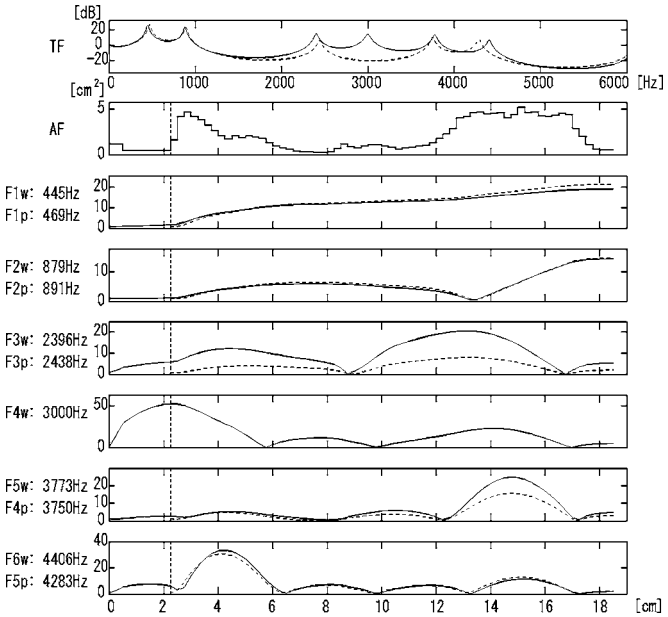


FIG. 12. Transfer function (TF), area function (AF) and resonance modes at each formant for the vowel /o/.

the energy concentration reaches a maximum at F4. This indicates that F4 is essentially affiliated with LC, and the pVT's contribution to the resonance is small. This result agrees with that of Fant and Pauli (1974), whose calculations of the acoustic energy distribution in the vocal tract shapes of six Russian vowels revealed that the main source of F4 was LC. Although LC is less than 14% of wVT in length, almost 70% of the total acoustic energy at F4 is concentrated in LC for the vowels /a/, /u/, and /o/. On the other hand, the concentration rate is lower for the front vowels /i/ and /e/.

5. Acoustic effect of piriform fossae on F4

Figure 15 shows the transfer functions calculated from the area functions used in Fig. 3 (dashed lines), the transfer functions including effects of the piriform fossae (bold lines), and the cepstrum based spectral envelopes averaged among eight speech samples (thin lines). The figure supports the results of Dang and Honda (1997): deep spectral minima appear at the frequency region from 4 to 5 kHz. Although introducing the piriform fossae changes the spectral shape above 4 kHz, it has a small acoustic effect on the F4 location.

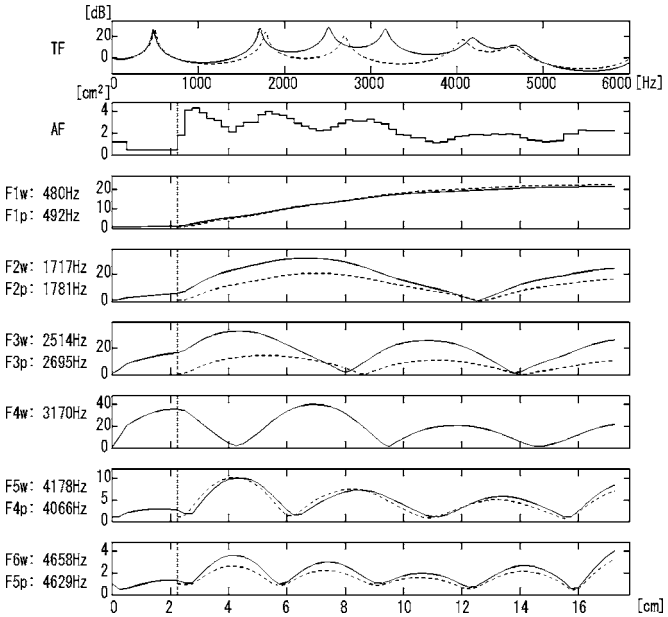


FIG. 11. Transfer function (TF), area function (AF), and resonance modes at each formant for the vowel /e/.

IV. CONCLUSION

The laryngeal cavity (LC) investigated in this study is a small airy space located at the bottom of the vocal tract just above the vocal folds. While its acoustic roles in forming spectral characteristics of vocal sounds have been discussed in the literature, its importance in determining the formant structure of vowels has been underexplored. The following summarizes what we obtained from acoustic simulations in the present study.

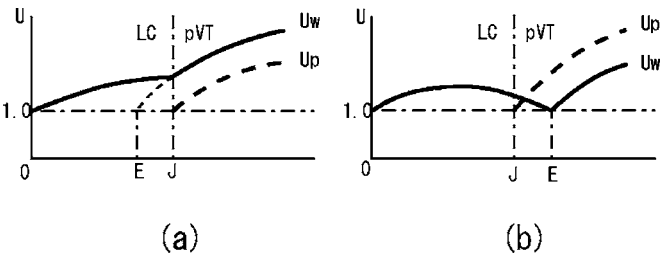


FIG. 13. Effective input end (E) of the vocal tract and the LC-pVT junction (J). (a) At the formants whose frequencies are lower than the resonance of LC, i.e., at F1, F2, and F3, the effective input end is located on the glottal side of the junction. (b) At the formants whose frequencies are higher than the resonance frequency of LC, i.e., at F5 and F6, the effective input end is on the labial side of the junction.

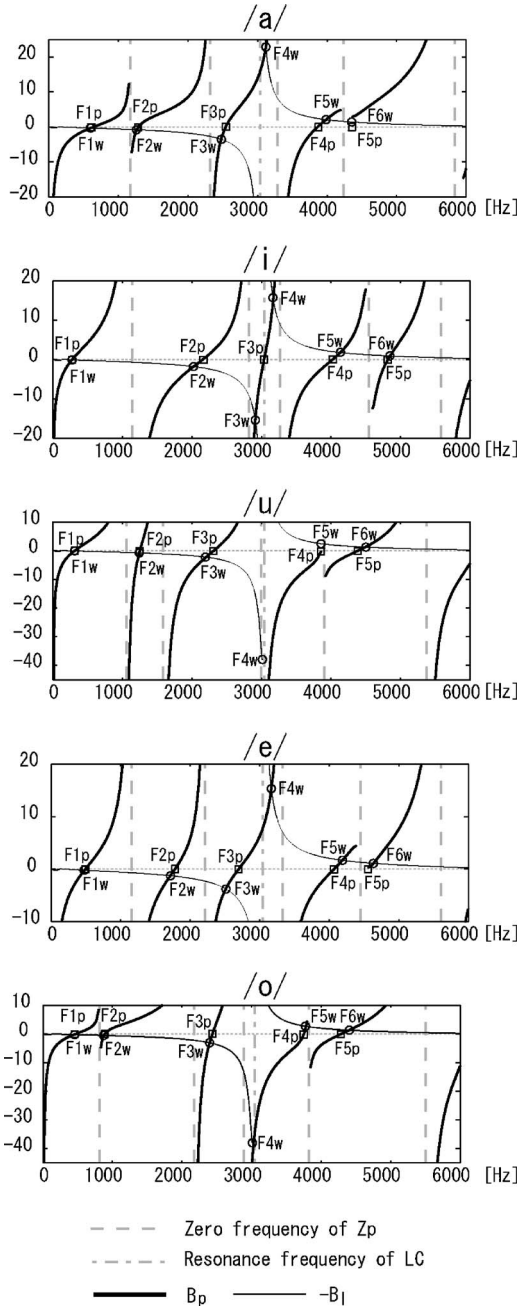


FIG. 14. The normalized susceptance of pVT; B_p, with the thick line and the negative normalized susceptance of LC $-B_l$ with the thin line. Small squares with F1p-F5p indicate zero crossings of B_p, representing the formants of pVT. The circles with F1w-F6w are placed on the intersections of B_p and $-B_l$, indicating the formants of wVT. The gray dashed lines are asymptotes of B_p, i.e., zero frequencies of the pVT's input impedance, where the open-tube resonances occur. The gray alternate long and short dashed line is an asymptote of $-B_l$, corresponding to the resonance frequency of LC.

(1) LC generates one of the formants of the vocal tract, which corresponds to the fourth one (F4) in this study. Because the formant is provided almost only by LC, the formant is mainly sensitive to the LC shape.

(2) The formant affiliated with LC appears between the resonance frequency of LC and the third zero frequency of the input impedance of the vocal tract proper (pVT).

(3) Because the formant is close to the resonance frequency of LC, the volume velocity increases at the LC-pVT junction to induce an open-tube resonance of pVT.

TABLE VI. The percentage of the amount of the acoustic energy (kinetic and potential energies) in LC relative to that in the vocal tract for each vowel and for each formant.

	F1	F2	F3	F4	F5	F6
/a/	7.4	2.6	10.2	67.4	13.8	23.7
/i/	3.8	13.9	36.1	37.3	10.3	5.7
/u/	3.9	2.3	13.0	70.7	6.0	8.1
/e/	3.7	6.7	22.8	53.5	10.8	13.5
/o/	6.3	1.7	7.1	73.0	4.0	10.2

(4) At the other formants, the volume velocity at the junction is so small that the junction can be considered as an effective input end of the vocal tract. pVT, therefore, resonates as a closed tube.

In the previous studies, LC was often treated as a straight closed tube (called the "larynx tube") neglecting the cavity formed by the ventricle. Anatomically, the real LC geometry when including the ventricle resembles a Helmholtz resonator with a long neck, and the acoustic effects of LC also agree with those of a Helmholtz resonator. Because of the long neck of LC, however, the resonance frequency estimated from the Helmholtz equation does not necessarily coincide with the calculation based on a transmission line model.

In this study, LC is found to generate a special formant F4, which is rather independent from other formants and determined almost only by LC geometry. Since the resonance of LC arises between two higher formants, it contributes to generating a cluster of formants. It has been known that in singing voice, F3, F4, and F5 tend to form a strong cluster called the "singing formant." The present study suggests that the control of LC geometry is a factor contributing to F3-F4 clustering of singing voice, while other factors include widening the piriform fossae (Sundberg, 1974) and creating a distinct division between the back and front portions of the vocal tract (Story, 2004).

In addition to providing a formant, LC lifts the amplitude of the transfer function by approximately 10 dB in the frequency region from 2 to 4 kHz, as shown in Fig. 4(/a/: 10.2 dB; /i/: 9.0 dB; /u/: 10.7 dB; /e/: 9.5 dB; /o/: 10.9 dB). It is interesting to find this local amplification mechanism by LC because this frequency region coincides with the peak of human auditory sensitivity (Fletcher, 1940). Considering that F4 as well as F5 is rather stable in running speech, LC may well act as an organ to transmit not linguistic but paralinguistic or nonlinguistic information. In fact, the same frequency range has been discussed as one of the determinants of voice quality (e.g., Sundberg 1974; Nawka *et al.*, 1997) or in relation to the perceptual (e.g., Kitamura and Akagi, 1997) and acoustic-phonetic (e.g., Mokhtari and Clermont, 1996) factors of speaker characteristics.

LC and the piriform fossae form the bottom of the vocal tract, and they determine the spectral envelope of the higher frequency region above 2 kHz: LC generates a peak at 3 kHz and the bilateral cavities of the piriform fossae cause dips at 4–5 kHz, as shown in Sec. III B 5, both affecting the lower formants only slightly. The geometrical variation of the hypopharyngeal cavities is large across speakers but small

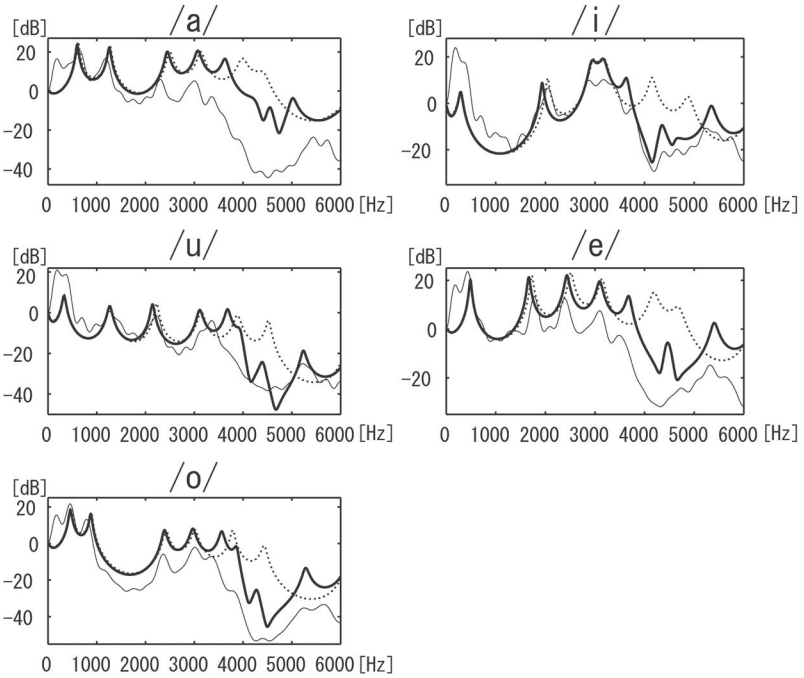


FIG. 15. Transfer functions of the vocal tract (dashed lines), those including the acoustic effects of the piriform fossae (bold lines), and cepstrum based speech spectra (thin lines).

across vowels within a speaker (Kitamura *et al.*, 2005). Therefore, these acoustic effects, or the "hypopharyngeal resonance," can be treated independently from the so-called vocal tract resonance, as proposed by the acoustic model of vowel production with hypopharyngeal cavity coupling (Honda *et al.*, 2004).

ACKNOWLEGMENT

This research was conducted as part of "Research on Human Communication" with funding from the National Institute of Information and Communications Technology of Japan.

Adachi, S. and Yamada, M. (**1999**). "An acoustical study of sound production in biphonic singing, Xöömij," J. Acoust. Soc. Am. **105**,2920–2932.

Badin, P., Perrier, P., Boë, L. J., and Abry, C. (**1990**). "Vocalic nomograms: Acoustic and articulatory consideration upon formant convergences," J. Acoust. Soc. Am. **87**,1290–1300.

Bartholomew, W. T. (**1934**). "A physical definition of "good voice-quality" in the male voice," J. Acoust. Soc. Am. **6**,25–33.

Chiba, T. and Kajiyama, M. (**1942**). *The Vowels-Its Nature and Structure* (Tokyo-Kaiseikan, Tokyo).

Caussé, R., Kergomard, J., and Lurton, X. (**1984**). "Input impedance of brass musical instruments-Comparison between experiments and numerical models," J. Acoust. Soc. Am. **75**,241–254.

Dang, J. and Honda, K. (**1997**). "Acoustic characteristics of the piriform fossa in models and humans," J. Acoust. Soc. Am. **101**,456–465.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fant, G. and Båvegård, M. (**1997**). "Parametric model of VT area functions: vowels and consonant," Speech, Music and Hearing Laboratory-Quarterly Progress and Status Report, Stockholm, Vol. **38**, pp. 1–21.

Fant, G. and Pauli, S. (**1974**). "Spatial characteristics of vocal tract resonance modes," in *Proc. Speech Comm. Sem.*, Stockholm, Sweden, 1–3, August Vol. **74**, pp. 121–132.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Honda, K., Takemoto, H., Kitamura, T., Fujita, S., and Takano, S. (**2004**). "Exploring human speech production mechanisms by MRI," IEICE Trans. Inf. Syst. **E87-D**,1050–1058.

Ingard, U. (**1953**). "On the theory and design of acoustic resonators," J. Acoust. Soc. Am. **25**,1037–1067.

Kitamura, T. and Akagi, M. (**1997**). "Significant cues in spectral envelope of isolated vowels for speaker identification," J. Acoust. Soc. Jpn. **53**,185–191.

Kitamura, T., Honda, K., and Takemoto, H. (**2005**). "Individual variation of the hypopharyngeal cavities and its acoustic effects," Acoust. Sci. & Tech. **26**,16–26.

Lewis, D. (**1936**). "Vocal resonance," J. Acoust. Soc. Am. **8**,91–99.

Mokhtari, P. and Clermont, F. (**1996**). "A methodology for investigating vowel-speaker interactions in the acoustic-phonetic domain," Proc. Australian International Conference on Speech Science and Technology, Adelaide, Australia, pp. 127–132.

Mrayati, M., Carre, R., and Guerin, B. (**1988**). "Distinctive regions and modes: a new theory of speech production," Speech Commun. **7**,257–286.

Nawka, T., Anders, L. C., Cebulla, M., and Zurakowski, D. (**1997**). "The speaker's formant in male voices," J. Voice **11**,422–428.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Story, B. H. (**2004**). "Vowel acoustics for speaking and singing," Acta Acust. **90**,629–640.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**2001**). "The relationship of vocal tract shape to three voice qualities," J. Acoust. Soc. Am. **109**,1651–1667.

Sundberg, J. (**1974**). "Articulatory interpretation of the 'singing formant'," J. Acoust. Soc. Am. **55**,838–844.

Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois University Press, DeKalb, IL).

Sundberg, J. (**1995**). "The singer's formant revisited," Speech Transmission Laboratory-Quarterly Progress and Status Report, Stockholm, 2-3/1995, pp. 83–96.

Sundberg, J. (**2003**). "Research on the singing voice in retrospect," Speech, Music and Hearing Laboratory-Quarterly Progress and Status Report, Stockholm, Vol. **45**, pp. 11–22.

Takemoto, H., Honda, K., Masaki, S., Shimada, Y., and Fujimoto, I. (**2003**). "Modeling of the inferior part of the vocal tract based on analysis of 3D cine-MRI data," Proc. Autumn Meet. Acoust. Soc. Jpn., pp. 281–282.

Takemoto, H., Honda, K., Masaki, S., Shimada, Y., and Fujimoto, I. (**2006**). "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data," J. Acoust. Soc. Am. **119**,1037–1049.

Williams, P. L., Bannister, L. H., Berry, M. M., Collins, P., Dyson, M., Dussek, J. E., and Ferguson, M. W. J. (**1995**). *Gray's Anatomy* (Churchill Livingstone, New York), 38th ed.

Cyclicity of laryngeal cavity resonance due to vocal fold vibration

Tatsuya Kitamura,[a] Hironori Takemoto, Seiji Adachi, Parham Mokhtari, and Kiyoshi Honda
ATR Human Information Science Laboratories, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan

(Received 24 October 2005; revised 14 July 2006; accepted 17 July 2006)

Acoustic effects of the time-varying glottal area due to vocal fold vibration on the laryngeal cavity resonance were investigated based on vocal tract area functions and acoustic analysis. The laryngeal cavity consists of the vestibular and ventricular parts of the larynx, and gives rise to a regional acoustic resonance within the vocal tract, with this resonance imparting an extra formant to the vocal tract resonance pattern. Vocal tract transfer functions of the five Japanese vowels uttered by three male subjects were calculated under open- and closed-glottis conditions. The results revealed that the resonance appears at the frequency region from 3.0 to 3.7 kHz when the glottis is closed and disappears when it is open. Real spectra estimated from open- and closed-glottis periods of vowel sounds also showed the on-off pattern of the resonance within a pitch period. Furthermore, a time-domain acoustic analysis of vowels indicated that the resonance component could be observed as a pitch-synchronized rise-and-fall pattern of the bandpass amplitude. The cyclic nature of the resonance can be explained as the laryngeal cavity acting as a closed tube that generates the resonance during a closed-glottis period, but damps the resonance off during an open-glottis period. [DOI: 10.1121/1.2335428]

PACS number(s): 43.70.Bk, 43.70.Aj [BHS] Pages: 2239–2249

I. INTRODUCTION

The laryngeal cavity is a part of the vocal tract that comprises the lower section of the pharynx above the glottis. It has often been argued that the resonance of this cavity contributes to an extra formant (Bartholomew, 1934), modification of the higher formants of vowels (Fant, 1960), or spectral prominence in a higher frequency region (Sundberg, 1974). In previous studies, the laryngeal cavity has been considered as a closed tube. During voice production, however, the bottom of the cavity opens and closes due to vocal fold vibration, possibly resulting in an on-off pattern of the resonance within a glottal cycle. The aim of the present study is to explore the cyclicity of laryngeal cavity resonance due to opening and closure of the glottis.

The laryngeal cavity is defined in this study as the lower part of the vocal tract including the vestibular and ventricular spaces of the larynx. The laryngeal ventricle is a small expansion of the tract located above the glottis between the true and false vocal folds, and the laryngeal vestibule is a narrow tubular space situated superior to the laryngeal ventricle and opening into the wide pharyngeal cavity. Acoustic roles of the cavity have been investigated by many researchers. Lewis (1936), for instance, examined resonance patterns of sung vowels and pointed out that the resonance at around 3.2 kHz is consistent among vowels for a male singer. Chiba and Kajiyama (1942) estimated that the resonance arises from the laryngeal cavity because the frequency approximately corresponds to the quarter-wavelength resonance frequency of the cavity, whose length spans from 2.0 to 2.7 cm for male subjects.

In other studies, it has been reported that varying the length or area of the laryngeal cavity mainly affects the frequencies of the third, fourth, and fifth formants, while its effects on the first and second formants are small (Fant, 1960, 1975, 1976, 1980; Fant and Pauli, 1975; Fant and Båvegård, 1997; Titze and Story, 1997; Imagawa *et al.*, 2003; Story, 2004). Fant and his colleagues reported in detail that the laryngeal cavity contributes to the fourth and fifth formants. Fant (1960) showed that the insertion of a laryngeal tube cavity into vocal tract area functions for Russian vowels increases the density of poles in the frequency region from 3 to 5 kHz and affects the frequencies of adjacent poles. Dependence of the fourth formant on the laryngeal cavity was clearly demonstrated by vocal tract energy function (Fant 1975; Fant and Pauli 1975; Fant 1976, 1980). Fant and Båvegård (1997) also reported the length of the laryngeal cavity could affect the fourth and fifth formants. Sundberg (1974) observed that the shapes of the pharyngeal and hypopharyngeal cavities are different between speech and singing, and suggested that acoustic isolation of the laryngeal cavity from the entire vocal tract is the condition necessary to give rise to the "singing formant." He also indicated that the cross-sectional area ratio between the pharyngeal cavity to the outlet of the laryngeal cavity must be at least 6:1 to generate that particular formant. Most recently, Takemoto *et al.* (2005) examined acoustic roles of the laryngeal cavity using magnetic resonance imaging (MRI)-based simulations and demonstrated that a quarter-wavelength resonance of the laryngeal cavity generates one of the formants in a male speaker, which disappears when the cavity is removed from the entire vocal tract. Their simulation also showed that the cavity is acoustically independent of the vocal tract exclud-

[a]Electronic mail: kitamura@atr.jp

ing the cavity at that resonance frequency. For that particular subject, the laryngeal cavity resonance was found to be the fourth formant for the five Japanese vowels.

It is generally assumed that the glottal source and the vocal tract resonance are independent or linearly separable factors in vowel production (Fant, 1960), where the vocal tract is considered as a closed tube from the closed glottal-end to the open lip-end. The studies described in the previous paragraph also considered the laryngeal cavity as a closed tube disregarding the effects of glottal opening on the laryngeal cavity resonance. It may be worthwhile, however, to reconsider that the laryngeal cavity resonance is not stable because the cavity is no longer a closed tube during the open-glottis period. The present study therefore attempts to explore the effects of the open and closed glottis on the laryngeal cavity resonance during vocal fold vibration. Based on vocal tract area functions obtained from MRI, we first examine the relationship between the glottal areas and vocal tract transfer functions using a transmission line model (Sec. II). We next estimate power spectral densities of vowel sounds made during the open- and closed-glottis periods by applying a short-term spectral analysis on synthetic and natural speech (Sec. III). Furthermore, the pitch-synchronous pattern of the laryngeal cavity resonance is extracted from natural vowels by means of a bandpass filter tuned to the resonance frequency to demonstrate the cyclic nature of the resonance in natural speech signals (Sec. IV). We then discuss the results (Sec. V) and offer concluding remarks (Sec. VI).

II. SIMULATION USING TRANSMISSION LINE MODEL

Acoustic effects of glottal opening on the laryngeal cavity resonance were estimated using a transmission line model. In this simulation, we used vocal tract area functions measured from volumetric magnetic resonance (MR) images for three male subjects, and vocal tract transfer functions were calculated under open- and closed-glottis conditions. Transfer functions of vocal tract area functions excluding the laryngeal cavity region were also computed to estimate the laryngeal cavity resonance by comparing with the transfer functions of the entire vocal tract area functions.

A. MR images and vocal tract area functions

MR images of three Japanese male subjects KH, TI, and YT were obtained during production of the five Japanese vowels (/a/, /e/, /i/, /o/, and /ɯ/) with a Shimadzu-Marconi ECLIPSE 1.5T Power Drive 250 at the ATR Brain Activity Imaging Center. Subjects KH and YT were Tokyo dialect speakers, and subject TI was a Kansai dialect speaker. They all produced the vowel /ɯ/ as a midvowel, with subject TI producing the vowel as a slightly posterior type.

Two techniques were used in this study to acquire high-quality MR images from the subjects: phonation-synchronized scanning and bone-conducting stimulus presentation (Nota *et al.*, in press). The phonation-synchronized scan is the method for acquiring MRI data only during vowel production (Masaki *et al.*, 1999; Takano *et al.*, 2005). With this method, our subjects are presented with a harmonic-complex-tone with a cyclic sequence (four beats on a 3 s cycle) while lying in the MRI unit. Each subject repeatedly breathes in at the first beat and phonates during the succeeding beats in exact timing with the sequence. This technique allows scanning only during production to avoid motion artifacts due to inhalation. The subjects were instructed to adjust their pitch frequency to the fundamental frequency of the harmonic-complex-tone at 110 Hz in order to reduce artifacts due to larynx movements. The bone-conducting stimulus presentation is used to avoid distortion of vowel gestures due to exposure to the intense machine noise during experiments. Subjects wore earplugs and were presented with the harmonic-complex-tone by a piezoelectric bone-conduction speaker. This makes it possible for the subject to monitor his/her own voice through bone conduction feedback, and thus to produce more natural-sounding vowels during scanning.

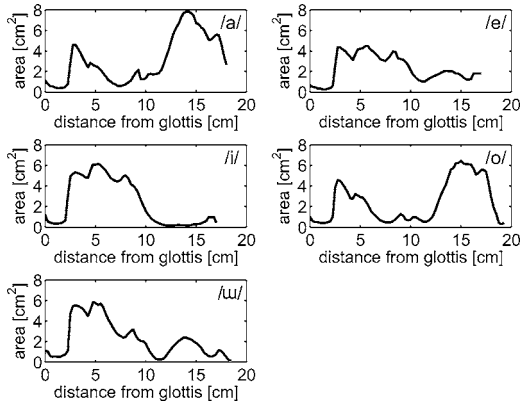


FIG. 1. Vocal tract area functions for the five Japanese vowels from subject KH extracted from MR images.

Each subject was positioned to lie supine on the platform of the MRI unit. A head-neck coil was then positioned over the subject's head and neck region. The imaging sequence was a sagittal fast spin echo series with 2.0-mm slice thickness, no slice gap, no averaging, a 256×256-mm field of view, a 512×512-pixel image size, 51 slices, 90° flip angle, 11-ms echo time, and 3000-ms repetition time.

Cross-sectional areas of the vocal tract along its midline were then measured at 2.5-mm intervals from the MR images according to Takemoto *et al.* (2005). Prior to measuring the area, volume data of the upper and lower jaws were superimposed on the MR images (Takemoto *et al.*, 2004) for accurate measurement taking into account the teeth. The bilateral piriform fossa cavities were excluded from the area functions in this study. Figures 1–3 illustrate the extracted area functions of the five vowels for subjects KH, TI, and YT, respectively.

B. Method for calculating transfer functions

Calculation of velocity-to-velocity transfer functions of the vocal tract area functions was based on a transmission line model. The transfer functions were calculated for the frequency region up to 5 kHz considering the glottal impedance and radiation impedance at the mouth. The glottal area A_g was set to 0.0 cm^2 (complete closure) and 0.2 cm^2. The

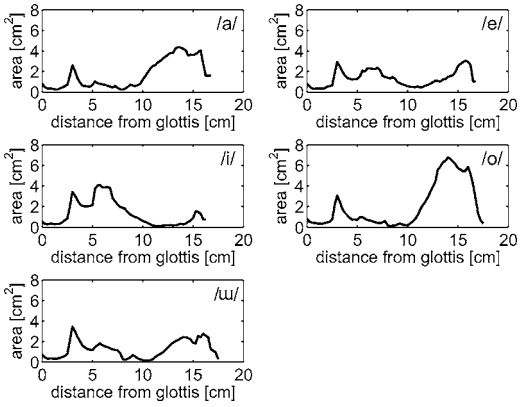


FIG. 2. Vocal tract area functions for the five Japanese vowels from subject TI extracted from MR images.

glottal impedance Z_g for the closed-glottis condition (A_g $=0.0$ cm^2) was set to an infinite value while that for the open-glottis condition ($A_g=0.2$ cm^2) was approximated by

$$Z_g = R_g + j\omega L_g = \left(\frac{12\mu d_g l_g^2}{A_g^3} + \frac{0.875}{A_g}\sqrt{2p_0\rho} \right) + j\omega \frac{\rho d_g}{A_g}, \tag{1}$$

where μ is the viscosity coefficient, d_g is the depth of the glottal slit, l_g is the length of the glottal slit, p_0 is the subglottal pressure, and ρ is the air density (Ishizaka and Flanagan, 1972; Flanagan, 1972). Since this equation does not consider acoustic effects of the trachea and lungs, subglottal formants do not appear under the open-glottis condition in the simulation. We assumed $\mu=1.88\times10^{-5}$ kg/(s m), d_g $=3$ mm, $l_g=18$ mm, $p_0=10$ cm H_2O, and $\rho=1.12$ kg/m^3.

The radiation impedance of the vocal tract Z_R was approximated by the following equation suggested by Caussé *et al.* (1984):

$$\frac{Z_R}{\rho c} = \frac{z^2}{4} + 0.0127z^4 + 0.082z^4 \ln z - 0.023z^6 + j(0.6133z - 0.036z^3 + 0.034z^3 \ln z - 0.0187z^5), \tag{2}$$

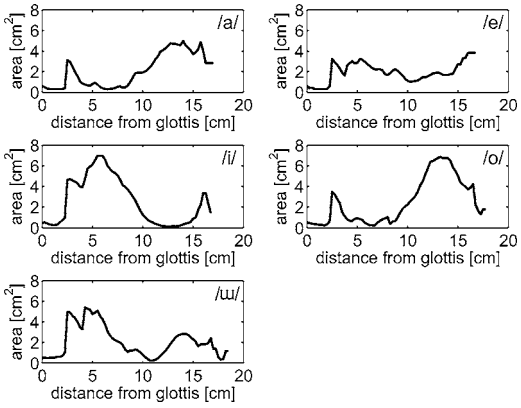


FIG. 3. Vocal tract area functions for the five Japanese vowels from subject YT extracted from MR images.

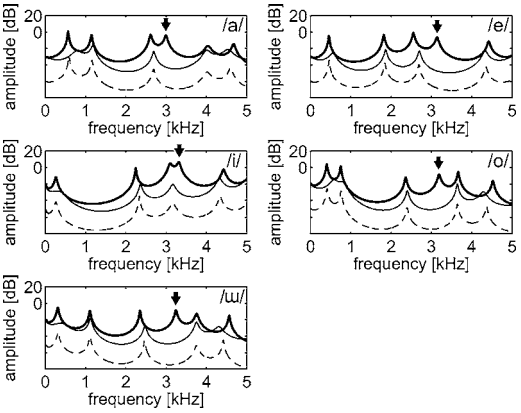


FIG. 4. Velocity-to-velocity transfer functions with the glottal area $A_g=0.0$ and 0.2 cm^2 for the five Japanese vowels from subject KH. The thick and thin lines show those with $A_g=0.0$ and 0.2 cm^2, respectively. The dashed lines denote transfer functions of the vocal tract without the laryngeal cavity. Arrows indicate the laryngeal cavity resonances.

$$z = kr, \tag{3}$$

where k is the wave number, r is the radius of the open end, and c is the speed of sound. We assumed $c=353.0$ m/s. It should be noted that Eq. (2) is valid for a frequency region satisfying $kr<1.5$. Since the maximum equivalent radius of the open end of the area functions is 11 mm (the vowel /e/ of subject YT), Eq. (2) is valid for the frequency region for the simulation (up to 5 kHz). In addition to the losses above, the model includes losses due to heat conduction, viscous friction, and vibration at the vocal tract wall.

We also estimated the laryngeal cavity resonance by comparing the transfer functions of the area functions with and without the laryngeal cavity. Similar to Takemoto *et al.* (2005), the laryngeal cavity was eliminated from the vocal tract so that the inlet of the pharyngeal cavity becomes the input end of the vocal tract without the laryngeal cavity. Hereafter, the vocal tract excluding the laryngeal cavity is referred to as the "vocal tract proper." Transfer functions were calculated for the vocal tract proper only under the closed-glottis condition.

C. Results

Figures 4–6 depict calculated velocity-to-velocity transfer functions of vocal tract area functions with two different glottal areas. Transfer functions of the vocal tract proper (vocal tract region without the laryngeal cavity) are also superimposed in the figures at a lower overall gain. Arrows in the figures point to the laryngeal cavity resonances. Comparisons of the transfer functions reveal that one of the formants appears with complete closure of the glottis and disappears in the open-glottis state. This formant resides at the frequency region from 3.0 to 3.7 kHz. Since this formant is missing when the laryngeal cavity is removed from the entire vocal tract, it can be regarded as the laryngeal cavity resonance. These observations indicate that the laryngeal cavity resonance appears during the closed-glottis period and that it disappears during the open-glottis period.

The formants adjacent to the laryngeal cavity resonance were found to shift toward the missing resonance frequency

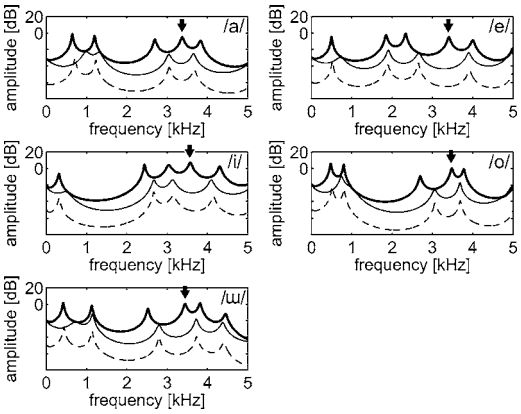


FIG. 5. Velocity-to-velocity transfer functions with the glottal area A_g=0.0 and 0.2 cm^2 for the five Japanese vowels from subject TI. The thick and thin lines represent those with A_g=0.0 and 0.2 cm^2, respectively. The dashed lines denote transfer functions of the vocal tract without the laryngeal cavity. Arrows indicate the laryngeal cavity resonances.

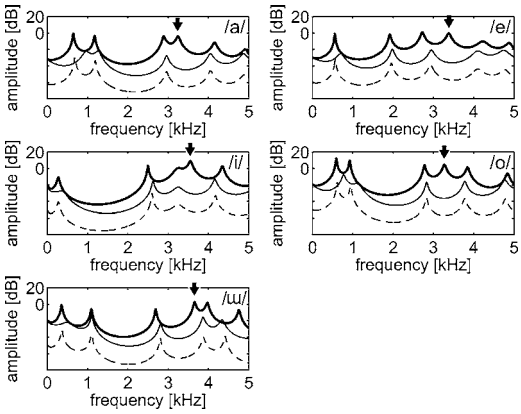


FIG. 6. Velocity-to-velocity transfer functions with the glottal area A_g=0.0 and 0.2 cm^2 for the five Japanese vowels of subject YT. The thick and thin lines denote those with A_g=0.0 and 0.2 cm^2, respectively. The dashed lines represent transfer functions of vocal tract without the laryngeal cavity. Arrows show the laryngeal cavity resonances.

when the glottis is open. The formant frequencies of those transfer functions were measured for the conditions of A_g =0.0 and 0.2 cm^2 and listed in Tables I–III. For example, in the case of the vowel /a/ from subject KH (Fig. 4), Table I shows that the lower three formants shift upward and the fifth and sixth formants shift downward in the open-glottis condition.

For all three subjects, the frequencies of the second and higher formants on all the transfer functions during the open-glottis period are nearly the same as those of the transfer functions of the vocal tract proper, while the first formant frequency of those transfer functions changes significantly because the formant is highly damped when the glottis is open. These results indicate that the resonance pattern for the open-glottis condition is generated only within the vocal tract proper, and that the laryngeal cavity resonance provides an extra formant to the resonance pattern of the vocal tract proper when the glottis is closed.

III. SPECTRAL OBSERVATION OF LARYNGEAL CAVITY RESONANCE DURING OPEN- AND CLOSED-GLOTTIS PERIODS

The results in the previous section suggested that the spectral pattern of vowels changes cyclically due to vocal fold vibration. A pitch-synchronous short-term spectral analysis was applied to open- and closed-glottis periods to explore the appearance and disappearance of the laryngeal cavity resonance due to the glottal conditions. To confirm the validity of the method with a small number of sample points, we first compared the analysis results from synthetic vowel samples and the transfer functions obtained previously. Then, we estimated spectra from natural vowel samples and compared the spectra from the two glottal conditions.

A. Synthetic speech data

Acoustic wave forms of five Japanese vowels were computed for the three male subjects to estimate power spectral densities (PSDs) during glottal open and closed periods. This synthesis was performed using Maeda's time-domain speech synthesizer (Maeda, 1982). The synthesizer generates speech wave forms from a vocal tract area function and a glottal area time-function $A_g(t)$, and consequently acoustic responses during the open and closed periods of the glottis can be accurately simulated. This method offers another advantage of ensuring noise-free wave forms, since a PSD estimation from data with a short window length for glottal open or closed periods is particularly sensitive to noise in the data samples. Thus, the use of the synthesizer guarantees ideal conditions for the analysis.

TABLE I. Formant frequencies of velocity-to-velocity transfer functions for the five Japanese vowels from subject KH for glottal area conditions A_g=0.0 and 0.2 cm^2.

Vowel	A_g (cm^2)	Formant frequency (Hz)					
/a/	0.0	566	1147	2612	2994	4036	4673
	0.2	787	1181	2692	4009	4527	
/e/	0.0	458	1819	2554	3141	4421	
	0.2	552	1865	2703	4330		
/i/	0.0	264	2243	3102	3312	4422	
	0.2	314	2363	3164	4319		
/o/	0.0	405	751	2363	3191	3666	4527
	0.2	590	775	2404	3644	4289	
/ɯ/	0.0	311	1108	2351	3241	3756	4562
	0.2	391	1121	2473	3751	4305	

TABLE II. Formant frequencies of velocity-to-velocity transfer functions for the five Japanese vowels from subject TI for glottal area conditions A_g=0.0 and 0.2 cm^2.

Vowel	A_g (cm^2)	Formant frequency (Hz)					
/a/	0.0	645	1199	2694	3372	3828	
	0.2	991	1307	3042	3651		
/e/	0.0	499	1851	2336	3409	4003	
	0.2	713	1906	2661	3902	4944	
/i/	0.0	314	2436	3039	3573	4303	
	0.2	471	2670	3136	4092		
/o/	0.0	478	800	2699	3481	3779	
	0.2	743	—	3057	3685		
/ɯ/	0.0	420	1130	2520	3442	3813	4446
	0.2	725	1147	2795	3721	4377	

The vocal tract area functions shown in Figs. 1–3 were used to synthesize vowels. $A_g(t)$ was a repeated series of a single-period Rosenberg wave (Rosenberg, 1971) of the form

$$f(t)=\begin{cases} a\left[3\left(\frac{t}{\tau_1}\right)^2-2\left(\frac{t}{\tau_1}\right)^3\right] & 0\leqslant t\leqslant \tau_1, \\ a\left[1-\left(\frac{t-\tau_2}{\tau_2}\right)^2\right] & \tau_1<t\leqslant \tau_1+\tau_2, \\ 0 & \tau_1+\tau_2<t\leqslant T, \end{cases} \tag{4}$$

where a=0.2 cm^2, τ_1=33, τ_2=23, and T=100. The fundamental frequency of the synthetic vowel was fixed at 110 Hz.

B. Natural speech data

The five vowels of the male subjects were recorded in an anechoic room at a sampling rate of 48 kHz with 16-bit resolution by a solid-state audio recorder (Marantz PMD670). Electroglottograph (EGG) wave forms were recorded simultaneously with the speech to estimate open- and closed-glottis periods. The direct current (dc) component of the EGG wave forms was removed by a highpass filter built into the recorder. The EGG wave forms were then shifted backward by 0.72 ms to compensate for the time lag between the EGG and speech wave forms. This time lag was computed using $(l+d)/c$ where l is the vocal tract length (17 cm), d is the distance between the lips and the microphone (8.5 cm), and c is the speed of sound (353.0 m/s).

C. Method for short-term spectral analysis

Because the duration of open- and closed-glottis periods of speech is too short to obtain fine fast Fourier transform spectra, we therefore adopted Burg's method (Marple, 1987), which is based on the autoregressive (AR) parametric wave form model, to estimate PSDs of the wave forms from the two periods. Since an appropriate model order for estimating a PSD is likely to differ between the two periods, the model order was chosen to minimize an error criterion using the minimum description length (MDL) developed by Rissanen (1983). The MDL for AR models of order p is defined as

$$\mathrm{MDL}[p]=N\ln(\hat{\rho}_p)+p\ln(N), \tag{5}$$

where N is the number of data samples and $\hat{\rho}_p$ is the estimated white noise variance. The term $p\ln(N)$ is the penalty factor for preventing the order p from increasing excessively (Marple, 1987).

With the above considerations, we analyzed the PSDs of open- and closed-glottis periods for the synthetic and natural data. The data were downsampled to 10 kHz and were preemphasized with a factor of 0.98. Glottal open and closed periods were then excerpted manually using the glottal area time function $A_g(t)$ for the synthetic speech and using the EGG wave forms for the natural speech. The glottal closed period of the natural speech was set from the maximum point to the minimum point of the differentiated EGG wave form within a glottal period. Since EGG measures the vocal fold contact area but not the glottal area (Baken and Orlikoff,

TABLE III. Formant frequencies of velocity-to-velocity transfer functions for the five Japanese vowels from subject YT for glottal area conditions A_g=0.0 and 0.2 cm^2.

Vowel	A_g (cm^2)	Formant frequency (Hz)					
/a/	0.0	645	1175	2890	3245	4150	4905
	0.2	970	1255	2960	4050	4870	
/e/	0.0	555	1920	2725	3380	4225	4875
	0.2	700	1985	2950	4100	4750	
/i/	0.0	275	2495	3260	3550	4345	
	0.2	345	2620	3250	4155		
/o/	0.0	590	925	2780	3275	3850	4840
	0.2	775	1095	2835	3780	4770	
/ɯ/	0.0	355	1095	2690	3655	3970	4755
	0.2	485	1105	2810	3865	4335	

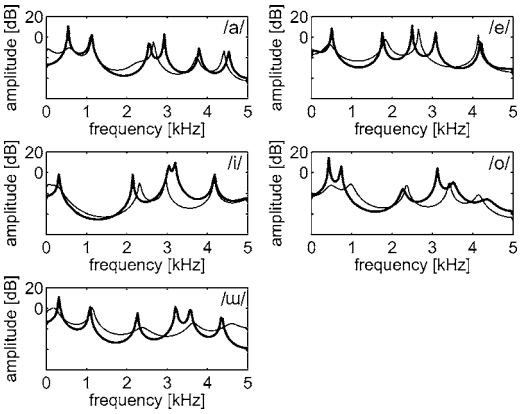


FIG. 7. Power spectral densities of open- and closed-glottis periods for synthetic speech from subject KH. The thick lines denote those during the closed-glottis period and the thin lines represent those during the open-glottis period.

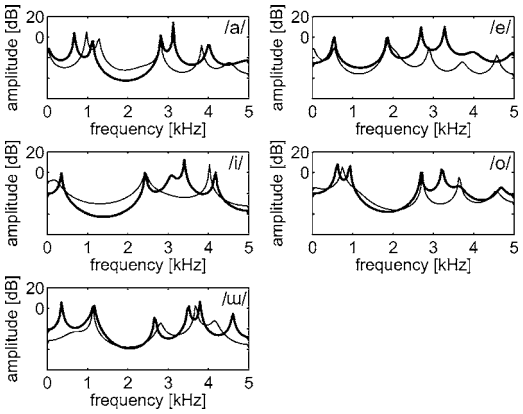


FIG. 9. Power spectral densities of open- and closed-glottis periods for synthetic speech from subject YT. The thick lines denote those during the closed-glottis period and the thin lines represent those during the open-glottis period.

2000), there is potential uncertainty in the estimation of the glottal open and closed periods. After the dc component of the samples from each period was eliminated, PSDs of the synthetic speech were estimated from a single period, and those of the natural speech were obtained from AR model parameters averaged over five successive periods. PSDs were also estimated on a pitch-asynchronous, frame-by-frame basis. The frame length was 3 ms and the frame period was 1.5 ms.

D. Results

Figures 7–9 show the estimated PSDs of the open- and closed-glottis periods of the synthetic vowels from the three subjects' transfer functions. The thick lines denote those during the closed-glottis period and the thin lines represent those during the open-glottis period. The shapes of the PSDs resemble those of the corresponding transfer functions shown in Figs. 4–6, although some of the formants are missing for the open-glottis period due to a high damping of the formant; thus, it is reasonable to consider that the PSDs are estimated accurately.

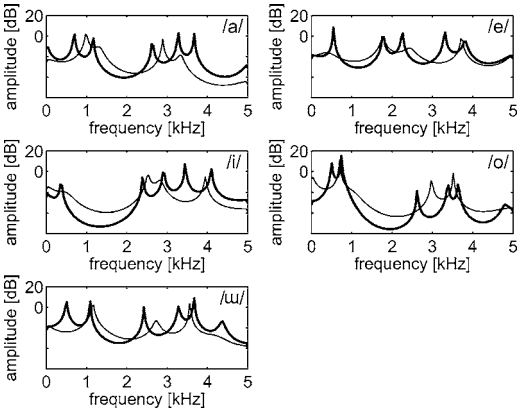


FIG. 8. Power spectral densities of open- and closed-glottis periods for synthetic speech from subject TI. The thick lines represent those during the closed-glottis period and the thin lines denote those during the open-glottis period.

As the results of the previous simulation show, one of the formants occurring in the closed-glottis period disappeared during the open-glottis period, and this formant is identical to those from the simulation. These results do not depend on the vowel type nor speaker. For example, in the case of the vowel /a/ of subject KH (Fig. 7), the fourth formant appears during the closed-glottis period and disappears during the open-glottis period just as the corresponding transfer function shown in Fig. 4.

The glottal opening affects the PSDs' pole frequencies. The frequencies of adjacent poles of the laryngeal cavity resonance tend to shift toward the diminished resonance frequency when the glottis opens, while a theoretical study suggested that pole frequencies and bandwidths increase due to the glottal loss, with the effect being greater at the lower frequencies (Flanagan, 1972). On the other hand, the frequencies and bandwidths of the lower two poles of the PSDs increase during the open-glottis conditions, consistent with the classical theoretical study.

Figure 10(a) shows the synthetic wave form for vowel

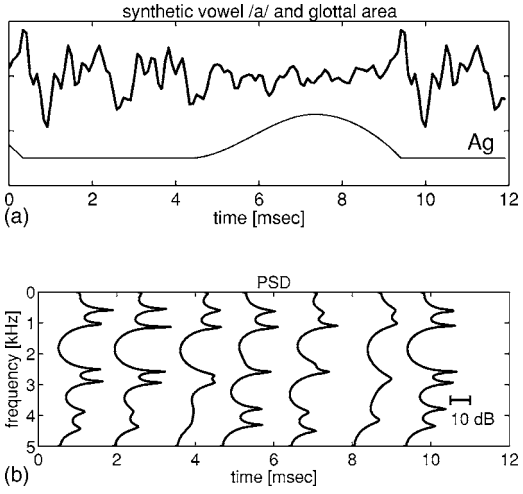


FIG. 10. PSD pattern within a pitch period of synthetic vowel /a/ from subject KH. The thick line in the upper panel (a) denotes the synthetic speech wave form and the thin line shows glottal area A_g used to generate the synthetic speech. The lower panel (b) shows PSDs estimated on a frame-by-frame basis with a 3-ms frame period and a 1.5-ms frame shift.

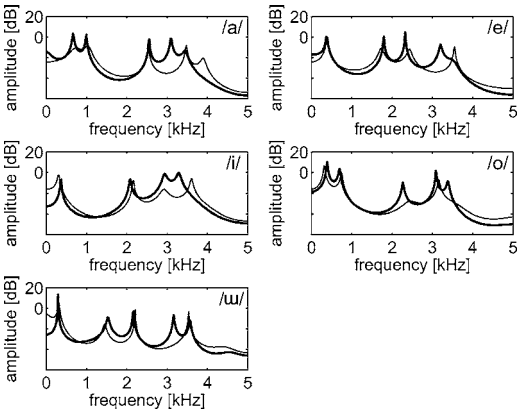


FIG. 11. Power spectral densities of open- and closed-glottis periods of natural speech from subject KH. The thin lines denote those during the open-glottis period and the thick lines represent those in the closed-glottis period.

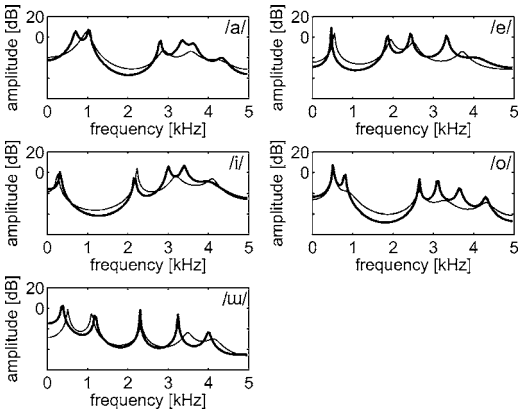


FIG. 13. Power spectral densities of open- and closed-glottis periods of natural speech from subject YT. The thin lines denote those during the open-glottis period and the thick lines represent those in the closed-glottis period.

/a/ from subject KH and the glottal area wave form used to generate the vowel sounds, and Fig. 10(b) illustrates its frame-based PSDs. The laryngeal cavity resonance at approximately 3 kHz becomes damped and disappears as the frame is shifted from the closed-glottis period to the open-glottis period, and the resonance appears again in the next closed-glottis period. The same results are observed for the other two subjects. These provide evidence that vowel spectra vary considerably within a pitch period.

Figures 11–13 depict PSDs obtained from the natural speech. These PSDs are different from those obtained from the synthetic speech because zeros generated by the piriform fossa affect spectra in the frequency region above approximately 3 kHz for natural speech (Dang and Honda, 1997). The formants in the frequency region from 3.0 to 3.6 kHz disappeared when the glottis was open for all vowels except the vowel /ɯ/ produced by subject YT. Note that, in the PSDs for the vowel /i/ from subject TI shown in Fig. 12, the fourth and fifth formants form a cluster, which appears as a single formant with a broad bandwidth during the closed-glottis period, while during the open-glottis period the fourth formant (the laryngeal cavity resonance) is greatly diminished. With the effective removal of the laryngeal cavity resonance, the number of spectral poles decreases when the glottis is open.

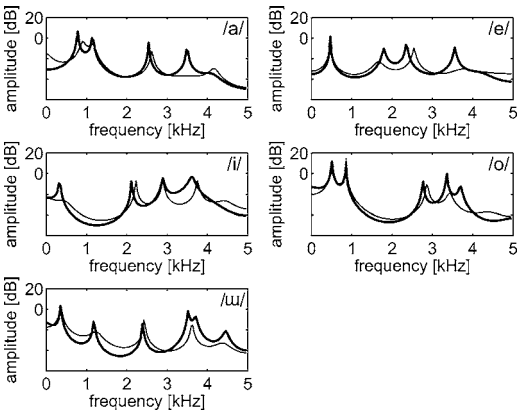


FIG. 12. Power spectral densities of open- and closed-glottis periods of natural speech from subject TI. The thin lines denote those during the open-glottis period and the thick lines represent those in the closed-glottis period.

The model orders for the analysis were optimized by minimizing the criterion function MDL are listed in Table IV, which indicates that the order tends to decrease a few orders when the glottis opens. In the case of subject KH (Fig. 11), for example, the optimized model order decreases from 13 to 11 for the vowel /a/ and from 12 to 10 for the vowel /e/. These spectral changes during vocal fold vibration reveal that the glottal conditions determine laryngeal cavity resonance resulting in a time-varying pattern of resonance within a pitch period in natural vowels as well as synthetic ones.

IV. TIME-DOMAIN OBSERVATION OF LARYNGEAL CAVITY RESONANCE OF OPEN- AND CLOSED-GLOTTIS PERIODS

The time-pattern of the laryngeal cavity resonance was extracted from natural vowels as a bandpass-filtered wave form. This digital filtering was performed by a bandpass filter in order to reveal the time-domain changes of the resonance component in the natural vowels. The resonance appeared and disappeared pitch synchronously as shown in the previous two sections, which allowed us to predict the on-off pattern of the amplitude of bandpass-filter output within a pitch period.

A. Vowel samples

Digital filtering was performed on the sustained vowels of the three male subjects. The vowels were recorded together with EGG wave forms at a sampling rate of 48 kHz with 16-bit resolution in the anechoic room, after which the EGG wave forms were shifted backward to compensate for the time lag between the EGG and speech wave forms as described in Sec. III.

B. Method

A finite impulse response bandpass filter defined with the window method (Rabiner and Gold, 1975; Rabiner *et al.*,

TABLE IV. Model order for short-term spectral analysis for open- and closed-glottis period of natural vowels optimized by minimizing the minimum description length.

Subject	Glottal condition	/a/	/e/	/i/	/o/	/ɯ/
KH	Closed	13	12	10	11	14
	Open	11	10	11	11	13
TI	Closed	11	10	13	12	13
	Open	11	10	11	11	11
YT	Closed	14	11	13	13	12
	Open	11	10	10	10	10

1979) was used to extract the component of the laryngeal cavity resonance. The method permits a flexible design to obtain linear-phase filters. If we denote the Fourier series coefficients of the frequency response of a digital filter as $h(n)$, for $-\infty \leqslant n \leqslant \infty$, and a window as $w(n)$, for $-N \leqslant n \leqslant N$, the impulse response of the windowed digital filter $\hat{h}(n)$ is given as

$$\hat{h}(n) = \begin{cases} w(n)h(n) & 0 \leqslant n \leqslant 2N, \\ 0 & \text{otherwise}, \end{cases} \tag{6}$$

where $\hat{h}(n)$ incorporates an ideal delay of N samples. We used a Hamming window for $w(n)$ and set N to 100.

The speech was processed in four bandpass filter conditions with the four passbands including the subject's second, third, fourth, and fifth formants. The fourth formant is the target of the filtering assuming that the laryngeal cavity resonance corresponds to the fourth formant for the subjects, and the other formant filterings are for comparison. The filter passbands for the subjects' vowels /a/ and /i/ are listed in Table V. The bandwidth of all bandpass filters is set to 0.4 kHz.

C. Results

Figures 14–16 show bandpass-filtered speech wave forms for 30 ms of steady phonation of the vowels /a/ and /i/, along with the speech and EGG wave forms for the subjects. The amplitude of the wave forms is normalized for each sample. The bandpass-filtered speech wave forms were then shifted forward to compensate for the delay caused by the filters. Since EGG wave forms index the changes in the contact area of the vocal folds, a positive excursion in the EGG wave form amplitude corresponds with a closed-glottis period.

The output wave forms of the fourth formant filter show a rise-and-fall pattern in amplitude in each vowel for all the subjects. This time pattern of the filtered wave forms corresponds to that of the EGG wave forms: the amplitude increases in the greater vocal fold contact area and decreases in the smaller area. This result indicates that the glottal closure indeed contributes to generating the laryngeal cavity resonance. The rise in amplitude of the filtered wave forms is more rapid and the fall in amplitude is more gradual, which also corresponds to the fact that glottal closure tends to be faster than glottal opening. This characteristic pattern of the laryngeal cavity resonance is not observed in the output wave forms of the second formant filter: the amplitude decreases gradually from the closed to open periods due to the gradual effect of vocal tract damping between instances of glottal closure. Also, the output wave forms of the third and fifth formant filters does not exhibit the rise-and-fall pattern except in cases where these formant frequencies are close to the fourth formant frequency. For example, in the case of the vowel /i/ from subject KH (Fig. 14), the third and fourth formant frequencies are close (2.95 and 3.25 kHz, respectively) as shown in Fig. 11, and the amplitude of these formants therefore depends on that of the fourth formant. The results indicate that the characteristic pattern of the laryngeal cavity resonance is caused not by frequency-dependent loss but by the glottal states. To summarize, the laryngeal cavity resonance is strongly affected by the glottal condition resulting in the characteristic on-off pattern, while the vocal tract proper resonance simply decays in a glottal cycle by means of damping.

V. DISCUSSION

In the classical vocal tract resonance model, it has been assumed that all formants are generated by resonance of the entire vocal tract and that no regional resonance takes place within the vocal tract (Chiba and Kajiyama, 1941). The acoustic role of the laryngeal cavity has been discussed in relation to generation of the "singing formant" (Sundberg,

TABLE V. Bandpass filter passbands including the subject's second ($F2$), third ($F3$), fourth ($F4$), and fifth formants ($F5$). The bandwidth of all bandpass filters is set to 0.4 kHz.

Subject	Vowel	$F2$ (kHz)	$F3$ (kHz)	$F4$ (kHz)	$F5$ (kHz)
KH	/a/	0.90–1.30	2.05–2.45	2.90–3.30	3.25–3.65
	/i/	1.90–2.30	2.75–3.15	3.05–3.45	3.40–3.80
TI	/a/	0.95–1.35	2.40–2.80	3.30–3.70	3.95–4.35
	/i/	2.00–2.40	2.70–3.10	3.40–3.80	3.55–3.95
YT	/a/	0.80–1.20	2.60–3.00	3.15–3.55	3.40–3.80
	/i/	2.00–2.40	2.80–3.20	3.20–3.60	3.80–4.20

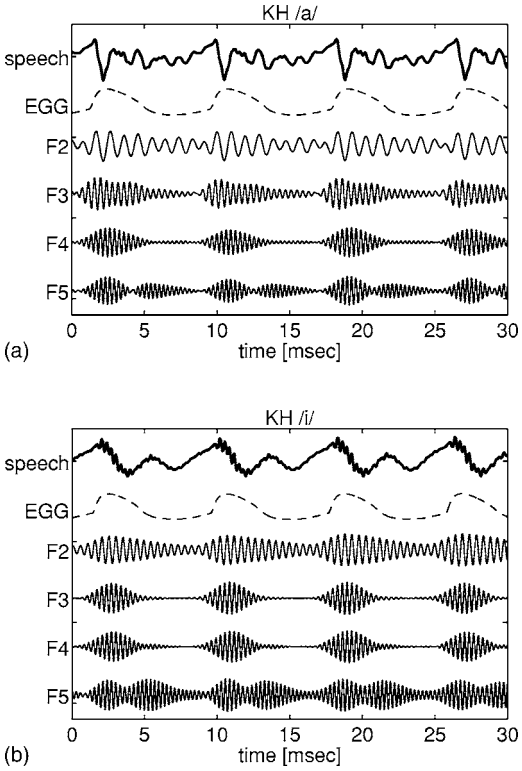


FIG. 14. Outputs of bandpass filters for which the passband includes the second, third, fourth, and fifth formants for steady phonation from subject KH. The upper and lower panels show the results for the vowels /a/ and /i/, respectively. The lines represent from top to bottom, speech wave form, EGG wave form, and bandpass-filtered speech around the formant frequencies.

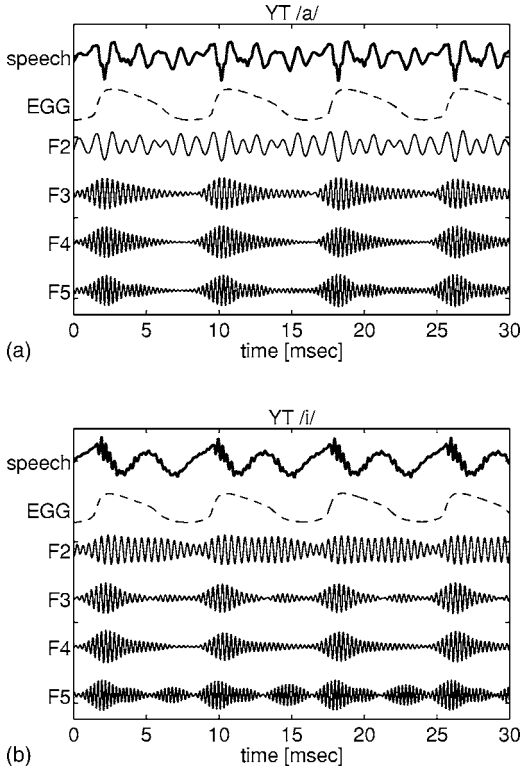


FIG. 16. Outputs of bandpass filters for which the passband includes the second, third, fourth, and fifth formants for steady phonation from subject YT. The upper and lower panels show the results for the vowels /a/ and /i/, respectively. The lines represent from top to bottom, speech wave form, EGG wave form, and bandpass-filtered speech around the formant frequencies.

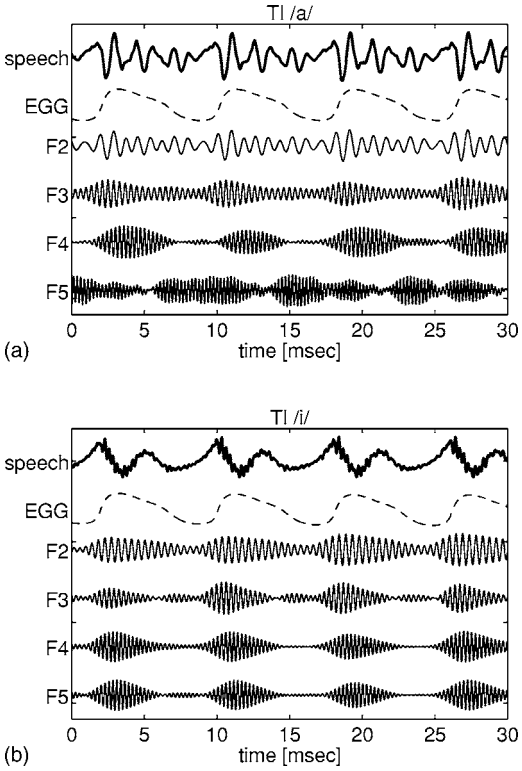


FIG. 15. Outputs of bandpass filters for which the passband includes the second, third, fourth, and fifth formants for steady phonation from subject TI. The upper and lower panels show the results for the vowels /a/ and /i/, respectively. The lines represent from top to bottom, speech wave form, EGG wave form, and bandpass-filtered speech around the formant frequencies.

1974), or small effects on higher formants (Fant, 1960, 1975, 1976, 1980; Fant and Pauli 1975; Fant and Båvegård, 1997; Titze and Story, 1997; Imagawa *et al.*, 2003; Story, 2004). Although it is suggested in the studies on the singing formant and those on the vocal tract energy function that the resonance of the laryngeal cavity can be independent from the rest of the vocal tract, the general understanding among the researchers is that the laryngeal cavity only modifies a few higher-resonance frequencies of the vocal tract. Previous studies presuppose that all the formants reflect resonance of a closed tube and note that glottal opening due to vocal fold vibration increases formant frequencies and bandwidths, greater in the lower frequencies, but only slightly in the higher frequencies (Tarnoczy, 1962; Fujimura and Lindqvist, 1971; Flanagan, 1972).

The comparisons between the vocal tract transfer functions computed under the open- and closed-glottis conditions (Sec. II) show that the laryngeal cavity resonance appears when the glottis is closed and disappears when the glottis is open. The resonance also disappears when the laryngeal cavity region is excluded in computation of vocal tract resonance, consistent with Takemoto *et al.* (2005). The results indicate that the formants do not correspond to resonance of the entire vocal tract but to respective resonances of the vocal tract proper and the laryngeal cavity; that is, a regional resonance takes place within the vocal tract. These facts lend support to the acoustic model of vowel production proposed

by Honda *et al.* (2004), which presumes the vocal tract to be a cascade filter of the vocal tract proper and the hypopharyngeal cavities.

The simulation results also show that the formants other than the laryngeal cavity resonance shift toward the missing resonance frequency under the open-glottis condition, in contradiction to the previous studies. The spectral pattern in the higher frequencies observed when the glottis is open resembles that of the vocal tract proper, while the lower formants are slightly different due to the glottal opening. Such results suggest that under the open-glottis condition the entire vocal tract is acoustically equivalent to the vocal tract proper terminated with a large loss at its closed end.

The laryngeal cavity resonance of the vocal tract transfer functions lies in the frequency region from 3.0 to 3.3 kHz for subject KH, from 3.4 to 3.6 kHz for subject TI, and from 3.2 to 3.7 kHz for subject YT, and it varies up to 0.5 kHz among the vowels for each subject. According to Takemoto *et al.* (2005), the resonance frequency depends on two factors: individual geometry of the vestibular and ventricular spaces of the larynx, and acoustic characteristics (the input impedance) of the vocal tract proper. Widening of the laryngeal vestibule area or narrowing of the laryngeal ventricle raises the resonance frequency, and varying the shape of the vocal tract proper also shifts the resonance frequency.

The short-term spectral analysis of open- and closed-glottis periods of the synthetic and natural speech (Sec. III) and the time-domain analysis of the natural speech (Sec. IV) reveal that the laryngeal cavity resonance appears when the glottis is closed and disappears when it is open. These results indicate that the vocal tract resonance pattern varies significantly within a pitch period, and spectra calculated from speech samples containing a couple of glottal cycles are thus the time-averaged ones of the respective spectra or the two glottal periods. The number of spectral poles tends to decrease when the glottis opens as noted in Table IV. These facts support the validity of approaches to improve the estimation of linear prediction coding parameters by selecting or overweighting closed-glottis periods (Steiglitz and Dickinson, 1977; Miyoshi *et al.*, 1987; Ma *et al.*, 1933).

VI. CONCLUSIONS

In this study, we observed the pitch synchronous pattern of the laryngeal cavity resonance due to vocal fold vibration. The investigation using the three methods explored the characteristic time-varying pattern of the resonance. In a simulation using the transmission line model, we examined the effects of the glottal area on the resonance, with the results showing that glottal closure contributes to generating the resonance. The measured spectra of the samples from the open- and closed-glottis periods also demonstrate the pitch-synchronous pattern of the resonance. The bandpass filter analysis reveals a rise-and-fall pattern of the spectrum amplitude level around the resonance while the other formant shows a gradual damping within a pitch period. These results can be interpreted as follows: the laryngeal cavity resonance is a regional acoustic phenomenon within the cavity that arises when the cavity acts as a closed tube. Therefore, the resonance appears when the glottis is closed and disappears when it is open. Thus, the laryngeal cavity resonance during vocal fold vibration is cyclic in nature.

Laryngeal cavity resonance modifies the spectral pattern in a frequency region around 3.0–3.7 kHz, where the human ear is most sensitive (Fletcher and Munson, 1933; Suzuki and Takeshima, 2004). Because the same frequency region is included in the spectral band containing individual information (Furui and Aakgi, 1985; Kitamura and Akagi, 1995), the resonance contributes not only to coloring vocal sounds but also to giving rise to individual vocal characteristics. Kitamura *et al.* (2005) reported, based on MRI observations, that the shape of the hypopharyngeal cavities is relatively stable regardless of the vowel while displaying a large degree of interspeaker variation, and they concluded that the hypopharyngeal resonance (i.e., the resonance of the laryngeal cavity and the antiresonance of the piriform fossa) constitutes a causal factor of speaker characteristics. In evidence, the hypopharyngeal resonance is observed to be more stable than other formants among vowels of each speaker, while they vary to a greater extent from speaker to speaker. Since the gross shape of the vocal tract also contributes to speaker characteristics (Yang and Kasuya, 1996; Apostol *et al.*, 2004), we consider it reasonable to state that the vocal tract proper and hypopharyngeal cavities together determine individual characteristics of vocal sounds, as proposed by Honda *et al.* (2004). Considering the fact that the laryngeal cavity resonance is only additive to and thus independent from the resonance of the vocal tract proper, separation of the two resonance components in the vocal tract could be possible through future work involving analysis of spectra obtained from open- and closed-glottis periods.

Although the laryngeal cavity resonance appeared in the frequency region around 3.0–3.7 kHz as the fourth formant for the male subjects of this study, for other speakers it may appear in a somewhat lower or higher frequency region as the third or fifth formant, depending on the length and shape of the laryngeal cavity in relation to the overall vocal tract length. Other issues that remain to be studied include the frequency and intraperiod temporal pattern of the laryngeal resonance for females, whose vocal tract length is generally shorter and whose proportion of laryngeal cavity to vocal tract length may therefore be different compared with males.

In the present study, we observed the laryngeal cavity resonance of the vowels in modal phonation, in which the vocal folds vibrate with full contact during glottal closure. The resonance may not always appear in speech produced with nonmodal phonation types, such as breathy phonation with air leakage at the glottis due to incomplete glottal closure. Thus the time pattern of the laryngeal cavity resonance could be used as an index to estimate intraspeaker differences in glottal conditions as observed by voice qualities, speaking styles, degrees of vocal effort, or vocal pathologies with incomplete glottal closure. Since the laryngeal cavity resonance may not be just on-and-off, but reflects geometrical changes of the ventricle during vocal fold vibration, a more detailed analysis of the time pattern may reveal microscopic dynamics of vocal fold vibration such as vertical motion of the folds during a glottal cycle. The relationship be-

tween glottal conditions and laryngeal cavity resonance patterns needs to be investigated to establish a method of evaluating glottal conditions through parametrization of the resonance's time pattern.

ACKNOWLEDGMENTS

This research was conducted as part of "Research on Human Communication" with funding from the National Institute of Information and Communications Technology. The authors wish to thank Dr. Shinji Maeda, CNRS, Dr. Hideki Kawahara, Wakayama University, and Hiroyuki Hirai, SANYO Electric Co., Ltd. for their helpful comments.

Apostol, L., Perrier, P., and Baily, G. (**2004**). "A model of acoustic inter-speaker variability based on the concept of formant-cavity affiliation," J. Acoust. Soc. Am. **115**, 337–351.

Baken, R. J., and Orlikoff, R. F. (**2000**). *Clinical Measurement of Speech and Voice*, 2nd ed., (Singular, San Diego), Chap. 10, pp. 413–427.

Bartholomew, W. T. (**1934**). "A physical definition of good voice-quality in the male voice," J. Acoust. Soc. Am. **6**, 25–33.

Caussé, R., Kergomard, J., and Lurton, X. (**1984**). "Input impedance of brass musical instruments—Comparison between experiment and numerical models," J. Acoust. Soc. Am. **75**, 241–254.

Chiba, T., and Kajiyama, M. (**1941**). *The Vowel—Its Nature and Structure* (Tokyo-Kaiseikan, Tokyo), Chap. 11, pp. 146–147.

Dang, J., and Honda, K. (**1997**). "Acoustic characteristics of the piriform fossa in models and humans," J. Acoust. Soc. Am. **101**, 456–465.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague, Paris), Chap. 2.3, pp. 119–122.

Fant, G., and Pauli, S. (**1975**). "Spatial characteristics of vocal tract resonance modes," in *Speech Communication*, Proc. Speech Communication Seminar, edited by G. Fant (Almqvist & Wiksell, Stockholm) Vol. **2**, 121–132.

Fant, G. (**1975**). "Vocal-tract area and length perturbations," STL-QPSR **4**, 1–14.

Fant, G. (**1976**). "Vocal tract energy functions and non-uniform scaling," ASJ Trans. of the Com. on Speech Res. **S76-24**, 1–18.

Fant, G. (**1980**). "The relation between area functions and the acoustical signal," Phonetica **37**, 55–86.

Fant, G., and Båvegård, M. (**1997**). "Parametric model of VT area functions: Vowels and consonants," TMH-QPSR **31**, 1–20.

Flanagan, J. L. (**1972**). *Speech Analysis Synthesis and Perception*, 2nd ed. (Springer, Berlin), pp. 63–65.

Fletcher, H., and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," J. Acoust. Soc. Am. **5**, 82–108.

Fujimura, O., and Lindqvist, J. (**1971**). "Sweep-tone measurements of vocal-tract characteristics," J. Acoust. Soc. Am. **49**, 541–558.

Furui, S., and Aakgi, M. (**1985**). "Perception of voice individuality and physical correlates," Tech. Rep. Hear. Acoust. Soc. Jpn. **H85-18**, 1–8.

Honda, K., Takemoto, H., Kitamura, T., Fujita, S., and Takano, S. (**2004**). "Exploring human speech production mechanisms by MRI," IEICE Trans. Inf. Syst. **E87-D**, 1050–1058.

Imagawa, H., Sakakibara, K., Tayama, N., and Niimi, S. (**2003**). "The effect of the hypopharyngeal and supra-glottic shapes on the singing voice," *Proc. Stockholm Music Acoust. Conf.* pp. 471–474.

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.

Kitamura, T., and Akagi, M. (**1995**). "Speaker individualities in speech spectral envelopes," J. Acoust. Soc. Jpn. (E) **16**, 283–289.

Kitamura, T., Honda, K., and Takemoto, H. (**2005**). "Individual variation of the hypopharyngeal cavities and its acoustic effects," Acoust. Sci. & Tech. **26**, 16–26.

Lewis, D. (**1936**). "Vocal resonance," J. Acoust. Soc. Am. **8**, 91–99.

Ma, C., Kamp, Y., and Williams, L. F. (**1993**). "Robust signal selection for linear prediction analysis of voiced speech," Speech Commun. **12**, 69–81.

Maeda, S. (**1982**). "A digital simulation method of the vocal-tract system," Speech Commun. **1**, 199–229.

Marple, S. L., Jr. (**1987**). *Digital Spectral Analysis with Applications* (Prentice-Hall, Englewood Cliffs, NJ), Chap. 8, pp. 206–238.

Masaki, S., Tiede, M., and Honda, K. (**1999**). "MRI-based speech production study using a synchronized sampling method," J. Acoust. Soc. Jpn. (E) **20**, 375–379.

Miyoshi, Y., Yamato, K., Mizoguchi, R., Yanagida, M., and Kakusho, O. (**1987**). "Analysis of speech signals of short pitch period by a sample-selective linear prediction," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**, 1233–1239.

Nota, Y., Kitamura, T., Honda, K., Takemoto, H., Hirata, H., Shimada, Y., Fujimoto, I., Shakudo, Y., and Masaki, S. (in press). "A bone-conduction system for auditory stimulation in MRI," Acoust. Sci. & Tech.

Rabiner, L. R., and Gold, B. (**1975**). *Theory and Application of Digital Waveform Processing* (Prentice-Hall, Englewood Cliffs, NJ), Chap. 3, pp. 88–105.

Rabiner, L. R., McGonegal, C. A., and Paul, D. (**1979**). "FIR windowed filter design program—WINDOW," in *Programs for Digital Waveform Processing* (IEEE, New York), Sec. 5.2, pp. 1–19.

Rissanen, J. (**1983**). "A universal prior for the integers and estimation by minimum description length," Ann. Stat. **11**, 417–431.

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

Steiglitz, K., and Dickinson, B. (**1977**). "The use of time-domain selection for improved linear prediction," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-25**, 34–39.

Story, B. H. (**2004**). "Vowel acoustics for speaking and singing," Acta. Acust. Acust. **90**, 629–640.

Sundberg, J. (**1974**). "Articulatory interpretation of the singing formant," J. Acoust. Soc. Am. **55**, 838–844.

Suzuki, Y., and Takeshima, H. (**2004**). "Equal-loudness-level contours for pure tones," J. Acoust. Soc. Am. **116**, 918–933.

Takano, S., Kinoshita, K., and Honda, K. (**2005**). "Measurement of cricothyroid articulation using high-resolution MRI and 3D pattern matching," *Proc. MAVEBA2005*, pp. 141–144.

Takemoto, H., Kitamura, T., Nishimoto, H., and Honda, K. (**2004**). "A method of tooth superimposition of MRI data for accurate measurement of vocal tract shape and dimensions," Acoust. Sci. & Tech. **25**, 468–474.

Takemoto, H., Adachi, S., Kitamura, T., Honda, K., and Mokhtari, P. (**2005**). "Acoustic characteristics of the laryngeal cavity," IEICE Tech. Rep., Speech **195**, 13–17.

Tarnoczy, T. H. (**1962**). "Vowel formant bandwidths and synthetic vowels," J. Acoust. Soc. Am. **34**, 859–860.

Titze, I. R., and Story, B. H. (**1997**). "Acoustic interactions of the voice source with the lower vocal tract," J. Acoust. Soc. Am. **101**, 2234–2243.

Yang, C.-S., and Kasuya, H. (**1996**). "Speaker individualities of vocal tract shapes of Japanese vowels measured by magnetic resonance images," *Proc. ICSLP96*, Vol. **2**, pp. 949–952.

Developmental and cross-linguistic variation in the infant vowel space: The case of Canadian English and Canadian French

Susan Rvachew, Karen Mattock, and Linda Polka
McGill University, Montreal, Quebec H3G 1A8, Canada

Lucie Ménard
Université du Québec à Montréal, C. P. 888, Succursale Centre-Ville, Montéal, Québec H3C 3P8, Canada

(Received 13 September 2005; revised 11 July 2006; accepted 12 July 2006)

This article describes the results of two experiments. Experiment 1 was a cross-sectional study designed to explore developmental and cross-linguistic variation in the vowel space of 10- to 18-month-old infants, exposed to either Canadian English or Canadian French. Acoustic parameters of the infant vowel space were described (specifically the mean and standard deviation of the first and second formant frequencies) and then used to derive the grave, acute, compact, and diffuse features of the vowel space across age. A decline in mean $F1$ with age for French-learning infants and a decline in mean $F2$ with age for English-learning infants was observed. A developmental expansion of the vowel space into the high-front and high-back regions was also evident. In experiment 2, the Variable Linear Articulatory Model was used to model the infant vowel space taking into consideration vocal tract size and morphology. Two simulations were performed, one with full range of movement for all articulatory paramenters, and the other for movement of jaw and lip parameters only. These simulated vowel spaces were used to aid in the interpretation of the developmental changes and cross-linguistic influences on vowel production in experiment 1. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2266460]

PACS number(s): 43.70.Ep, 43.70.Fq, 43.70.Kv [BHS] Pages: 2250–2259

I. INTRODUCTION

A. Rationale

The acoustic characteristics of vowels produced by English-learning infants have been described in a number of prior studies (Buhr, 1980; Gilbert *et al.*, 1997; Kent and Murray, 1982; Robb *et al.*, 1997; Rvachew *et al.* 1996; Sussman *et al.* 1999; 1996). These studies have revealed a strong preference for central vowels, with very little developmental change in the location of the center of the vowel space, and a very gradual expansion of the range of vowels produced with age. These characteristics of infant vowels are interpreted as reflecting the limitations imposed by the structure of the infant's vocal tract and immature speech motor control. The process by which the young child overcomes these physiological limitations to acquire the vowel system of the ambient language is not well understood however. The purpose of this study was to shed some light on this developmental process by shifting the focus from the universal characteristics of the infant vowel space to individual differences in infant vowel production. Specifically, we describe the vowels produced by a relatively large number of infants drawn from a broad age range and two language backgrounds.

B. Background

1. Acoustic characteristics of infant vowels

Kent and Murray (1982) measured the formant frequencies of vocalic utterances produced by 21 English-learning infants aged 3, 6, and 9 months. Mean first formant ($F1$) and second formant ($F2$) values remained relatively stable across the three age groups of infants, (approximately 900–1000 Hz for $F1$ and 3000 Hz for $F2$). However, the range of first and second formant frequencies progressively increased with age, indicating an expansion of the vowel space. A preference for midfront or central vowels was observed throughout the 3-to 9-month age range. Buhr (1980) reported similar findings for a single infant who demonstrated a gradual growth of the vowel space along the $F1-F2$ dimensions between 16 and 64 weeks of age. Reshaping of the vowel space was also observed with the acute region becoming more defined at an earlier age than the grave corner of the vowel space.

Rvachew *et al.* (1996) described the vowels produced by nine infants, followed longitudinally from 6 to 18 months of age. The mean and standard deviation of first formant frequencies were stable throughout this period. A large and steady increase in the range of $F2$ values was observed during the period of the study. The mean $F1$ observed by Rvachew *et al.* (1996) was similar to that reported by Kent and Murray (1982) for younger infants but the mean $F2$ was considerably lower at approximately 2400 Hz. A small reduction of the mean $F2$ was observed during the latter half of the observation period.

Robb *et al.* (1997) described the vowels produced by 20 children aged 4 to 25 months in a cross-sectional study. Contrary to expectations, no decrease with age in mean $F2$ or $F1$ was apparent. In a longitudinal study of four infants between the ages of 15 and 36 months, Gilbert *et al.* (1997) did find a significant lowering of $F1$ and $F2$, but only between 24 and 36 months.

 0001-4966/2006/120(4)/2250/10/$22.50

2. Role of physiological limitations

Several factors can explain the observed acoustic shifts in vowel production across the infancy period: anatomical growth, motor control development, auditory (peripheral and central) abilities, and other cognitive factors. The structure of the human vocal tract is obviously an important determinant of the acoustic characteristics of speech sounds. Studies using magnetic resonance imaging (MRI) confirm the long-standing impression that vocal tract development does not involve a simple linear increase in vocal tract length (Fitch and Giedd, 1999). Major developmental changes in vocal tract structure that occur shortly after birth include the descent of the larynx, lengthening of the pharyngeal cavity, and a sharper angle between the oral and pharyngeal cavities (Kent and Vorperian, 1995). Kent *et al.* (1999) described the growth of the supralaryngeal vocal tract in one infant who received repeated MRI scans between birth and 30 months of age. They found that changes in the size of vocal tract structures were generally coordinated, even during growth spurts at 1 and 4 months of age and 12 and 15 months of age. Most of the increase in vocal tract length in the infants' first year could be explained by the descent of the larynx and tongue, whereas the lengthening of the hard palate made a greater contribution to vocal tract growth during the second year of life.

The influence of the morphology of the vocal tract on the acoustic characteristics of infant vowels has been investigated in studies in which the Variable Linear Articulatory Model was used to synthesize vowels that would be produced by vocal tracts having the dimensions observed for different age groups, specifically a 4-week-old infant, 2-, 4-, 8-, and 12-year-old children, a 16-year-old adolescent, and a 21-year-old adult male (Ménard *et al.*, 2002; 2004). Listener judgments of the resulting vowels indicated that the infant's vocal tract anatomy does not prevent the production of the full range of vowels used in the ambient language. At the same time, infant vocal tract anatomy does at least partly explain infant production preferences: When the maximal vowel space is plotted for the infant and adult vocal tracts, a larger portion of the infant vowel space corresponds to vowels that would be perceived to be low or front vowels, when compared to the adult vowel space. It is also important to note that while it is possible to produce vowels with an infant vocal tract that are perceptually equivalent to adult vowel categories, in many cases the infant would need to employ different articulatory gestures than the adult to achieve the same perceptual outcome.

The finding that the full range of vowel contrasts can be produced with the modeled infant vocal tract assumes adult-like levels of motor control, which is obviously not the case in natural speech. Green *et al.* (2000) measured temporal and spatial coupling of upper lip, lower lip, and jaw movements, during the production of [baba], [mama], and [papa], in 1-, 2-, and 6-year-old children and adults. Jaw movements were dominant, although poorly controlled with respect to force, in the 1-year-old children. By age 2, lip movements were more integrated with the jaw movement. Between ages 2 and 6 years, progressive differentiation of the rigid coupling of upper and lower lip movements was observed. The comparison of movement patterns for 6-years-old children and adults indicated continual refinements in movement control and coordination. Green *et al.* (2002) confirmed the developmental pattern of increased integration of lip movement control into a previously stabilized pattern of jaw movements between 1 and 2 years of age. These data are consistent with the prediction that the development of speech production involves an initial dominance of the "mandibular frame" followed by a progressive differentiation of articulator movements. However, the limitations imposed by immature speech motor control on the development of the infant's speech production abilities do not preclude a role for the auditory environment in shaping the nature of the infant's vocalizations.

3. Role of the auditory environment

Auditory input is clearly critical to the normal development of speech, right from birth. The canonical babbling stage is delayed or never achieved by infants with sensory-neural hearing impairment (see Oller, 2000 for a review of this literature) because hearing impairment interferes with the child's access to both self-produced and other-produced speech (Koopmans-van Beinum *et al.*, 2001). The specific phonetic content of adult speech may shape infant speech production patterns. Kuhl and Meltzoff (1996) manipulated the phonetic content of speech input to the infant in the laboratory by presenting one of three point vowels to different groups of infants aged 12, 16, or 20 weeks. Infants of all ages shifted the acoustic characteristics of their vowels toward the modeled vowel category.

Another strategy for examining the role of speech input is to study cross-linguistic variation in speech production. de Boysson-Bardies, *et al.* (1989) described the acoustic characteristics of the vowel space of 20 10-month-old infants being raised in monolingual French-, English-, Algerian-, and Cantonese-speaking families. They found support for the influence of the ambient language environment on the vowel formants, with variation in mean $F1$ and $F2$ frequencies being greater between language groups than within language groups. English-learning infants' mean $F2$ values were slightly higher than French-learning infants' mean $F2$, but mean $F1$ values were similar for the English- and French-learning infants. Their data suggest that there are systematic and language-specific differences in the articulatory movements produced by infants from different language backgrounds during the first year of life. However this study described only a single age group and a replication has not been published.

The purpose of the current study was to replicate these findings with infant learners of Canadian English (CE) or Canadian French (CF). Recent studies of the adult vowel productions of these languages indicate that CF and CE vowels are characterized by significant acoustic-phonetic differences even where there is phonological overlap (Escudero and Polka, 2003; LaCharite and Paradis, 1997; Martin, 2002). Specifically, the CF /i/ is more diffuse relative to CE /i/, and the CF /u/ is more grave relative to CE /u/. The CF /a/ is slightly less compact than CE /a/. The most acute vowel in CE is [æ], a vowel that is produced allophonically but not phonemically in CF. The acute corner of the CF vowel space

appears to be less acute in comparison with CE. These data on vowels produced by adult speakers do not lead to specific predictions about the potential differences between the vowel spaces produced by infant learners of CE and CF because of the differences in the procedures used to obtain and describe speech samples produced by adult and infant talkers. None the less, the fact that there are significant differences in the acoustic-phonetic characteristics of the adult CE and CF vowel spaces supports the hypothesis that there may be cross-linguistic differences in the acoustic characteristics of vowels produced by infants who are exposed to one of these languages.

C. The current studies

The purpose of experiment 1 was to systematically examine developmental changes and cross-linguistic differences in the first and second formant frequencies of vowels. In this cross-sectional study, we recorded speech samples from 23 infants exposed to Canadian French and 20 infants exposed to Canadian English, aged between 10 and 18 months. Acoustic analyses were used to describe the frequency locations of the center and the corners of each infant's vowel space. Although phonetic transcriptions facilitate a direct comparison of infant and adult phonetic repertoires, this type of analysis was avoided. Oller (2000) has questioned the validity of phonetic transcription for the description of infant speech on a number of grounds, three of which are particularly relevant to this investigation. First, phonetic transcriptions of infant vowels are notoriously unreliable, especially for the identification of back vowels (e.g., Davis and MacNeilage, 1995). Second, phonetic transcriptions are subject to listener biases that are particularly acute when listening to non-native speech sounds. Third, phonetic descriptions of infant speech imply, unrealistically, that infant vocalizations are composed of the same articulatory features that characterize adult produced phonemes. As Ménard *et al.* (2002, 2004) explained, listener perceptions of infant speech do not reliably point to the underlying articulatory gestures that produced the perceived vowel. Thus, for this study we describe vowels in terms of raw acoustic parameters, specifically the mean and standard deviation of the $F1$ and $F2$, and in terms of features that are simple linear combinations of the raw acoustic values, namely the acute-grave and compact-diffuse features. These features may be more closely associated with the goals of vowel production than the raw acoustic values, which vary significantly as a function of vocal tract size and shape. Other researchers (e.g., Kuhl *et al.*, 1997) have used these parameters to describe vowel production.

In experiment 2, vowel spaces were modeled on the basis of a simulation of the infant vocal tract at 6, 12, and 18 months of age. The vowel space that would be produced by these vocal tracts was derived in order to aid in the interpretation of the data obtained in experiment 1. This modeling study offered an opportunity to study the sole effect of vocal tract growth on acoustic data.

On the basis of the data reported by de Boysson-Bardies *et al.* (1989), we expected cross-linguistic differences in the infant's vowel productions. Specifically, a progressive divergence of the center of the vowel space for French and English infants, particularly with respect to the $F2$ dimension, was predicted. Changes in the first and second formant frequencies, for either language group, that are greater than would be predicted from simple growth of the vocal tract (as indicated by experiment 2) would lend further support to the hypothesis that the phonetic content of adult speech input has an influence on infant speech output during the first 18 months of life. Language-general changes in the vowel space were also expected, especially with respect to the overall size of the vowel space.

II. EXPERIMENT 1

A. Method

1. Participants

Forty-three typically developing infants from predominately middle-class families were recruited from birth registries for the Montréal region. Each infant was no younger than 300 days and no older than 570 days. All infants were reportedly born between 38 and 42 weeks gestation following uncomplicated pregnancies, with no known history of ear infections or hearing impairment, and were healthy on the day of testing. A parent questionnaire about language use in the home (e.g., by parent, siblings, television, radio), and in the speech directed to their infant from others (e.g., grandparents, babysitter, daycare worker) confirmed that 23 infants were being raised by monolingual CF speaking families, and 20 infants were being raised by monolingual CE speaking families. Thirty-two of the 43 infants passed several audiological screenings (tympanometry and otoacoustic emissions) performed by an audiologist beginning at 2–3 months of age (these infants were initially recruited for another study in our lab). The remaining infants passed a tympanometric screening on the day of the speech sample recording.

a. Speech sample recordings. Samples of the infants' vocalizations were recorded during a play session between mother and infant, either in a sound proof booth in the laboratory or in the infant's home. Mothers were instructed to interact with their infant in the usual manner using a set of quiet toys. Recording sessions continued until the infant produced 60 utterances perceived to meet the utterance selection criteria (described below), or until 30 minute had lapsed, whichever came first. The speech samples were obtained using a portable DAT recorder and a Sennheiser microphone affixed to the infant's clothing at the shoulder. Following recording, all speech utterances were digitized at 22 050 Hz using Time Frequency Response software (Avaaz Innovations) installed on IBM PC hardware equipped with a Creative Labs Live Drive.

b. Acoustic analysis. Each utterance was assigned an Infraphonological code (i.e., canonical syllable, fully resonant vowel, quasiresonant vowel, marginal syllable, squeal, raspberry or growl, using the criteria described in detail by Oller, 1986). Isolated vowels and vowels contained within canonical syllables were selected for formant analysis if vowel or syllable duration was less than 500 ms, and if the vowel had normal phonation, full resonance, and at least two

measurable formants. These utterance types comprised 25% of the sample. The remaining utterances (28% marginal syllables, 47% "other" including quasiresonant vowels, squeals, growls, and raspberries) were not submitted to formant analysis. Seven vowels were discarded from the data set because either $F1$ ($n=2$) or $F2$ ($n=5$) was three standard deviations greater than the mean value. A total of 1190 utterances (665 French, 525 English) met the criteria. Vowel formant analyses were performed blind to the age and language background of the infant. To determine $F1$ and $F2$ frequencies, a 20 ms segment at the middle of the steady state portion was submitted to linear predictive coding (LPC) autocorrelation analysis with a window size of 256 points, 50% overlap, 98% preemphasis, Hanning window, and model order of 12. Model order was increased or decreased accordingly to obtain reliable measurements of some vowels where the formants were difficult to measure. Formant locations for all vowels were confirmed with narrowband short-time FFT spectrograms (512 points). The Peterson and Barney (1952) norms were also referred to in order to confirm that the obtained frequency values roughly approximated the expected relationship between formants, given the perceived quality of the vowel (e.g., a vowel sounding to be /u/-like would be expected to yield $F1$ and $F2$ values that were close in frequency). Replicate acoustic analysis of 299 vowels (25% of full sample) were conducted by a second individual trained in speech acoustics who was blind to the age, language background of the infants, as well as the measurements obtained by the first coder. Vowels were reanalyzed using the same measurement parameters as the first coder (see above). Intraclass correlations between the independently identified formant frequencies were 0.96 and 0.94 for $F1$ and $F2$ respectively. All first and second formant frequencies in hertz were converted to the mel scale (Stevens *et al.* 1937) using the formula

$$F_{\text{mels}} = (1127.010\,481)\ln\left(1 + \frac{F_{\text{hertz}}}{700}\right).$$

c. Statistical analyses. Each infant's vowel space was described using the following eight summary statistics, all expressed in mels: (1) $MF1$—mean of the first formant frequencies; (2) SD $F1$—standard deviation of the first formant frequencies; (3) $MF2$—mean of the second formant frequencies; (4) SD $F2$—standard deviation of the second formant frequencies; (5) *Grave*—minimum value of $(F1+F2)/2$; (6) *Acute*—maximum value of $(F1+F2)/2$; (7) *Compact*—minimum value of $F2-F1$; and (8) *Diffuse*—maximum value of $F2-F1$. The extraction of these summary statistics from an infant's vowel space is illustrated in Fig. 1. The figure shows $F1$ and $F2$ coordinates for each vowel produced by the infant. Superimposed are two bars that indicate the location of the center vowel and the standard deviation of the first and second formant frequencies as a measure of dispersion of formant values around the center vowel. Arrows on the figure indicate the vowels that represent the most *grave, acute, compact* and *diffuse* values in the vowel space.

Regression analysis was used to examine the main effect of language group, the main effect of infant age, and the interaction of language and age on each summary statistic. These analyses revealed interaction effects for many of the summary statistics and consequently, simple regression analyses are reported for the effect of age on each summary statistic, independently for each language group.

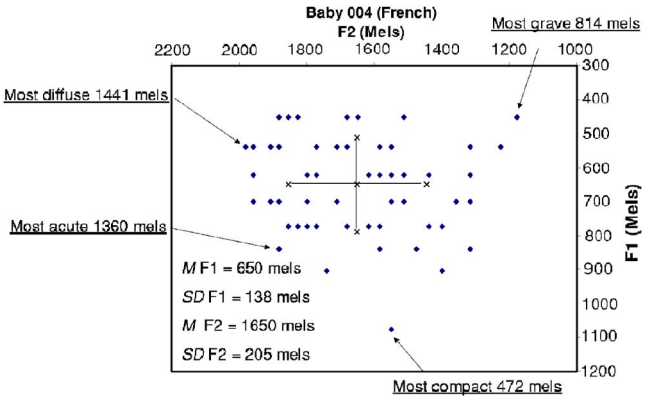


FIG. 1. The vowel space of one French infant. $F1$ and $F2$ coordinates of each vowel produced by the infant are plotted. The bars represent mean and standard deviation of the center vowel. The most grave, acute, compact, and diffuse vowels are indicated by arrows.

B. Results

1. Parameters

Figure 2 (top left) shows a significant decline in $MF1$ from 962 to 730 mels for the French group [$B=-0.86$; SE $=0.30$; $F(1,21)=8.02$, $p=0.01$]. The smaller decline from 913 to 814 mels for the English group was not statistically significant [$B=-0.37$; SE$=0.27$; $F(1,18)=1.89$, $p=0.19$]. Figure 2 (top right) illustrates a small decline in SD $F1$ that was not significant for the French [$B=-0.10$; SE$=0.09$; $F(1,21)=1.39$, $p=0.25$] or the English [$B=-0.05$; SE$=0.08$; $F(1,18)=0.41$, $p=0.53$] group. Figure 2 (bottom left) depicts a significant decline in $MF2$, from 1714 to 1523 mels, for the English group [$B=-0.71$; SE$=0.24$; $F(1,18)=8.64$, $p=0.01$]. The much smaller decline for the French group, from 1667 to 1636 mels, was not statistically significant [$B=-0.11$; SE$=0.26$; $F(1,21)=0.18$, $p=0.68$]. Figure 2 (bottom right) illustrates a significant increase in SD $F2$, from 130 to 245 mels, for the English group [$B=0.43$; SE$=0.13$; $F(1,18)=10.91$, $p=0.00$]. The SD $F2$ for the French group was relatively stable throughout the age range, with a nonsignificant increase from 157 to 175 mels [$B=0.07$; SE $=0.14$; $F(1,21)=0.23$, $p=0.63$].

2. Features

Figure 3 (top left) shows an increase in the maximum value of the *diffuse* feature for both groups, specifically from 1101 to 1423 for the French group [$B=1.19$; SE$=0.58$; $F(1,21)=4.19$, $p=0.05$] and from 1115 to 1311 for the English group [$B=0.73$; SE$=0.59$; $F(1,18)=1.53$, $p=0.23$]. These data suggest a trend toward expansion of the vowel space into the acoustic area associated with widely spaced first and second formant frequencies, at least for the French group.

Figure 3 (top right) depicts a decrease in the *grave* fea-

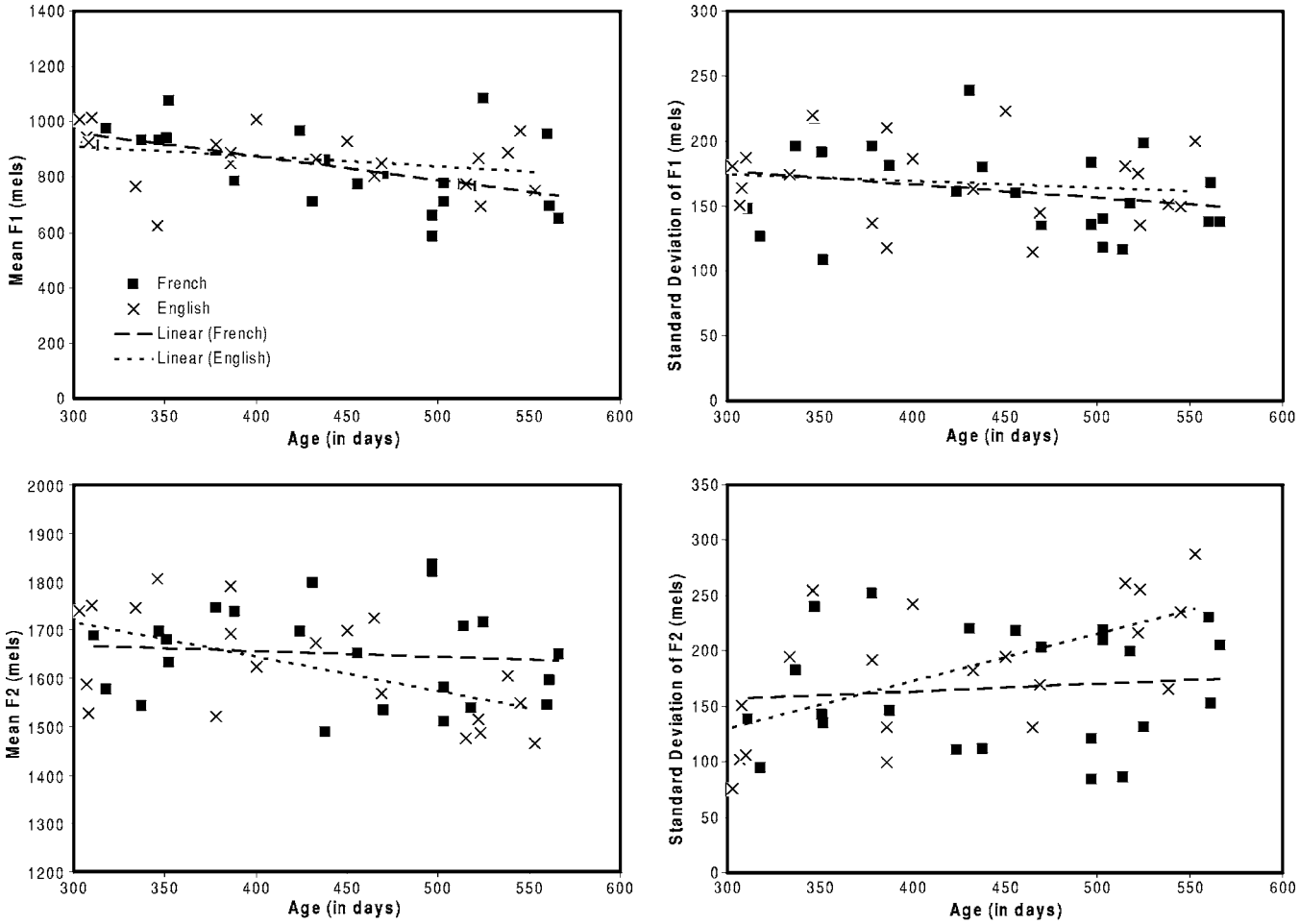


FIG. 2. Speech sample mean of each of four parameters (in mels) plotted for each infant as a function of age and language group, specifically *M F*1 (top left), SD *F*1 (top right), *M F*2 (bottom left), and SD *F*2 (bottom right).

ture that was from 1056 to 907 mels for the French group [$B=-0.55$; SE=0.27; $F(1,21)=4.27$, $p=0.05$] and from 1101 to 843 mels for the English group [$B=-0.96$; SE =0.29; $F(1,18)=10.67$, $p=0.01$]. These values indicate an age-related expansion of the vowel space into the acoustic area associated with low frequency and closely spaced first and second formant frequencies.

Figure 3 (bottom left) illustrates a significant decrease in the *acute* feature from 1545 to 1384 mels for the French group [$B=-0.60$; SE=0.19; $F(1,21)=9.65$, $p=0.01$] and a smaller decrease from 1516 to 1420 mels for the English group [$B=0.36$; SE=0.18; $F(1,18)=4.10$, $p=0.06$]. The decreasing values highlight an age-related compression of the vowel space in the acoustic area associated with relatively high and closely spaced first and second formant frequencies, at least for the French group.

Figure 3 (bottom right) suggests an interaction of age and language group for the *compact* feature. However, the increase from 334 to 440 for the French group was not significant [$B=0.39$; SE=0.55; $F(1,21)=0.51$, $p=0.48$]; the decline from 434 to 304 for the English group was not significant either [$B=-0.48$; SE=0.30; $F(1,18)=2.55$, $p=0.128$].

C. Discussion

Acoustic analyses of vowels produced by infants aged approximately 10 to 18 months indicated the presence of developmental changes that were common to both language groups as well as some significant differences across language groups, as illustrated in Fig. 4. Cross-linguistic variation was apparent in the frequency location of the center of the infants' vowel spaces. Specifically, the French-learning infants demonstrated a significant decline with age in the *MF*1, whereas the English-learning infants produced a significant decline with age in the *MF*2. The English-learning infants demonstrated a significant increase in the dispersion of second formant frequencies as age increased. The French-learning infants did not produce a reliable age-related change in SD *F*2 with age. Neither group showed age-related changes in SD *F*1.

Both groups showed a developmental expansion of the size of the vowel space along the diffuse-grave dimension although expansion into the diffuse corner was greater for the French group and expansion into the grave corner was greater for the English group. These findings suggest a developmental expansion into the areas of the vowel space traditionally associated with tongue retraction and advancement in adult articulation. Compression of the vowel space in the acute corner, particularly marked for the French group, suggests less extreme jaw opening gestures with age.

The distribution of the infants' vowels within this global vowel production space appears to differ across the language groups. At the same time, English-learning and French-

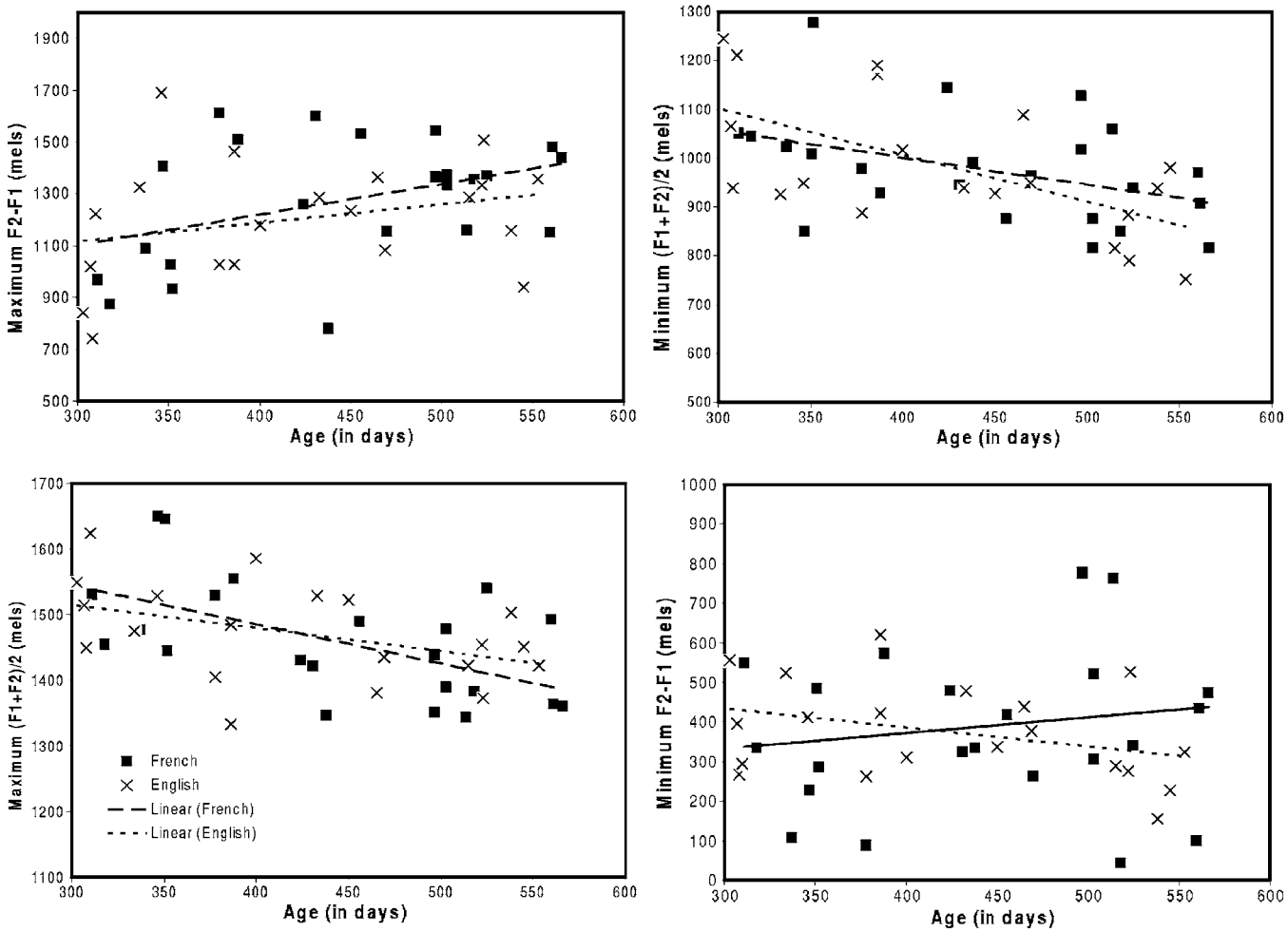


FIG. 3. Feature value for each infant's corner vowels plotted as a function of age and language group, specifically *diffuse* (top left), *grave* (top right), *acute* (bottom left), and *compact* (bottom right).

learning infants demonstrated expansion of the vowel space between 10 and 18 months. In order to interpret these developmental changes in relation to the growth of the vocal tract during this developmental period, a simulation experiment was conducted.

III. EXPERIMENT 2

A. Method

The maximal vowel spaces that could be produced by an infant, aged 6-, 12-, or 18-months of age, were modeled using the Variable Linear Articulatory Model (VLAM). This model integrates the growth data currently available (Goldstein 1980) into a previous model already existing for the adult (Maeda, 1979; 1990). The latter is based on a statistical analysis of 519 midsagittal cineradiographic images of a French speaker uttering ten sentences (Bothorel *et al.*, 1986). The analysis revealed that seven articulatory parameters (P_i, $i \in \{1, \ldots, 7\}$) could account for 88% of the variance of the tongue contours (Boë *et al.*, 1995): labial protrusion, labial aperture, tongue tip position, tongue body position, tongue dorsum position, jaw height, and larynx height. Each parameter is adjustable at a value in the range of ±3.5 standard deviations around the mean values for this articulator in the cineradiographic images. These parameters control the position of the articulators in the model, and hence the midsagittal contour. The cross-sectional area function is computed from the midsagittal contour following the Heinz and Stevens (1965) formula and the transfer function is calculated following the Badin and Fant (1984) model. VLAM integrates nonuniform vocal tract growth, in the longitudinal dimension, by two scaling factors: one for the oral cavity and another for the pharyngeal cavity, the zone in-between being interpolated. The values of the factors, from 0.3 to 1.2, were calibrated year by year and month by month based on Goldstein's (1980) length data.

Two sets of maximal vowel spaces were simulated, with the first set representing very limited articulatory movement (henceforth limited range simulation) and the second representing the full range of movement for all seven articulatory parameters (henceforth full range simulation). The limited range simulation was accomplished by generating the acoustic properties of all vowels that could be produced given the full range of variation in the jaw and lip height movements, while holding tongue tip position, tongue body position, and larynx height in the neutral position. The full range simulation was accomplished by generating the acoustic properties of all vowels that could be produced given the full range of variation in all seven articulatory parameters.

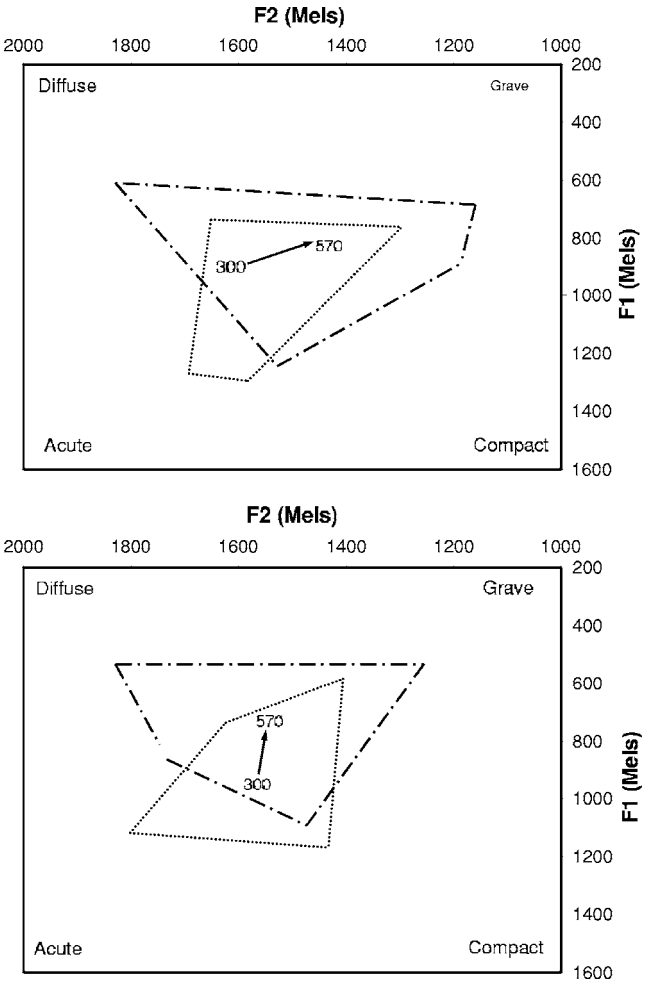


FIG. 4. Graphic summary of the findings for the English-learning infants (top) and French-learning infants (bottom). On both charts the arrow indicates the movement of the center of the vowel space as age increases from 300 to 570 days, the dotted-line quadrilaterals trace the periphery of the vowel space at 300 days, and the dashed-line quadrilaterals trace the periphery of the vowel space at 570 days.

B. Results and discussion

The resulting vowel spaces for the 6-, 12-, and 18 -month vocal tracts were described using the same procedures outlined above for experiment 1. These values for the simulated vowel spaces are shown in Table I, with the limited range simulation represented in the upper half and the full range simulation represented in the lower half of the table. Changes in these values with age, especially as shown for the limited range simulation, largely reflect increasing length of the vocal tract. As would be expected, *MF*1 and *MF*2 decrease with age although the decreases shown are quite small, less than 50 mels on average. Changes in the corners of the vowel space with age are also quite small, decreasing less than 50 mels for each feature. Most change occurred to the grave and acute features. The compact feature shows the smallest degree of age related change in the simulation as it did for the acoustic measures reported in experiment 1. Changes in the acoustic characteristics of the vowel space from the limited range simulation to the full range simulation reflect the increase in the range of vowels that the infant could produce, given full range of movement

TABLE I. Summary statistics for the limited range and full range simulated vowel space parameters for the 6-, 12-, and 18-month-old vocal tract.

Summary statistic	6 months	12 months	18 months
Limited range simulation			
Mean F1	431.00	299.24	383.62
SD F1	15.50	12.09	13.71
Mean F2	960.84	940.98	918.05
SD F2	35.97	42.20	46.38
Grave	669.02	641.43	624.29
Acute	725.26	700.62	683.82
Compact	488.15	491.54	484.38
Diffuse	671.82	680.12	684.47
Full range simulation			
Mean F1	407.73	384.18	373.04
SD F1	30.32	24.55	26.57
Mean F2	914.13	882.02	847.60
SD F2	105.36	92.08	103.97
Grave	526.63	534.88	489.84
Acute	746.93	722.16	707.09
Compact	276.89	283.10	258.86
Diffuse	765.93	722.56	724.04

of all articulators. As shown in the lower part of Table I, small decreases in *Mf*1 and *MF*2 occurred although a large increase in SD *F*2 is shown. Substantial decreases in the grave and compact features and a small increase in the diffuse feature are also apparent when comparing the limited range with the full range simulation. Thus these simulations modeled the expected expansion of the infant vowel space and suggest that this expansion occurs as a consequence of improved speech motor control.

IV. GENERAL DISCUSSION

A. Developmental changes

The most obvious developmental change for the English and French infants was an expansion of the vowel space, especially with respect to the grave and diffuse features. In addition to being consistent with previous findings (e.g., Buhr, 1980; Gilbert *et al.*, 1997; Kent and Murray, 1982; Robb *et al.*, 1997; Rvachew *et al.*, 1996), these changes are a predictable consequence of developmental changes in the infant's ability to control tongue tip, tongue body, and tongue dorsum position, independently of jaw height, as indicated by the simulations reported in experiment 2 and shown in Table I, when comparing the limited and full range simulations.

Improved control of the jaw during the infant period should manifest itself as less extreme jaw displacement during the opening and closing phases of syllable production (Green *et al.*, 2000). More control of the jaw in the midopen position should result in less extreme acuteness values (i.e., reductions in maximum $F1+F2/2$) and greater compactness values (i.e., decreases in minimum $F2-F1$). Figure 3 (bottom left) confirms a statistically significant reduction in acuteness values. No clear developmental effects on compactness values were observed however.

B. Cross-linguistic differences

A priori predictions about likely cross-linguistic differences were much more difficult to formulate because no previous studies have compared vowels produced by Canadian-English- and Canadian-French-learning infants. de Boysson-Bardies *et al.* (1989) reported a similar mean *F*1 but a slightly higher mean *F*2 for the vowel spaces of 10-month-old infants exposed to British English compared to the vowel spaces of infants exposed to Parisian French. The 10-month-old infants enrolled in this study showed a similar pattern of differences in mean formant frequencies. We expected to see a linear divergence of the center of the vowel spaces with age; however, an unexpected interaction of age and language group was observed: *MF*2 decreased in the Canadian-English group while the *MF*2 remained stable across age for the Canadian-French group. Cross-linguistic differences were also observed in *MF*1 as the children grew older: *MF*1 decreased in both groups but the decrease was much greater for the Canadian-French infants than for the Canadian-English group. Although some decrease in *MF*1 and *MF*2 is expected as the infant's vocal tract lengthens, the decrease in *MF*1 observed for the Canadian-French group and the decrease in *MF*2 observed for the Canadian-English group were each much larger than would be predicted on the basis of vocal tract growth alone (see Table I, experiment 2). It is possible that the observed interaction of age and language group on *MF*2 might be due to differences in vowel inventory. English has fewer front vowels and no rounded front vowels, in contrast to French which has three rounded front vowels and three unrounded front vowels, with the front vowels having higher *F*2 than back vowels. The *F*2 decrease in the CE group might be due to the combined effects of vocal tract growth and increased frequency of back vowels, thus decreasing *F*2. While a similar decrease in *F*2 may occur in the CF group for the same reasons, the effect may be balanced by a greater frequency of front vowels, resulting in a stable *MF*2 across age in this language group.

Our findings of a decrease in formant frequencies with age is not consistent with the earlier findings of Gilbert *et al.* (1997) and Robb *et al.* (1997) who found that *F*1 and *F*2 remained stable across age. This difference in findings across studies may be accounted for by utterance selection. In their papers, Gilbert *et al.* (1997) and Robb *et al.* (1997) acknowledge that possible changes in *F*1 and *F*2 may have been obscured by nasal resonance in the younger children's vocalizations. In the present study we controlled for this by only analyzing vowels with normal phonation and full resonance.

A cross-linguistic difference was also observed for the SD *F*2, with the Canadian-English infants showing an increase with age in the range of *F*2 values (as has been reported in other studies; e.g., Robb *et al.*, 1997; Rvachew *et al.*, 1996). The Canadian-French infants did not show this pattern of change for SD *F*2 however.

C. Future directions

These results demonstrate significant developmental changes in the shape of the infant vowel space as well as significant impacts of the auditory environment on the frequency location of the center of the infant vowel space. These patterns of developmental change and cross-linguistic differences appear to emerge after 12 months of age but are clearly evident before 18 months of age. The observed individual differences in vowel production are undoubtedly explained by a complex interaction of factors, including changing vocal tract morphology, developing speech motor control, and the child's intake of self- and other-produced speech. More research is required to understand how these factors determine infant speech output.

Recent technological advances, such as magnetic resonance imaging and computational modeling, allow us to make predictions about the impact of changing vocal tract morphology on the acoustic characteristics of speech output. More direct observation of infant articulatory movements are required however, in order to better model the impact of limited speech motor control along with the limitations imposed by the size and shape of the infant's vocal tract. Kinematic studies of jaw and lip movements indicate that maturation of speech motor control is not a linear process. For example, after examining the correlation between the spatiotemporal trajectories for adult and child jaw movements during bisyllable productions, Green *et al.* (2002) concluded that jaw movements appear to be more adultlike at the end of the first year, than at the end of the second year. Reduced stability of jaw movements at the later age may be due to the challenge of developing independent control of other articulators or the challenge of producing speech for communicative purposes.

The cross-linguistic differences that were observed in this study are difficult to interpret. Presumably, the speech that is heard by the infant provides targets for speech production that shape the specific characteristics of the infant's speech output. The exact nature of these targets is unknown. Although the acoustic characteristics of adult-produced Canadian-English and Canadian-French vowels have been described (Escudero and Polka, 2003), these kinds of descriptions are not well suited to the task of understanding the target for infant speech production. First, these descriptions are based on adult-directed speech, and it has been shown that the acoustic-phonetic nature of infant-directed speech is significantly different from that of adult-directed speech (Kuhl *et al.*, 1997). In particular, the talker's vowel space when addressing an infant is larger, with more extreme point vowels, in comparison with the vowel space produced in an adult-directed register. Second, adult-produced speech is usually described in relation to specific phonetic targets. For example, Escudero and Polka (2003) found that the Canadian-English [u] is considerably less grave than the Canadian-French [u], while the Canadian-English [æ] is more acute that the Canadian-French [æ]. However, since it is not possible to ask infants to produce a specific vowel, the infant vowel space is always described in terms of more global characteristics as we have done here. We are currently engaged in an effort to describe infant-directed speech from Canadian-French- and Canadian-English-speaking parents, in terms of the center and corners of their vowels spaces. This kind of information may facilitate the development of

specific predictions about cross-linguistic differences in infant vowel production across different language groups.

The infant's access to self-produced speech must also be considered. The development of speech motor control requires that the infant develop a mapping between auditory-perceptual targets, articulatory gestures, and the acoustic-phonetic product of those articulatory movements (Callan *et al.*, 2000). This model of speech development highlights the importance of feedback of the infant's own speech. A better understanding of how the infant processes this feedback is necessary if we are to predict patterns of developmental change in speech production. Investigating the role of visual speech (e.g., the visual cues for production of different vowels) is another avenue for further research.

V. CONCLUSIONS

In this study we described the vowel spaces produced by infants in terms of the center and the corners of the vowel space. The infants were drawn from two language groups, Canadian English and Canadian French, and covered a broad age range, from 10 to 18 months. The findings were interpreted in relation to simulations of vowel production given differences in vocal tract length and speech motor control. Some individual differences in the vowel spaces, such as an expansion of the vowel space into the diffuse and grave regions, were associated with ageing of the infant. These developmental changes appear to reflect maturation of the vocal tract and speech motor control. Other differences were associated with the infant's ambient language environment. Infants exposed to Canadian French demonstrated a decline in mean first formant frequencies whereas infants exposed to Canadian English showed a decline in mean second formant frequencies with age. The divergence of the vowel spaces between the two language groups emerged between 12 and 18 months of age. In order to understand the mechanism by which the ambient speech environment influences infant speech production, future research should attempt to link the characteristics of infant vowels to the infant's perception of both adult- and self-produced speech.

ACKNOWLEDGMENTS

This research was supported by research grants from the Canadian Language and Literacy Research Network and the Natural Sciences and Engineering Research Council of Canada to the first author and a postdoctoral fellowship to the second author from the Centre for Research in Language, Mind and Brain. The authors thank the infants and their families for their participation, Voula Tsagaroulis for audiometric testing, Jade Heilmann for recruiting and data collection, Pi-Yu Chiang, Jessica Whittley, Shani Abada, and Heather McKinnon for digitizing the speech samples, and Marie Desmarteau and Sara Turner for acoustic analysis.

Badin, P., and Fant, G. (**1984**). "Notes on vocal tract computation," *Speech Transmission Laboratory—Quarterly Progress Status Report*, Vol. **2–3**, 53–108.

Boë, L.-J., Gabioud, B., and Perrier, P. (**1995**). "The SMIP: An interactive articulatory-acoustic software for speech production studies," Bulletin de la Communication Parlée **3**, 137–154.

Bothorel, A., Simon, P., Wioland, F., and Zerling, J. P. (**1986**). *Cinéradiographie des Voyelles et Consonnes du Français [Cineradiographic study of French vowels and consonants]*, Institut de Phonétique de Strasbourg, Strasbourg, France.

Buhr, R. D. (**1980**). "The emergence of vowels in an infant," J. Speech Lang. Hear. Res. **23**, 73–94.

Callan, D. E., Kent, R. D., Guenther, F. H., and Vorperian, H. K. (**2000**). "An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system," J. Speech Lang. Hear. Res. **43**, 721–738.

Davis, B. L., and MacNeilage, P. F. (**1995**). "The articulatory basis of babbling," J. Speech Lang. Hear. Res. **38**(6), 1199–1211.

de Boysson-Bardies, B., Halle, P., Sagart, L., and Durand, C. (**1989**). "A crosslinguistic investigation of vowel formants in babbling," J. Child Lang **16**, 1–17.

Escudero, P., and Polka, L. (**2003**). "A cross-language study of vowel categorization and vowel acoustics: Canadian English versus Canadian French," 15th International Congress of the Phonetic Sciences, Barcelona, Spain, Vol. **1**, pp. 861–864.

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.

Gilbert, H. R., Robb, M. P., and Chen, Y. (**1997**). "Formant frequency development: 15 to 36 months," J. Voice **11**(3), 260–266.

Goldstein, U. G. (**1980**). "An articulatory model for the vocal tract of growing children," Doctoral dissertation, Massachusetts Institute of Technology, Boston.

Green, J. R., Moore, C. A., Higashikawa, M., and Steeve, R. W. (**2000**). "The physiologic development of speech motor control: Lip and jaw coordination," J. Speech Lang. Hear. Res. **43**(1), 239–255.

Green, J. R., Moore, C. A., and Reilly, K. J. (**2002**). "The sequential development of jaw and lip control for speech," J. Speech Lang. Hear. Res. **45**(1), 66–79.

Heinz, J. M., and Stevens, K. N. (**1965**). "On the relations between lateral cineradiographs, area functions, and acoustic spectra of speech," *Proceedings of the 5th International Congress of Acoustics*, Liege, Belgium, Vol. **1** (Elsevier, Amsterdam), p. A44.

Kent, R. D., and Murray, A. D. (**1982**). "Acoustic features of infant vocalic utterances at 3, 6, and 9 months," J. Acoust. Soc. Am. **72**, 353–365.

Kent, R. D., and Vorperian, H. K. (**1995**). "Anatomic development of the craniofacial-oral laryngeal systems: A review," J. Med. Speech-Language Pathology **3**, 145–190.

Kent, R. D., Vorperian, H. K., Gentry, L. R., and Yandell, B. S. (**1999**). "Magnetic resonance imaging procedures to study the concurrent anatomic development of vocal tract structures: preliminary results," Int. J. Gynecol. Pathol. **49**, 197–206.

Koopmans-van Beinum, F. J., Clement, C. J., and van den Dikkenberg-Pot, I. (**2001**). "Babbling and the lack of auditory speech perception: a matter of coordination?" Developmental Science **4**(1), 61–70.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., and Lacerda, F. (**1997**). "Cross-language analysis of phonetic units in language addressed to infants," Science **277**, 684–686.

Kuhl, P. K., and Meltzoff, A. N. (**1996**). "Infant vocalizations in response to speech: Vocal imitation and developmental change," J. Acoust. Soc. Am. **100**(4), 2425–2438.

LaCharite, D., and Paradis, C. (**1997**). "Category preservation and proximity versus phonetic approximation in loanword adaptation," Linguistic Inquiry **36**, 223–258.

Maeda, S. (**1979**). "An articulatory model of the tongue based on a statistical analysis," J. Acoust. Soc. Am. **65**, S22.

Maeda, S. (**1990**). "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model," *Speech Production and Speech Modelling*, edited by W. L. Hardcastle and A. Marchal, (Kluwer Academic, Dodrecht, The Netherlands), 131–149.

Martin, P. (**2002**). "Le système vocalique du français du Québec. De l'acoustique à la phonologie," La Linguistique **38**(2), 71–88.

Ménard, L., Schwartz, J., and Boë, L. (**2002**). "Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood," J. Acoust. Soc. Am. **111**(4), 1892–1905.

Ménard, L., Schwartz, J., and Boë, L. (**2004**). "Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized vowels from birth to adulthood," J. Speech Lang. Hear. Res.

47, 1059–1080.
Oller, D. K. (**1986**). "Metaphonology and infant vocalizations," in *Precursors of Early Speech* edited by B. Lindblom and R. Zetterstrom (Stockton New York), pp. 21–35.
Oller, D. K. (**2000**). *The Emergence of the Speech Capacity* (Lawrence Erlbaum Associates, Mahwah, New Jersey).
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, pp. 175–184.
Robb, M. P., Chen, Y., and Gilbert, H. R. (**1997**). "Developmental aspects of formant frequency and bandwidth in infants and toddlers," Folia Phoniatr Logop **49**, 88–95.
Rvachew, S., Slawinski, E. B., Williams, M., and Green, C. L. (**1996**). "Formant frequencies of vowels produced by infants with and without early onset otitis media," Can. Acoust. **24**, 19–28.
Stevens, S. S., Volkmann, E. B., and Newman, J. (**1937**). "The mel scale equates the magnitude of perceived differences in pitch at different frequencies," J. Acoust. Soc. Am. **8**, 185–190.
Sussman, H. M., Duder, C., Dalston, E., and Caciatore, A. (**1999**). "An acoustic analysis of the development of CV coarticulation: A case study," J. Speech Lang. Hear. Res. **42**(5), 1080–1096.
Sussman, H. M., Minifie, F. D., Buder, E. H., Stoel-Gammon, C., and Smith, J. (**1996**). "Consonant-vowel interdependencies in babbling and early words: Preliminary examination of a locus equation approach," J. Speech Lang. Hear. Res. **39**, 424–433.

Contribution of low-frequency acoustic information to Chinese speech recognition in cochlear implant simulations

Xin Luo[a] and Qian-Jie Fu
Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 25 October 2005; revised 21 July 2006; accepted 24 July 2006)

Chinese sentence recognition strongly relates to the reception of tonal information. For cochlear implant (CI) users with residual acoustic hearing, tonal information may be enhanced by restoring low-frequency acoustic cues in the nonimplanted ear. The present study investigated the contribution of low-frequency acoustic information to Chinese speech recognition in Mandarin-speaking normal-hearing subjects listening to acoustic simulations of bilaterally combined electric and acoustic hearing. Subjects listened to a 6-channel CI simulation in one ear and low-pass filtered speech in the other ear. Chinese tone, phoneme, and sentence recognition were measured in steady-state, speech-shaped noise, as a function of the cutoff frequency for low-pass filtered speech. Results showed that low-frequency acoustic information below 500 Hz contributed most strongly to tone recognition, while low-frequency acoustic information above 500 Hz contributed most strongly to phoneme recognition. For Chinese sentences, speech reception thresholds (SRTs) improved with increasing amounts of low-frequency acoustic information, and significantly improved when low-frequency acoustic information above 500 Hz was preserved. SRTs were not significantly affected by the degree of spectral overlap between the CI simulation and low-pass filtered speech. These results suggest that, for CI patients with residual acoustic hearing, preserving low-frequency acoustic information can improve Chinese speech recognition in noise. [DOI: 10.1121/1.2336990]

PACS number(s): 43.71.Es, 43.71.Hw [ARB] Pages: 2260–2266

I. INTRODUCTION

Contemporary cochlear implant (CI) devices provide good speech recognition in quiet to many patients. However, speech understanding in noise remains difficult for even the best-performing CI patients, because of the limited spectro-temporal fine-structure cues provided by the CI device (e.g., Fu *et al.*, 1998a; Stickney *et al.*, 2004; Qin and Oxenham, 2003). Combined electric and acoustic stimulation (EAS) in CI patients with residual acoustic hearing may preserve low-frequency spectro-temporal fine-structure cues that are important to speech recognition in noise, music perception, and tone reception for tonal languages (e.g., Mandarin Chinese). "Short-electrode" arrays have been designed to preserve residual low-frequency acoustic hearing in the implanted ear (e.g., Turner *et al.*, 2004). For CI patients implanted with standard electrode arrays, bilateral EAS can be achieved by combining a CI with a hearing aid (HA) in the nonimplanted ear (e.g., Tyler *et al.*, 2002; Ching *et al.*, 2004; Kong *et al.*, 2005). Despite potential incompatibilities between the two modes of hearing, EAS has been shown to benefit CI users' speech recognition in noise, melody recognition, and quality of hearing in daily life (e.g., Turner *et al.*, 2004; Ching *et al.*, 2004; Kong *et al.*, 2005). Bilateral EAS has also been shown to provide better sound localization, as interaural differences in time, phase, and intensity are partially preserved (e.g., Tyler *et al.*, 2002).

Improved reception of low-frequency acoustic cues with residual hearing may help CI users to track pitch changes and segregate speakers, thereby improving speech recognition in the presence of interfering speech. For example, Kong *et al.* (2005) found that, for CI patients who use an HA in the nonimplanted ear, while the HA alone did not provide speech understanding, combined use of the CI and HA significantly improved CI speech recognition in the presence of competing speech. Similarly, Turner *et al.* (2004) found that CI speech recognition in a background of two competing talkers was significantly improved by preserving residual low-frequency acoustic hearing in the implanted ear; however, unilateral EAS did not significantly improve CI speech recognition in the presence of steady-state, speech-shaped noise. These results suggest that preserving low-frequency fine-structure cues may help CI users to segregate and then group high-frequency envelope information, and that EAS can benefit CI speech recognition in the presence of competing speech.

The reception of tonal information is important for Chinese speech recognition by CI users (e.g., Fu *et al.*, 1998b; Luo and Fu, 2004a). However, most CI speech processing strategies do not provide sufficient spectral or temporal cues for tone recognition; the fundamental frequency ($F0$) and its harmonics are not explicitly encoded and temporal pitch cues in electric hearing are generally weak. With the latest CI technology, native Mandarin-speaking CI patients can correctly recognize only 50%–70% of Mandarin-Chinese tones, comparable to that of normal-hearing (NH) listeners listening to 4-channel acoustic CI simulations (e.g., Fu *et al.*, 2004;

[a]Author to whom correspondence should be addressed. Electronic mail: xluo@hei.org

Wei *et al.*, 2004). To improve pitch and tone perception by CI users, researchers have tried to increase the low-frequency spectral resolution (Geurts and Wouters, 2004), enhance the temporal envelope cues associated with pitch and tonal patterns (e.g., Geurts and Wouters, 2001; Luo and Fu, 2004b; Green *et al.*, 2005), or encode tonal information by varying the stimulation rate based on the *F*0 extracted from speech signal (Lan *et al.*, 2004). However, only limited improvements in tone recognition were observed with these approaches; most likely, even enhanced temporal envelope cues remain relatively weak in electric hearing. Because residual acoustic hearing provides more salient low-frequency pitch cues, EAS may greatly improve CI users' Chinese tone recognition and, in turn, Chinese sentence recognition (e.g., Fu *et al.*, 1998b; Luo and Fu, 2004a).

The present study investigated the contribution of low-frequency acoustic information to NH subjects' Chinese speech recognition while listening to an acoustic simulation of bilateral EAS. To precisely control the amount of low-frequency acoustic cues, low-pass filtered speech was presented to one ear while a 6-channel sine-wave CI simulation was presented to the other ear (for the sine-wave CI simulation, see Dorman *et al.*, 1997). In the first experiment, speech reception thresholds (SRTs) for Chinese sentences in steady-state, speech-shaped noise [defined as the signal-to-noise ratios (SNRs) needed to produce 50%-correct whole sentence recognition] were measured in six native Mandarin-speaking NH subjects. The input acoustic frequency range was fixed for the CI simulation (100–6000 Hz), while the cutoff frequency for low-pass filtered speech was varied from 0 to 1000 Hz. Chinese tone, vowel, and consonant recognition were measured at fixed SNRs for the same simulated EAS conditions.

It is unclear whether spectral overlap between the acoustic and electric hearing may enhance or impair speech recognition. In the first experiment of the present study, as the cutoff frequency for low-pass filtered speech was increased, the spectral overlap between the CI simulation and low-pass filtered speech was also increased. In a second experiment, SRTs for Chinese sentences in steady-state, speech-shaped noise were measured for conditions with no spectral overlap between the CI simulation and low-pass filtered speech. The lowest frequency limit of the input acoustic frequency range for the CI simulation was varied according to the cutoff frequency for low-pass filtered speech. Thus, for all conditions in the second experiment, there was no spectral overlap between the CI simulation and low-pass filtered speech; only the distribution of acoustic frequency information between the two simulated modes of hearing was varied.

II. EXPERIMENT 1: CHINESE SPEECH RECOGNITION WITH DIFFERENT CUTOFF FREQUENCIES FOR LOW-PASS FILTERED SPEECH

A. Methods

1. Subjects

Six young adult native Chinese-speaking listeners (three males and three females) participated in the present experiments. All subjects were normal hearing and had pure-tone thresholds better than 20 dB HL at octave frequencies from 125 to 8000 Hz in both ears. All subjects were paid for their participation.

2. Stimuli and speech processing

Sentences drawn from the Mandarin Hearing in Noise Test (HINT; Soli, 2003) were used to measure SRTs. One male speaker produced 240 Chinese sentences of easy to moderate difficulty, with ten keywords per sentence; the *F*0 of the speaker ranged from 75 to 180 Hz. Vowel and consonant stimuli were drawn from the Chinese Standard Database, recorded by Wang (1993). Two male and two female speakers each produced four tones for the six Mandarin Chinese single-vowel syllables (/a/, /o/, /e/, /i/, /u/, /ü/), resulting in a total of 96 vowel tokens, which were used for both the Chinese tone and vowel recognition tests. One male and one female speaker each produced four tones for /u/ in a consonant-/u/ context, for the 21 Mandarin Chinese initial consonants (/b/, /c/, /ch/, /d/, /f/, /g/, /h/, /j/, /k/, /l/, /m/, /n/, /p/, /q/, /r/, /s/, /sh/, /t/, /x/, /z/, /zh/), thereby creating a set of 152 lexically meaningful combinations, which were used for the Chinese consonant recognition test. The Mandarin HINT sentences were digitized using a 16-bit A/D converter at a 24-kHz sampling rate, while the Chinese vowel and consonant stimuli were sampled at a 16-kHz sampling rate, both without high-frequency pre-emphasis. Steady-state, speech-shaped noise was created by low-pass filtering white noise (−12 dB/octave above 1200 Hz).

Prior to speech processing, the speech signal was combined with noise at the target SNRs; the combined speech and noise stimuli were normalized to have the same long-term root-mean-square (rms) amplitude (65 dB). After normalization, stimuli were processed by an acoustic simulation of bilateral EAS. In one ear, a six-channel sine-wave vocoder was used to simulate CI speech processing. After pre-emphasis (first-order Butterworth high-pass filter at 1200 Hz), the input speech and noise signal was bandpass filtered into six channels (fourth-order Butterworth filters). The overall input acoustic frequency range was fixed (100–6000 Hz); the corner frequencies of the six analysis bands, calculated according to the Greenwood's (1990) formula, were 100, 274, 573, 1083, 1955, 3448, and 6000 Hz. The temporal envelope from each band was extracted by half-wave rectification and low-pass filtering (fourth-order Butterworth filter at 500 Hz), and was used to modulate a sine wave generated at center frequency of the analysis band. The amplitude-modulated sine waves from all six channels were summed to produce the CI simulation signal. In the other ear, the input speech and noise signal was low-pass filtered to simulate residual low-frequency acoustic hearing. The signal was processed by a 40th-order Butterworth low-pass filter; the cutoff frequency was varied from 0 to 1000 Hz in steps of 250 Hz to preserve different amounts of low-frequency acoustic information. Note that the 0-Hz low-pass filter cutoff frequency corresponded to CI-only processing, and served as the baseline condition. After processing by the low-pass filter and the CI simulation, the output speech and noise signal was normalized to have the same long-term rms amplitude as the input signal.

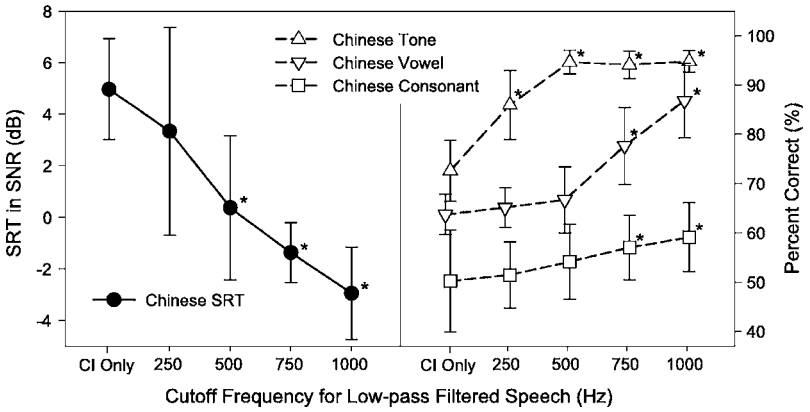


FIG. 1. Speech reception thresholds (SRTs) for Chinese sentences (left panel) and recognition scores for Chinese phonemes and tones (right panel) in steady-state, speech-shaped noise, as a function of the cutoff frequency for low-pass filtered speech in the bilateral EAS simulation. The symbols represent the mean values, and the error bars represent 1 standard deviation. The asterisks indicate EAS performance that was significantly better than that with the CI simulation alone (i.e., the "CI only" condition).

3. Procedures

Subjects were seated in a double-walled sound-treated booth and stimuli were presented via headphone (Sennheiser HDA200) at 65 dB SPL in each ear. For each recognition task, the test order of speech-processing conditions was randomized for each subject. No feedback was provided for any test session. To familiarize subjects with the speech-processing and test procedures, a practice session containing 20 sentences was provided prior to the sentence recognition test, and a preview and two practice sessions (in quiet and in noise) were provided prior to the phoneme and tone recognition tests. In the practice and preview sessions, stimuli processed by the six-channel CI simulation were presented to both ears of subjects, which were similar but not equal to the EAS simulation used in the present study.

For Chinese sentences in steady-state, speech-shaped noise, SRTs were measured as a function of the cutoff frequency for low-pass filtered speech in the bilateral EAS simulation; a one-down, one-up adaptive procedure was used (Van Tasell and Yanz, 1987). Mandarin HINT sentences were divided into 12 lists, with 20 sentences per list. For each condition, a list was randomly selected (without repetition) to measure the SRT. In each trial, a sentence was randomly selected (without repetition) from within the list and presented to the subject. Subjects were instructed to repeat the sentence as accurately as possible. The initial SNR was 0 dB; the SNR was adjusted according to subject response. If the whole sentence was repeated correctly, the SNR was reduced by 2 dB; if the whole sentence was not repeated correctly, the SNR was increased by 2 dB. The mean of final eight reversals in SNR was calculated as the SRT. If there were less than eight reversals in SNR, another run using another randomly selected sentence list and refined initial SNR values was conducted to obtain enough (at least eight) reversals in SNR.

After measuring SRTs, Chinese phoneme and tone recognition were measured at fixed SNRs, as a function of the cutoff frequency for low-pass filtered speech in the bilateral EAS simulation. To avoid any floor or ceiling effects on performance, Chinese tone and vowel recognition were measured at 1.65 dB SNR (the median SNR for the SRTs measured in the sentence recognition tests, across all subjects and conditions). Chinese consonant recognition was measured at 5 dB SNR, because subjects were more susceptible to noise for consonant recognition than for tone and vowel recognition. Closed-set identification tasks were used to measure Chinese tone (4-choices), vowel (6-choices), and consonant (21-choices) recognition. In each trial, a stimulus was randomly selected from the token list (without repetition) and presented to the subject; subjects responded by clicking on one of the response choices shown on screen. Responses were collected and scored in terms of percent correct.

B. Results

Figure 1 shows SRTs for Chinese sentences (left panel) and recognition scores for Chinese vowels, consonants, and tones (right panel) in steady-state, speech-shaped noise, as a function of the cutoff frequency for low-pass filtered speech in the bilateral EAS simulation. Simulated bilateral EAS significantly improved Chinese sentence recognition in steady-state, speech-shaped noise, relative to performance with the CI simulation alone. A one-way repeated-measures analysis of variance (ANOVA) showed that SRTs significantly improved when contra-lateral low-frequency acoustic cues were added to the CI simulation [$F(4,20)=16.84$, $p<0.001$]. *Post-hoc* Bonferroni *t*-tests showed that SRTs were significantly better with cutoff frequencies of 500 Hz or higher for low-pass filtered speech [$p<0.01$]. Simulated bilateral EAS also significantly improved Chinese tone recognition in steady-state, speech-shaped noise, relative to performance with the CI simulation alone [one-way repeated-measures ANOVA: $F(4,20)=36.16$, $p<0.001$]. *Post-hoc* Bonferroni *t*-tests showed that tone recognition was significantly better with cutoff frequencies of 250 Hz or higher for low-pass filtered speech [$p<0.001$]. Similarly, Chinese phoneme recognition in steady-state, speech-shaped noise was also significantly improved with bilateral EAS simulation, relative to performance with the CI simulation alone. One-way repeated-measures ANOVAs showed that adding contralateral low-frequency acoustic information to the CI simulation significantly improved vowel [$F(4,20)=49.52$, $p<0.001$] and consonant recognition [$F(4,20)=8.83$, $p<0.001$]. *Post-hoc* Bonferroni *t*-tests showed that both vowel and consonant recognition were significantly better with cutoff frequencies of 750 Hz or higher for low-pass filtered speech [$p<0.01$]. Even at a higher SNR (5 dB), subjects' consonant recognition scores were much lower than their tone and vowel recognition scores, which were obtained at 1.65 dB SNR. A possible explanation is that, although the overall SNR for

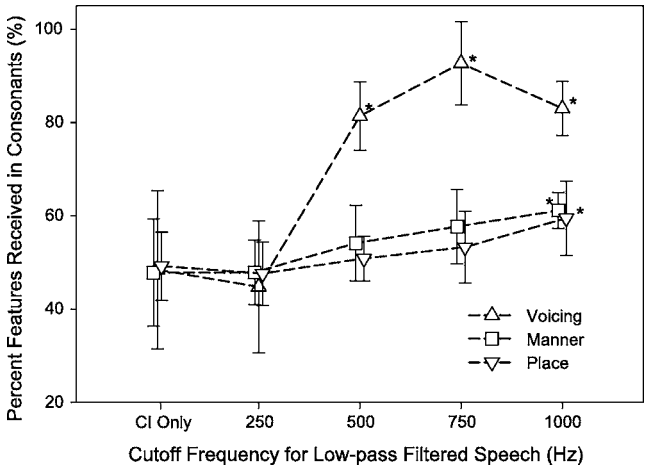


FIG. 2. Percent information received for Chinese consonant recognition in steady-state, speech-shaped noise, as a function of the cutoff frequency for low-pass filtered speech in the bilateral EAS simulation. The symbols represent the mean data for voicing, manner, and place of articulation. The error bars represent 1 standard deviation. The asterisks indicate EAS performance that was significantly better than that with the CI simulation alone (i.e., the "CI only" condition).

consonant-/u/ syllables was 5 dB, the local SNR for consonants was much lower than 1.65 dB, because in the consonant-/u/ syllables, the rms amplitude of consonants was around 10 dB lower than that of the vowel /u/.

Consonant recognition performance was further analyzed in terms of the amount of information received for production-based features of voicing, manner, and place of articulation (Miller and Nicely, 1955); the results are shown in Fig. 2. One-way repeated-measures ANOVAs showed that adding contralateral low-frequency acoustic information to the CI simulation significantly improved the reception of voicing $[F(4,20)=30.25,\ p<0.001]$, place $[F(4,20)=8.54,\ p<0.001]$, and manner cues $[F(4,20)=6.07,\ p=0.002]$. *Post-hoc* Bonferroni *t*-tests showed that the reception of voicing cues significantly improved with cutoff frequencies of 500 Hz or higher for low-pass filtered speech $[p<0.001]$. However, the reception of both place and manner cues significantly improved only when the cutoff frequency for low-pass filtered speech was 1000 Hz or higher $[p<0.01]$.

Note that, although the number of subjects (six) in the present study was small, the power of estimates in all the above statistical analyses was quite high (>0.91).

III. EXPERIMENT 2: CHINESE SENTENCE RECOGNITION WITH NO SPECTRAL OVERLAP BETWEEN THE CI SIMULATION AND LOW-PASS FILTERED SPEECH

A. Methods

1. Subjects

The same six subjects in the first experiment participated in the second experiment.

2. Stimuli, procedures, and speech processing

SRTs for Chinese sentences were measured in steady-state, speech-shaped noise using the same materials and procedures as in the first experiment; note that unique sentence lists were used. The speech processing in the bilateral EAS simulation was similar to that in the first experiment, except that the lowest frequency limit of the overall input acoustic frequency range for the CI simulation was varied according to the cutoff frequency for low-pass filtered speech. For example, when the cutoff frequency for low-pass filtered speech was 750 Hz, the overall input acoustic frequency range was 750–6000 Hz for the CI simulation; the corner frequencies of the six channels were re-calculated according to the Greenwood's (1990) formula for the CI simulation. Table I shows the corner frequencies for the six CI channels, for all experimental conditions.

B. Results

Figure 3 shows SRTs for Chinese sentences in steady-state, speech-shaped noise, as a function of the cutoff frequency for low-pass filtered speech in the bilateral EAS simulation. The solid line and filled circles show SRTs when there was spectral overlap between the CI simulation and low-pass filtered speech (same plot as in Fig. 1, left panel); the overall input acoustic frequency range for the CI simulation was fixed (100–6000 Hz), while the cutoff frequency for low-pass filtered speech was varied. The dashed line and open circles show SRTs when there was no spectral overlap between the CI simulation and low-pass filtered speech; the overall input acoustic frequency range for the CI simulation was varied according to the cutoff frequency for low-pass filtered speech. A two-way repeated-measures ANOVA showed that SRTs significantly improved with increasing cutoff frequency for low-pass filtered speech $[F(3,15)$

TABLE I. The corner frequencies of the six frequency channels used in the CI simulations. The lowest frequency limit of the overall input acoustic frequency range for the CI simulation was equal to the cutoff frequency for low-pass filtered speech.

	Cutoff frequency for low-pass filtered speech (Hz)				
	0	250	500	750	1000
	Corner frequencies of CI channels (Hz)				
Channel 1	100–274	250–479	500–794	750–1089	1000–1370
Channel 2	274–573	479–842	794–1223	1089–1556	1370–1860
Channel 3	573–1083	842–1414	1223–1846	1556–2200	1860–2508
Channel 4	1083–1955	1414–2317	1846–2754	2200–3088	2508–3365
Channel 5	1955–3448	2317–3745	2754–4076	3088–4312	3365–4499
Channel 6	3448–6000	3745–6000	4076–6000	4312–6000	4499–6000

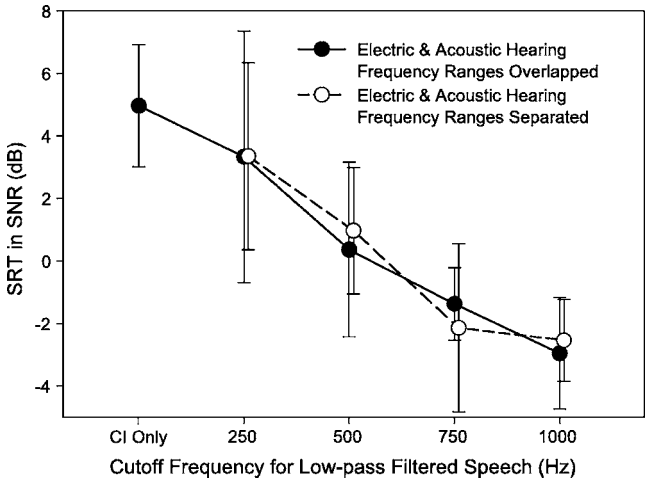


FIG. 3. Speech reception thresholds (SRTs) for Chinese sentences in steady-state, speech-shaped noise, as a function of the cutoff frequency for low-pass filtered speech in the bilateral EAS simulation. The filled symbols and solid line show the simulated EAS data with spectral overlap between the CI simulation and low-pass filtered speech (same plot as in Fig. 1, left panel). The open symbols and dashed line show the simulated EAS data with no spectral overlap between the CI simulation and low-pass filtered speech. The symbols represent the mean values, and the error bars represent 1 standard deviation.

$=20.46$, $p<0.001$]; however, there was no significant effect of spectral overlap between the CI simulation and low-pass filtered speech [$F(1,5)=0.05$, $p=0.84$], and there was no significant interaction between the cutoff frequencies for low-pass filtered speech and the degree of spectral overlap between the CI simulation and low-pass filtered speech [$F(3,15)=0.75$, $p=0.54$].

IV. GENERAL DISCUSSION

Although low-frequency acoustic hearing alone is not sufficient for speech recognition in noise (e.g., Kong *et al.*, 2005), in the present study, the addition of contralateral low-frequency acoustic information to a six-channel CI simulation significantly improved Chinese speech recognition in steady-state, speech-shaped noise. Perceptually, the spectral details contained in low-pass filtered speech may have helped listeners to better separate speech from noise, thereby improving speech recognition relative to the CI simulation alone, in agreement with the hypotheses of Turner *et al.* (2004) and Kong *et al.* (2005). Acoustically, speech cues available below 1000 Hz can well explain the pattern of the present results: improved reception of $F0$ information (mostly below 500 Hz) with low-pass filtered speech provided significantly better tone recognition, while improved reception of first formant ($F1$) information (from around 500 to 1000 Hz) with low-pass filtered speech provided significantly better vowel recognition. For tone recognition, performance asymptoted with cutoff frequencies above 500 Hz for low-pass filtered speech. For vowel recognition, no significant improvement was observed with cutoff frequencies below 500 Hz for low-pass filtered speech. For consonant recognition, the pattern for reception of voicing cues was similar to that of tone recognition, while the pattern for reception of place cues was similar to that of vowel recognition. These patterns of results point to the dichotomy in the distribution of speech cues available below 1000 Hz.

Chinese tone and phoneme recognition both contributed to Chinese sentence recognition in the present study. Chinese tone recognition significantly improved when the cutoff frequency for low-pass filtered speech was 250 Hz or higher, while Chinese phoneme recognition significantly improved only when the cutoff frequency was 750 Hz or higher. It is likely that the improved SRTs at the 500-Hz cutoff frequency for low-pass filtered speech were due to improved tone recognition, and that the further improvements in SRTs at the 750- and 1000-Hz cutoff frequencies were due to improved phoneme recognition. Despite the significantly improved Chinese tone recognition at the 250-Hz cutoff frequency for low-pass filtered speech, SRTs did not significantly improve until the cutoff frequency was 500 Hz or higher. This discrepancy may have been due to the variability in sentence recognition performance among the relatively small number of subjects. Correlation analyses were performed to confirm the contribution of Chinese tone and phoneme recognition to Chinese sentence recognition. At each cutoff frequency for low-pass filtered speech, the geometric mean of Chinese tone, vowel, and consonant recognition scores was compared to the mean SRT; the correlation was highly significant [$r^2=0.98$, $p<0.001$]. When the geometric mean of only Chinese vowel and consonant recognition scores was compared to the mean SRT, the significance of the correlation was reduced [$r^2=0.89$, $p=0.02$]. These results are in agreement with those previous studies (Fu *et al.*, 1998b; Luo and Fu, 2004a), which show the importance of tonal information to Chinese speech recognition by CI users.

The present simulation results for Chinese speech recognition extend the observations of Kong *et al.* (2005), who showed that for patients using an HA in conjunction with the CI, bilateral EAS provided a significant advantage over electric or acoustic hearing alone for recognition of English sentences in the presence of a competing talker. Despite similar findings in the two studies, one should be cautious when relating the present simulation results to real EAS performance. First, simply low-pass filtering speech in NH listeners cannot fully simulate the residual acoustic hearing available to some CI users. Severe-to-profound hearing loss is likely to affect both ears in CI users, and can have many effects beyond the reduction of the functional bandwidth of hearing. For example, pitch discrimination is likely to be impaired (Faulkner *et al.*, 1992), and frequency resolution is also likely to be poor (Faulkner *et al.*, 1990; Moore, 1996), resulting in poorer speech perception in noise. Second, although acoustic CI simulations have been successful in estimating the limits of CI patient performance (e.g., Shannon *et al.*, 1995; Dorman *et al.*, 1997), sine-wave vocoders may introduce speech information that is not available to real CI users. For example, using a 500-Hz low-pass envelope filter for temporal envelope extraction will produce spectral sidebands around the sine-wave carriers that may contain $F0$ information. Because real CI patients do not have access to these spectral sideband pitch cues, the addition of low-

frequency acoustic information would likely provide greater benefits for Chinese speech recognition than shown in the present EAS simulations.

The present results are somewhat different from those of Turner *et al.* (2004), who found that English spondee word recognition in steady-state, speech-shaped noise was not significantly improved by adding low-frequency (<500 Hz) acoustic information to the CI simulations. Two factors may have contributed to this difference in results. First, the two studies used different languages and materials to test speech recognition (English spondee recognition for Turner *et al.*, and Chinese sentence recognition in the present study). Chinese speech recognition depends more strongly on pitch and tonal information than does English speech recognition. Thus, while additional *F*0 cues (via low-pass filtered speech) might significantly improve Chinese speech recognition in steady-state, speech-shaped noise (as in the present study), they would have little effect on English speech recognition (as in Turner *et al.*). Second, different speech processing conditions were tested in the two studies. Turner *et al.* used 16-channel CI simulations, which produced much lower SRTs (~−15 dB) than are typically observed with CI patients (~2 dB) or observed with the six-channel CI simulations used in the present study (~5 dB); the data from Turner *et al.* may have overestimated the spectral resolution of real CI users and, in turn, underestimated the contribution of low-frequency acoustic information to speech recognition in steady-state, speech-shaped noise.

Although Dorman *et al.* (2005) found that speech performance for unilateral EAS was best when the frequency gap between the two simulated modes of hearing was minimized, the degree of spectral overlap between the simulated electric and acoustic hearing did not significantly affect speech performance for bilateral EAS in the present study. Within the overlapping frequency ranges between the CI simulation and low-pass filtered speech, the reduced spectro-temporal resolution in the CI simulation speech was most likely insufficient to be susceptible to interference by contralateral, high-resolution acoustic information. It is also possible that listeners may attend to the better signal and ignore the poorer representation when bilateral speech cues overlap in frequency. These results suggest that, for bilateral EAS, the input frequency ranges for the CI and low-frequency acoustic hearing do not have to be spectrally separated to provide optimal performance.

V. CONCLUSIONS

In an acoustic simulation of bilateral EAS, Chinese speech recognition in steady-state, speech-shaped noise significantly improved as access to low-frequency acoustic information was increased. Differential effects for the amount of low-frequency acoustic information on Chinese tone, vowel, consonant, and sentence recognition were found.

(1) Chinese tone recognition significantly improved when the cutoff frequency for low-pass filtered speech was 250 Hz and asymptoted when the cutoff frequency was 500 Hz or higher.
(2) Chinese phoneme recognition did not significantly improve until the cutoff frequency for low-pass filtered speech was 750 Hz or higher.
(3) For Chinese sentences, SRTs improved with increasing amounts of low-frequency acoustic information, and significantly improved when the cutoff frequency for low-pass filtered speech was 500 Hz or higher.
(4) Sentence recognition with bilateral EAS was not significantly affected by the degree of spectral overlap between the CI simulation and low-pass filtered speech.

These results suggest that preserving low-frequency acoustic information may greatly enhance Chinese speech recognition by CI patients.

ACKNOWLEDGMENTS

We are grateful to all subjects for their participation in these experiments. We thank John J. Galvin III for editorial assistance. We would also like to thank Dr. Fan-Gang Zeng, Dr. Andrew Faulkner, and an anonymous reviewer for their constructive comments on an earlier version of this paper. Research was supported in part by NIH (DC-004993).

Ching, T. Y. C., Incerti, P., and Hill, M. (**2004**). "Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears," Ear Hear. **25**, 9–21.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**, 2403–2411.

Dorman, M. F., Spahr, A. J., Loizou, P. C., Dana, C. J., and Schmidt, J. S. (**2005**). "Acoustic simulations of combined electric and acoustic hearing (EAS)," Ear Hear. **26**, 371–380.

Faulkner, A., Rosen, S., and Moore, B. C. J. (**1990**). "Residual frequency selectivity in the profoundly hearing-impaired listener," Br. J. Audiol. **24**, 381–392.

Faulkner, A., Ball, V., Rosen, S., Moore, B. C. J., and Fourcin, A. J. (**1992**). "Speech pattern hearing aids for the profoundly hearing-impaired: Speech perception and auditory abilities," J. Acoust. Soc. Am. **91**, 2136–2155.

Fu, Q.-J., Hsu, C.-J., and Horng, M.-J. (**2004**). "Effects of speech processing strategy on Chinese tone recognition by Nucleus-24 cochlear implant patients," Ear Hear. **25**, 501–508.

Fu, Q.-J., Shannon, R. V., and Wang, X.-S. (**1998a**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.

Fu, Q.-J., Zeng, F.-G., Shannon, R. V., and Soli, S. D. (**1998b**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **104**, 505–510.

Geurts, L., and Wouters, J. (**2001**). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," J. Acoust. Soc. Am. **109**, 713–726.

Geurts, L., and Wouters, J. (**2004**). "Better place-coding of the fundamental frequency in cochlear implants," J. Acoust. Soc. Am. **115**, 844–852.

Green, T., Faulkner, A., Rosen, S., and Macherey, O. (**2005**). "Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification," J. Acoust. Soc. Am. **118**, 375–385.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Kong, Y.-Y., Stickney, G. S., and Zeng, F.-G. (**2005**). "Speech and melody recognition in binaurally combined acoustic and electric hearing," J. Acoust. Soc. Am. **117**, 1351–1361.

Lan, N., Nie, K.-B., Gao, S.-K., and Zeng, F.-G. (**2004**). "A novel speech-processing strategy incorporating tonal information for cochlear implants," IEEE Trans. Biomed. Eng. **51**, 752–760.

Luo, X., and Fu, Q.-J. (**2004a**). "Importance of pitch and periodicity to Chinese-speaking cochlear implant patients," in Proceedings of IEEE Int. Conf. Acoust., Speech, Signal Processing, 2004, Vol. 4, pp. 1–4.

Luo, X., and Fu, Q.-J. (**2004b**). "Enhancing Chinese tone recognition by

manipulating amplitude envelope: Implications for cochlear implants," J. Acoust. Soc. Am. **116**, 3659–3667.
Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.
Moore, B. C. J. (**1996**). "Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids," Ear Hear. **17**, 133–161.
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**, pp. 446–454.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Soli, S. D. (2003). "Hearing in Noise Test for Mandarin Chinese," House Ear Institute, Los Angeles, CA.
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.
Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (**2004**). "Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing," J. Acoust. Soc. Am. **115**, 1729–1735.
Tyler, R. S., Parkinson, A. J., Wilson, B. S., Witt, S., Preece, J. P., and Noble, W. (**2002**). "Patients utilizing a hearing aid and a cochlear implant: Speech perception and localization," Ear Hear. **23**, 98–105.
Van Tasell, D. J., and Yanz, J. L. (**1987**). "Speech recognition threshold in noise: Effects of hearing loss, frequency response, and speech materials," J. Speech Hear. Res. **30**, 377–386.
Wang, R.-H. (**1993**). "The standard Chinese database," University of Science and Technology of China, internal materials.
Wei, C.-G., Cao, K.-L., and Zeng, F.-G. (**2004**). "Mandarin tone recognition in cochlear-implant subjects," Hear. Res. **197**, 87–95.

Formant transitions in fricative identification: The role of native fricative inventory

Anita Wagner,[a] Mirjam Ernestus, and Anne Cutler
Max Planck Institute for Psycholinguistics, Nijmegen, 6500 AH, The Netherlands

(Received 1 August 2005; revised 28 June 2006; accepted 14 July 2006)

The distribution of energy across the noise spectrum provides the primary cues for the identification of a fricative. Formant transitions have been reported to play a role in identification of some fricatives, but the combined results so far are conflicting. We report five experiments testing the hypothesis that listeners differ in their use of formant transitions as a function of the presence of spectrally similar fricatives in their native language. Dutch, English, German, Polish, and Spanish native listeners performed phoneme monitoring experiments with pseudowords containing either coherent or misleading formant transitions for the fricatives /s/ and /f/. Listeners of German and Dutch, both languages without spectrally similar fricatives, were not affected by the misleading formant transitions. Listeners of the remaining languages were misled by incorrect formant transitions. In an untimed labeling experiment both Dutch and Spanish listeners provided goodness ratings that revealed sensitivity to the acoustic manipulation. We conclude that all listeners may be sensitive to mismatching information at a low auditory level, but that they do not necessarily take full advantage of all available systematic acoustic variation when identifying phonemes. Formant transitions may be most useful for listeners of languages with spectrally similar fricatives. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2335422]

PACS number(s): 43.71.Es, 43.71.Hw [ARB] Pages: 2267–2277

I. INTRODUCTION

Do formant transitions contribute to listeners' identification of fricatives? These dynamic cues are crucial for the identification of stops, but despite decades of research (Harris, 1958; Heinz and Stevens, 1961; LaRiviere, Winitz, and Herriman, 1975; Jongman, 1989; Jongman, Wayland, and Wong, 2000), no clear answer has emerged for fricatives. Salient static cues are present in the fricative spectrum, and may suffice for phoneme identification. We report a study which contributes to this discussion by testing the hypothesis that the contribution of formant transitions is language specific and depends on the presence of spectrally similar fricatives in the listener's native phoneme inventory.

Fricatives are produced with a narrow constriction in the oral cavity. The turbulence of the airflow passing this constriction generates the characteristic sound of frication. The exact location of the narrow passage and the size and form of the cavity in front of the constriction define the acoustic characteristics of the fricative (Stevens, 1998). These energy peaks and minima in a fricative's spectrum serve listeners as primary cues for fricative identification (Stevens, 1998). The salience of those spectral poles, however, differs among fricatives, and previous research (e.g., Harris, 1958) suggests that listeners need additional cues to identify some but not all fricatives. Whereas sibilants have very pronounced spectral peaks and are identified primarily on the basis of these poles, dental and labiodental fricatives have a more diffuse energy spectrum and may require additional cues for accurate identification. Two contextual sources of such cues have been found (Whalen, 1981): formant transitions, which may be perceptually integrated with cues from the fricative spectrum; and the quality of the surrounding vowels, including the resulting slight modifications of the fricative spectrum itself.

It is unclear, however, whether formant transitions indeed contribute to the identification of fricatives, since the results from previous research are conflicting. Harris (1958) studied the identification of English fricatives in different vocalic contexts. In a fricative categorisation experiment, she presented American students with natural tokens of consonant (C) vowel (V)-syllables containing the fricatives /f v θ ð s z ʃ ʒ/ combined with the vowels /a ɪ u e/. These syllables were spliced such that every fricative was combined with every vowel as produced in the context of each of the fricatives. Thus, the formant transitions in some tokens contained misleading information with respect to the identity of the fricative. Participants accurately categorized /s/ and /ʃ/ in the combination of just the frication part from the sibilant with each of the vowels, independently of the fricative context from which these vowels were extracted. In contrast, stimuli with frication from /f/ or /θ/ were often confused with each other. In fact, the /f/ tended to be categorized as /f/ only when combined with a vowel originally produced after /f/, but as /θ/ when followed by any other vowel. Apparently, the English listeners recognized the sibilants/s/ and /ʃ/ by their frication part alone, while the dental fricatives /f/ and /θ/ were accurately categorized only when followed by correct formant transitions.

Similar results were obtained by Heinz and Stevens (1961) with synthesized English voiceless fricatives. American listeners identified /s ʃ f θ/ in isolation, and achieved satisfactory identification rates for /s ʃ/, but they could not

[a]Electronic mail: anita.wagner@mpi.nl

distinguish between /f/ and /θ/. The identification scores improved when the fricatives were combined with the synthetic vowel /a/, including approximated transition movements; especially the distinction between /f/ and /θ/ was more reliably perceived.

More recent studies, however, failed to replicate these results. Jongman (1989) asked English listeners to identify fricatives by listening either to portions of the frication alone, or to the whole frication, or to complete syllables (all eight English fricatives except /h/; produced by an American speaker with the vowels /a i u/). A portion of the frication longer than 40 ms appeared to be sufficient for listeners to identify all fricatives accurately, including the oft-confused fricatives /f/ and /θ/. No improvement of fricative identification resulted from inclusion of the vowel. Jongman *et al.* (1998) further supported this conclusion in a production study. They analyzed the variances of locus equations (Fruchter and Sussman, 1997) of English fricatives followed by the vowels /i e æ a o u/ as produced by 20 speakers. On this parameter /f v/ differed significantly from /s z ʃ ʒ θ ð/, but the three places of articulation represented in the latter set did not differ. Jongman *et al.* (1998) concluded that locus equations cannot sufficiently cue fricative place of articulation.

LaRiviere, Winitz, and Herriman (1975), too, queried the role of formant transitions in fricative identification. They compared identification of syllables made up of /f θ s ʃ/ and /a i u/, with the identification of the same syllables with deleted formant transitions. Listeners could reliably identify all fricatives in transitionless syllables, and the authors thus concluded that formant transitions do not necessarily contribute to fricative identification. LaRiviere *et al.* also found that /θ/ was the most difficult fricative to identify. They explain possible, but not necessary, perceptual benefit from the following vowel as arising from the information that it carries about the speaker's vocal tract, which contributes to the process of speaker normalisation.

Klaassen-Don (1983) also found no evidence that formant transitions contribute to fricative identification. In a gating experiment with Dutch fricatives, she presented naturally produced CV and VC strings including the fricatives /f v s z ʃ x/ and the vowels /a i u/. The syllables were produced in isolation or were excerpted from running speech. Formant transitions proved to be valuable cues for liquids and stops, but their contribution in fricative identification was negligible. Klaassen-Don reached the conclusion that "vowel transitions do not contain perceptually relevant information about adjacent fricatives in Dutch" (Klaassen-Don, 1983, p. 79).

Finally, in a series of production and perception experiments, Borzone de Manrique and Massone (1981) investigated the identification of Argentinian Spanish fricatives by native listeners. The perceptual power of the most prominent noise frequency bands was tested by bandpass filtering the fricatives /s f ʃ x/. The identifications showed that /s/ is the most robust fricative, whereas /f/ requires a wide noise band to be accurately identified. In further experiments, the authors concentrated on the role of the vocalic environment for fricative identification by Argentinian listeners. Their stimuli consisted of frication and vocalic parts spliced out of naturally produced CV syllables and of transitionless CV syllables, which they constructed by combining natural fricatives and vowels produced in isolation. For Argentinian listeners the frication part alone was sufficient to identify all fricatives, with the exception of the velars /x ɣ/. The absence of transitions in the vowel biased the listeners to the fricative that is realized with the least transition movements into the following vowel. For instance, the formant transitions following /f/ are shorter before /u/ than before /i/, and the authors observed a higher number of /f/ categorizations for syllables consisting of frication and /u/ rather than frication and /i/.

In short, the literature shows that formant transitions proved to be useful cues in some experiments but of little use in others. Importantly, the experiments involved listeners of different native languages. We hypothesize that the solution to the conflicting results is that listeners' attention to formant transitions for fricative identification is language specific and modulated by the presence of perceptually similar fricatives in the native phoneme inventory. Languages differ widely in how many fricatives they include, and how similar these fricatives are. More fricatives in a given perceptual space may reduce the distinctiveness of individual fricatives. To maximize the distinctiveness of fricatives in denser perceptual spaces, listeners may learn to integrate additional cues to attain accurate percepts of these fricatives.

If listeners of different native languages indeed differ in the use they make of transitional cues, we can further ask whether listeners who do exploit transitional information do so for all native fricatives, or only for contrasts which are perceptually similar. Listeners' language experience may tune the perceptual system to select relevant cues efficiently for each fricative: If more salient cues suffice to distinguish a given phoneme contrast, native listeners may make no use of the information in formant transitions. Thus our second hypothesis is that attention to formant transitions can be restricted to those fricatives that are difficult to distinguish spectrally. The fricative pair /f θ/ seems, on the evidence cited above, to be difficult to distinguish for English listeners. For Argentinian listeners, without /f θ/ in their native phoneme inventory, a different pair of fricatives appears to be potentially confusable: /x ɣ/. We assume that listeners will learn the most efficient way to identify all native fricatives, and that it might not be beneficial for them to use the cues in formant transitions for fricatives that can be identified accurately on the basis of the fricative spectrum alone.

In the present study, listeners of different languages heard pseudowords containing either coherent or misleading information in the formant transitions surrounding fricatives. In four experiments participants performed phoneme monitoring, a task that has been used to investigate a wide range of psycholinguistic issues (see Connine and Titone, 1996, for a review). In phoneme monitoring, listeners hear spoken input, e.g., lists of words, nonwords, or syllables, and respond as soon as they detect a prespecified target phoneme. Phoneme monitoring is especially promising as a paradigm for testing our hypothesis because it has been shown to be sensitive to formant transitions: Detection of a phoneme is more

TABLE I. The fricative inventories of the languages studied according to the place of articulation.

	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Alveolopalatal	Velar	Glottal
Dutch	f v		s z	(ʃ)			x	h
German	f v		s z	ʃ ʒ			x	h
Spanish	f	θ	s				x	
English	f v	θ ð	s z	ʃ ʒ				h
Polish[1]	f v		s z	ʃʲ ʒʲ	ʂ ʐ	ɕ ʑ	x	

[1]Polish postalveolar fricatives /ʃ ʒ/ are traditionally described as laminal alveolar (Jassem, 2003), and the alveolopalatal /ɕ ʑ/ are considered as their palatalized counterparts. Hamann (2003) argues that Polish postalveolar fricatives should be considered as retroflex; in addition Zygis and Hamann (2003) claim that the alveolopalatal and the palatalized postalveolar fricatives in Polish should be considered two separate sounds, as they are distinguished by native and non-native listeners. This view is adopted in our description of the Polish fricative repertoire.

difficult when its context is cross spliced and thus bears mismatching coarticulatory information (Martin and Bunnell, 1981; McQueen, Norris, and Cutler, 1999). Moreover, the task is sensitive to cross-language differences in speech processing. Otake *et al.* (1996) and Weber (2001) showed effects of language-specific phonotactic constraints in phoneme monitoring for nasals and fricatives, respectively. Similarly, with the same task Costa, Cutler, and Sebastián-Gallés (1998) showed that processing of acoustic variation is affected by native phoneme inventory constitution.

If listeners depend on formant transitions in fricative identification, then mismatching formant transitions should increase errors and slow reaction times in phoneme monitoring. In contrast, listeners whose fricative identification is governed mostly by the primary static cues in the noise spectrum should be less affected by misleading formant transitions, either in reaction speed or error rate.

We tested five languages: German and Dutch, which both have only spectrally distinct fricatives, and (Castilian) Spanish, English, and Polish, which all have pairs of fricatives in which the distribution of noise peaks across the spectrum is very similar, so that the members of the pair are perceptually less distinctive. Spanish and English contrast with Polish with respect to which spectrally similar fricatives appear in the phoneme inventory. Table I sketches the fricative inventories of the five languages.

Experiment I contrasted Spanish with Dutch and German. Spanish, as we saw, has the confusable pair /f θ/. The spectra of the labiodental and dental fricatives are relatively flat; the energy is distributed in each case across frequencies from circa 2–10 kHz with no defined spectral peaks (Jongman, Wayland, and Wang, 2000). We therefore expected Spanish listeners to pay more attention to formant transitions than Dutch or German listeners, whose languages contain no spectrally similar fricatives. The fricatives in the experiment were the labiodental /f/ and the alveolar /s/. Since of these only /f/ is spectrally confusable with another fricative in Spanish, we further expected Spanish listeners to be particularly affected by mismatching formant transitions for /f/.

II. EXPERIMENT I

A. Method

1. Materials

Three- and four-syllable pseudowords made up of the phonemes /p b t d k f s a i u e/ (e.g., *tikusa* and *dokupafi*) were recorded by a native speaker of Dutch. Note that no fricatives other than /f/ or /s/ appeared in the stimuli. The fricative identification was part of a larger phoneme monitoring experiment with various phonemes as targets. Only the results for the fricative targets will be reported here.

We created 12 pseudowords with the target /f/ and 12 pseudowords with the target /s/. The fricatives were preceded and followed by /a i u/. The target appeared always in the last syllable; stress was always on the first syllable. In addition, for every target fricative 12 filler items were created with the fricative in the penultimate syllable, and 12 filler items without the fricative.

The stimuli were recorded in a sound-attenuated room directly to computer and down-sampled to 22.05 kHz (16 bit resolution). With Praat software cross-spliced and identity-spliced versions of the pseudowords were created. Identity-spliced fricatives were replaced by the same fricative taken from another token of the same pseudoword (e.g., /s/ in *tikusa* by /s/ of another *tikusa*). Cross-spliced fricatives were replaced by the other fricative produced in the same context (e.g., /s/ in *tikusa* by /f/ from *tikufa*). Segmentation points for the fricatives were defined visually, on the basis of oscillograms and sonagrams. The end of harmonic structure of the preceding vowel and the beginning of harmonic structure in the fading noise of the fricative were defined as the splicing points. At zero-crossing points the coherent stochastic noise parts of the fricative were excised. The spliced stimuli were examined auditorily to ensure that no audible discontinuities had resulted from the manipulation.

2. Procedure

Participants sat in a sound-attenuated room in front of a computer screen, and heard both cross-spliced and identity-spliced stimuli over headphones. Each pseudoword appeared only once in a session. Trials were blocked by target phoneme, with the order of blocks counterbalanced across participants. Participants were informed orally about the possible targets in advance; during the experiment a letter on the computer screen designated the current target. Participants were instructed to press a key immediately upon detecting in the nonword the sound represented by the displayed letter. Every target block of stimuli was followed by a break, the duration of which was controlled by the participants. From item onset, listeners had 2000 ms to respond. Failures to respond, and responses over 2000 ms, were defined as timeout

TABLE II. Average percentages of timeouts and mean RTs in milliseconds (ms) for the three languages and the two fricatives in both splicing conditions in experiment I. The absolute numbers of timeouts and the total numbers of trials are given in parentheses.

	Fricative	Dutch	German	Spanish
Mean percentage of timeouts	/s/ identity spliced	4.3% (4/93)	1.8% (2/115)	2.7% (4/170)
	/s/ cross spliced	4.3% (4/93)	3.5% (4/115)	2.7% (3/167)
	/f/ identity spliced	2.0% (2/93)	1.8% (2/115)	4.6% (6/169)
	/f/ cross spliced	2.1% (2/93)	1.0% (1/115)	45.2% (55/145)
Mean RT	/s/ identity spliced	488.22	440.82	544.04
	/s/ cross spliced	512.23	428.8	562.52
	/f/ identity spliced	531.27	442.22	618.6
	/f/ cross spliced	540.50	475.67	666.8

errors. The experiment was self-paced: The next stimulus was presented 1000 ms after the participant's response or timeout, and it was preceded by a beep tone.

3. Participants

Eighteen Dutch regular students, and 21 German and 23 Spanish exchange students from the Radboud University Nijmegen took part in this experiment. They were paid for their participation. None reported any speech or hearing disorders.

B. Results

Two items, one for each fricative target, were missed by more than 40% of the participants and therefore excluded from the analysis. The average timeouts (mean percentages of targets not correctly detected within 2000 ms) and reaction times (RTs) for the remaining items for the three languages, the two fricatives and the two splicing conditions are shown in Table II.

1. Timeouts

We analyzed the timeouts by means of a loglinear analysis with the number of timeouts and nontimeouts for each stimulus as the dependent variable and language (Dutch, German, and Spanish), splicing (identity splicing and cross splicing), and fricative (/s/ and /f/) as independent variables. All main effects were significant (language: $F(2,129)=30.22$, $p<0.001$; splicing: $F(1,127)=33.47$, $p<0.001$; and fricative: $F(1,128)=29.16$, $p<0.001$). These main effects were modulated by an interaction between language and fricative [$F(2,125)=15.48, p<0.001$]. Importantly, we also observed the predicted interactions between language and splicing [$F(2,123)=6.63, p<0.001$], and between language, fricative, and splicing [$F(2,120)=4.29, p<0.015$]. Splicing did not affect the number of timeout errors for the Dutch and German listeners, but the Spanish listeners were severely disturbed by misleading formant transitions [$F(1,41)=48.42, p<0.001$]. The effect of splicing for Spanish was restricted to /f/ [interaction between splicing and fricative for Spanish $F(1,40)=11.32, p<0.001$].

2. RTs

Latencies were measured from onset of the target fricative, defined as onset of the disharmonic structure in the stimulus wave form. Latencies below 150 ms were excluded from analysis (0.3% of the data). Analyses of variance were conducted for participants ($F1$) and items ($F2$), with language, splicing, and fricative as independent variables.

The main effects of language and fricative were significant in both analyses [language: $F1(2,58)=7.14$, $p<0.01$, $F2(2,105)=55.42$, $p<0.001$; fricative: $F1(1,174)=31.49$, $p<0.001$, $F2(1,21)=8.53, p<0.001$], while splicing was significant only in the analysis by participants [$F1(1,174)=5.29, p<0.05$]. The interaction of language with fricative was significant in the analysis by participants [$F1(2,174)=31.60, p<0.001$]. More importantly, in the analysis by participants we also observed the interaction between language and splicing [$F1(2,174)=5.12, p<0.01$]. This interaction failed to reach significance in the analysis by items.

C. Summary and discussion

We found language-specific patterns in the use of formant transitions in fricative identification. Only Spanish listeners were affected by misleading formant transitions. Apparently, they were attending to cues that were neglected by the Dutch and German listeners. Recall that the German and Dutch phoneme repertoires do not contain spectrally similar fricatives, while Spanish includes the two spectrally similar fricatives /f/ and /θ/. Even though /θ/ was not in the stimulus set, Spanish listeners paid attention to the formant transitions for /f/. They did not do so for /s/, which is spectrally distinct from the other fricatives in Spanish. These data support the hypothesis that listeners make use of formant transitions especially for fricatives that are spectrally similar to other fricatives in their native phoneme repertoire. Further, the results indicate that listeners do not necessarily take advantage of all acoustic information transmitted in the signal. The German and Dutch listeners showed no effects of the mismatching information that led Spanish listeners into errors.

However, Dutch participants had the advantage of listening to native phoneme realizations, while the Spanish listened to a foreign realization. The fact that German listeners showed the same pattern of results as the Dutch listeners may reflect a closer resemblance of German phonemes to Dutch than to Spanish phonemes. An alternative explanation for the cross-language differences might therefore be that listeners pay attention to more or to different cues when listening to a foreign pronunciation.

TABLE III. Average percentages of timeouts and mean RTs in ms for the three languages and the two fricatives in both splicing conditions in experiment II. The absolute numbers of timeouts and the total numbers of trials are given in parentheses.

	Fricative	Dutch	German	Spanish
Mean percentage of timeouts	/s/ identity spliced	0% (0/180)	2.2% (5 /180)	1.1% (2/172)
	/s/ cross spliced	0% (0/180)	2.7% (4/180)	0% (0/173)
	/f/ identity spliced	1.1% (2/180)	1.1% (0/178)	2.3% (4/172)
	/f/ cross spliced	1.6% (3/180)	2.2% (4/180)	27.4% (47/173)
Mean RT	/s/ identity spliced	461.54	474.34	474.93
	/s/ cross spliced	463.05	490.81	473.89
	/f/ identity spliced	550.67	569.06	601.93
	/f/ cross spliced	552.43	568.67	661.33

Experiment II was designed to test this second explanation. Experiments II and I differed principally in the native language of the speaker who recorded the stimuli: Dutch in experiment I, Spanish in experiment II. In experiment II, the Spanish listeners were thus presented with a familiar pronunciation, while the Dutch and German listeners were confronted with an unfamiliar realization of phonemes.

III. EXPERIMENT II

A. Method

1. Materials and procedure

The stimulus set from experiment I was now recorded by a native speaker of Spanish. In addition, 30 new fillers were created for each target with the target in the penultimate syllable or with the target missing. These fillers did not contain the phonemes /b/ and /d/, since Spanish phonotactics allows voiced bilabial and alveolar stops only in certain positions, and these consonants would therefore lead to a marked pronunciation by the Spanish speaker. The procedure was as in experiment I.

2. Participants

Twenty-four Dutch regular, and 24 German and 24 Spanish exchange students from the Radboud University Nijmegen were paid to take part in this experiment. None had participated in experiment I, and none had any known speech or hearing disorders.

B. Results

We defined and analyzed timeout errors and reaction latencies in the same way as in experiment I. No data point was below 150 ms, the common phoneme monitoring cutoff value (see, e.g., McQueen *et al.*, 1999), and therefore no reaction time data were excluded from the analysis. Table III shows the results of this experiment.

1. Timeouts

All main effects were significant [language: $F(2,177)=28.32$, $p<0.001$; splicing: $F(1,176)=28.49$, $p<0.001$; fricative: $F(1,175)=42.50$, $p<0.001$]. These main effects were modulated by interactions of language and splicing [$F(2,173)=5.39, p<0.001$], language and fricative [$F(2,171)=13.68, p<0.001$], and splicing and fricative [$F(1,170)=6.3, p<0.005$]. The interaction between language, splicing, and fricative narrowly missed significance [$F(2,168)=2.4, p<0.1$]. Splicing affected the number of timeout errors for the Spanish listeners [$F(1,58)=38.4, p<0.001$] only, and especially for the detection of /f/ [interaction of splicing and fricative for Spanish $F(1,56)=10.41, p<0.001$]. These results replicate those of experiment I.

2. RTs

The main effects of language, splicing, and fricative were significant in both the participant and the item analyses [language: $F1(2,58)=7.2$, $p<0.01$, $F2(2,112)=11.56$, $p<0.001$; splicing: $F1(1,207)=5.79$, $p<0.05$, $F2(1,140)=4.94$, $p<0.05$; fricative: $F1(1,207)=42.45, p<0.001$, $F2(1,28)=25.45, p<0.001$]. Also the interaction of language and fricative was significant in both analyses [$F1(2,207)=9.27, p<0.001, F2(2,140)=8.63, p<0.001$].

C. Summary and discussion

Experiment II further supports the hypothesis that Spanish listeners are affected by misleading formant transitions for fricative identification, while German and Dutch listeners are not. We ascribe these language differences to the different structures in the phoneme inventories of these languages, more precisely to the presence or absence of spectrally similar fricatives. Moreover, the finding that the Spanish only appeared to attend to formant transitions surrounding the labiodental fricative /f/ supports the hypothesis that the use of these cues is restricted to spectrally similar fricatives.

We obtained the same results for stimuli produced by a Dutch speaker (experiment I) and by a Spanish speaker (experiment II). Thus, experiments I and II together suggest that the native language of the speaker, or, in other words, the listeners' familiarity with the presented realization of the phonemes, does not alter the role of formant transitions in listeners' identification. We conclude that listeners also apply the native strategy when listening to a foreign pronunciation.

To explore further whether the presence of acoustically similar fricatives in a language's phoneme repertoire results in attention to formant transitions, we performed a third experiment with English native listeners. Since English is a Germanic language, it is in many respects more like Dutch and German than like Spanish. However, English has, like

TABLE IV. Average percentages of timeouts and mean RTs in ms for the English listeners and the two fricatives in both splicing conditions in experiment III. The absolute numbers of timeouts and the total numbers of trials are given in parentheses.

Fricative	/s/ identity spliced	/s/ cross spliced	/f/ identity spliced	/f/ cross spliced
Mean percentage of timeouts	6.2 (11/177)	9.3 (16/176)	9.3 (16/175)	17.4 (30/173)
Mean RT	562.43	560.37	611.14	627.3

Spanish, both labiodental /f/ and the spectrally similar dental fricative /θ/ in its phoneme inventory. If our hypothesis is correct, English listeners should also attend to transitional cues, in particular for /f/.

IV. EXPERIMENT III

A. Method

1. Materials and procedure

The materials were as in experiment II, i.e., the stimuli recorded by a native speaker of Spanish. The procedure and data analysis were as in the preceding experiments, with the exception that the target phoneme was not presented on screen. Grapheme-phoneme correspondences are often ambiguous in English; thus /f/ can be spelled as in "foal" or as in "phone," /s/ can also be represented by the letter "c," as in "cedar," and the letter "s" can stand for /s/, as in "basic," for /z/, as in "cousin," or for nothing as in "debris." Therefore, we specified the target in recorded instructions at the beginning of every block of pseudowords, instead of in visual target representations.

2. Participants

Twenty-seven students from the participant pool of the Laboratory of Experimental Psychology of the University of Sussex took part in this experiment. They were native speakers of English and none reported any speech or hearing disorders.

B. Results

Mean timeouts and RTs are shown in Table IV.

1. Timeouts

Both splicing (cross-spliced versus identity-spliced items) and fricative (/s/ versus /f/) were significant [splicing: $F(1,58)=5.76$, $p<0.05$; fricative: $F(1,57)=5.95$, $p<0.05$]. The interaction did not reach significance. The English listeners missed more items in the cross-spliced condition, and more /f/ than /s/.

2. RTs

0.4% of the data was below 150 ms, and was excluded from the analysis. Only fricative was significant in both analyses [$F1(1,78)=12.66, p<0.001, F2(1,56)=2.89, p<0.05$]. Listeners responded less rapidly to /f/ than to /s/.

C. Summary and discussion

English listeners also appear to pay attention to formant transitions. The crucial interaction between fricative and splicing was not significant, and therefore at this point we cannot decide with certainty whether English listeners make use of transition cues only for identification of /f/. However, the data suggest that English listeners, like Spanish listeners, are particularly affected in the case of /f/ (note that the effect of cross splicing, though statistically robust for both fricatives for these listeners, was twice as strong in the timeout errors for /f/ as for /s/—87% increase as opposed to 47%). Both English and Spanish listeners have learnt to distinguish between /f/ and /θ/, two highly confusable fricatives. This apparently made them more attentive to the additional acoustic cues in the formant transitions.

Previous research has shown that the labiodental fricative is hard to identify on the basis of spectral characteristics alone (Harris, 1958; Jongman *et al.*, 1998). So far we have shown that some listeners attend to transitional cues for this fricative. Our hypothesis, however, is that listener's use of transitional information in fricative identification reflects not just inherent distinctiveness of fricatives, but the presence of spectrally confusable pairs in the native fricative inventory. On this hypothesis, even fricatives which are generally easy to identify should encourage use of transitional information in a language which contains more fricatives with similar spectra.

The /s/ has been shown to be perceptually very salient because of the acoustic make-up of its noise spectrum (Wang and Bilger, 1973). During the articulation of /s/ air jets are created as the airflow passes the edges of the teeth; this results in relatively high intensity peaks in the high-frequency range of the spectrum, which serve as reliable cues and make this fricative acoustically robust. Listeners should nevertheless also exploit formant transitions to identify /s/, we predict, if other fricatives are close to /s/ in their native perceptual space.

We tested this in Polish, which has 11 fricatives [f v s z ʃʲ ʒʲ ʂ ʐ ɕ ʑ x]. The dental fricative is not present, so that /f/ is acoustically distinct from all other fricatives. The presence of the postalveolar, alveolopalatal, and palatal retroflex fricatives may, however, reduce the perceptual saliency of /s/. In acoustic terms, the /s/ typically has energy peaks in the frequency range between 3 and 7 kHz. The postalveolar /ʃ/ exhibits energy peaks in the frequencies between 1.5 and 5 kHz, while the Polish alveolopalatal /ɕ/ has its energy maxima in the range between 2 and 6 kHz. Finally, the retroflex Polish fricative shows its high energy peaks around 1 and 4 kHz (Jassem, 1968). This concentration of several fricatives with energy distributions in the same spectral range might hinder the identification of these fricatives in Polish. We therefore expect Polish listeners to pay attention to formant transitions for /s/.

TABLE V. Average percentages of timeouts and mean RTs in ms for the Polish listeners and the two fricatives in both splicing conditions in experiment IV. The absolute numbers of timeouts and the total numbers of trials are given in parentheses.

Fricative	/s/ identity spliced	/s/ cross spliced	/f/ identity spliced	/f/ cross spliced
Mean percentage of timeouts	5.5 (10/180)	12.7 (23/180)	0 (0/180)	3.3 (6/180)
Mean RT	652.09	654.54	688.1	676.6

V. EXPERIMENT IV

A. Method

1. Materials and procedure

Materials were as in experiments II and III, procedure was as in experiment II, and data analysis was as in all the preceding experiments.

2. Participants

Twenty-four students at the Uniwersytet Śląski in Katowice, all native Polish speakers, were paid to take part in this experiment. None reported any speech or hearing disorders.

B. Results

Table V shows the average timeouts and RTs.

1. Timeouts

Both main effects were again significant: splicing $[F(1,58)=10.19, p<0.01]$ and fricative $[F(1,57)=21.92, p<0.001]$. The interaction between fricative and splicing narrowly failed to reach significance $[F(1,56)=3.73, p<0.06]$. More timeouts occurred for the cross-spliced items, and for /s/ (9.16% versus 1.6% for /f/). Furthermore, the effect of splicing appeared smaller for /f/ than for /s/.

2. RTs

The main effect of fricative was significant in the analysis by participants only $[F1(1,69)=5.65, p<0.05]$. As Table V shows, the Polish RTs were relatively long.

C. Summary and discussion

Like Spanish and English listeners, Polish listeners are affected by misleading formant transitions. The phoneme repertoires of all three languages contain spectrally similar fricatives, and the results are thus in line with our hypothesis that listeners learn to direct their attention to subtle acoustic cues for fricative identification if required by their native phoneme repertoire. Furthermore, we can reject the possibility that listeners only take advantage of formant transitions in order to identify the spectrally diffuse and therefore perceptually less salient labiodental fricative. Even though we found no significant interaction between splicing and fricative for Polish listeners, the error data indicate that in contrast to all the other listener groups Polish listeners missed four times as many cross-spliced /s/ items than /f/ items. Especially the spectrally salient /s/ requires attention to formant transitions if this fricative can easily be confused with other fricatives in the listeners' phoneme repertoire.

On which level may such language-specific differences occur? We used the term attention to refer to listeners' learned selection of acoustic cues for phoneme identification, without assuming that listeners differ in sensitivity at the auditory level. Differences in sensitivity would imply that Dutch and German have "lost" such sensitivity. However, listeners are known to display sensitivity to foreign-language contrasts which fall entirely outside the range of the native phoneme repertoire (Best, McRoberts, and Sithole, 1988). Thus the effects that we have observed may reflect strategic listening choices which have no implications for the underlying sensitivity. If so, Dutch listeners, too, may perceive the acoustic mismatches if their attention is drawn to them. We tested this possibility in experiment V.

Furthermore, the phoneme inventories we have tested differ in whether or not they offer an alternative category in the case of an ambiguous fricative of a particular kind. In experiment V we also tested the effects of this response availability. We used an untimed open-choice identification task, with Dutch and Spanish listeners. If no response alternatives are given, participants are expected to choose a phoneme category from their native inventory. Spanish listeners may identify at least some of the cross-spliced /f/ tokens as /θ/. Dutch listeners, in contrast, should identify all tokens of cross-spliced /f/ as /f/. By asking subjects to judge the goodness-of-fit of the stimuli, we examined the extent to which both Dutch and Spanish listeners perceive mismatch effects of cross splicing.

VI. EXPERIMENT V

A. Method

1. Materials

Materials were the target-bearing VCV-strings of all 60 items used in experiment II, including the identity-spliced and cross-spliced targets (e.g., from the experimental item *tikufa* we presented the fragment *ufa*).

2. Procedure

Participants, seated in a sound-attenuated room, were presented with the VCVs over headphones. They were instructed to write down the intervocalic consonant, and to judge on a scale from 1 to 8 whether it was a poor or a good example of this consonant. After the test, participants identified the letters they used to describe the consonants by writing down a native example word containing each letter used.

3. Participants

Thirty-one students from the Radboud University Nijmegen took part in this experiment. Fourteen were native Dutch regular students, and 17 were native Spanish exchange students. They were paid for their participation. None reported any speech or hearing disorders.

B. Results

Dutch listeners always identified each of the stimuli as either /f/ or /s/. Spanish listeners, on the other hand, showed greater response variance. Five of the 17 Spanish listeners reported hearing exclusively /f/ and /s/, while the remaining 12 participants included other consonants in their responses. All cross-spliced /s/ were identified as /s/, but the responses for /f/ varied, including /b/, /d/, /m/ and, most frequently, the dental fricative /θ/. One item was identified by none of these 12 Spanish participant as /f/, but as a poor example of /θ/. All in all nine cross-spliced /f/ were identified by at least five Spanish participants as a consonant belonging to a category other than /f/.

The average ratings for the items which were correctly identified as either an /s/ or an /f/ were: for identity-spliced /s/, Dutch 3.95, Spanish 4.81; for cross-spliced /s/, Dutch 3.94, Spanish 4.67; for identity-spliced /f/, Dutch 3.78, Spanish. 4.53; for cross-spliced /f/, Dutch 3.01, Spanish 3.73. We analyzed the averaged ratings in an analysis of variance. We found main effects of language $[F(1,56)=120.77, p<0.001]$, splicing $[F(1,56)=21.96, p<0.001]$, and fricative $[F(1,56)=37.01, p<0.001]$ and an interaction between splicing and fricative $[F(1,56)=15.25, p<0.001]$. In general Spanish listeners rated the stimuli as better examples than Dutch listeners, probably because they were presented with their native phoneme realizations. The cross-spliced /f/ items were rated as poorer examples than the identity-spliced /f/ by both listener groups.

C. Discussion

Experiment V showed that the acoustic mismatch in the cross-spliced /f/ tokens turned them into poorer instances of /f/. While Dutch listeners just perceived these /f/ tokens as poorer members of the /f/ category, Spanish listeners identified some of these tokens as belonging to another category, most frequently as a /θ/. Thus the availability of an alternative category may be a crucial factor in determining whether the mismatch between fricative noise and formant transitions results in the perception of a different category. Although Dutch listeners seem to accept the cross splicing as allophonic variation of /f/, the goodness ratings showed that they too were sensitive to the acoustic mismatch.

We reanalyzed the timeout errors from experiment II, including for /f/ only the six items which the Spanish participants had always identified as /f/ when cross spliced. In this new analysis, the significant three-way interaction between language, splicing, and fricative no longer reached significance. This may be because that three-way interaction had been principally carried by the nine items which produced variable responses in experiment V; alternatively, of course, it could simply result from reduction of statistical power.

In an additional analysis we included the average Dutch ratings as a predictor for the Spanish timeout errors in experiment II. Splicing remained statistically significant $[F(1,57)=42.12, p<0.001]$. This result suggests that even though Dutch listeners perceive the acoustic manipulation in the stimuli, the cross splicing of the /f/ is definitely more harmful for the Spanish than for the Dutch listeners.

VII. GENERAL DISCUSSION

Many studies have investigated the contribution of formant transitions to fricative identification. Some studies reported robust effects whereas others failed to find any perceptual relevance of formant transitions for fricatives. In four phoneme detection experiments, we tested the hypothesis that attention to formant transitions as cues for fricative identification differs as a function of the presence of perceptually confusable fricatives in the listeners' native language. The targets in the detection experiments were /s/ and /f/ surrounded by either misleading (cross-splicing condition) or by coherent (identity-splicing condition) formant transitions. The stimuli were presented to Dutch, German, Spanish, English, and Polish listeners.

Our results support the hypothesis. First, target fricatives surrounded by misleading formant transitions were missed more often than fricatives with coherent formant transitions. This finding confirms previous work (Harris, 1958; Heinz and Stevens, 1961) showing that English listeners attend to formant transitions for some fricatives. More importantly, however, we observed a language-specific pattern of taking these acoustic cues into account for phoneme identification. Native listeners of Dutch and German, both languages without spectrally confusable fricatives, were not affected by misleading formant transitions. In contrast, listeners of Spanish and English, languages with the spectrally similar labiodental /f/ and dental /θ/ fricatives, and Polish, a language with spectrally similar sibilants, were affected by misleading formant transitions.

On the basis of the languages in which we found formant transitions to be used, we further queried whether attention to formant transitions is restricted to the spectrally similar contrasts only or whether it generalizes to nonconfusable fricatives. We found that transition cues were restricted to /f/ for the Spanish listeners. For Polish listeners, the crucial interaction between splicing and fricative narrowly failed to reach significance ($p=0.053$). But, as shown in Table V, the effect of splicing was greater for /s/ than for /f/. For English, the interaction between splicing and fricative did not reach significance, even though the effect is numerically greater for /f/ than for /s/. This may indicate that English listeners were also affected by misleading formant transitions for /s/. This is not incompatible with our hypothesis, if we take into consideration that English, in contrast to Spanish, has a postalveolar fricative category, which is spectrally more similar to /s/ than to /f/. Thus, with respect to our second hypothesis, we can tentatively conclude that attention to formant transition is restricted to spectrally similar fricative categories. Which fricatives are spectrally similar, of course, is a function of all fricative contrasts in a language, and their distribution in the perceptual space.

The pattern in our data, and in English in particular, might of course also have been affected by the particular splicing manipulation we applied to our stimuli. The frication noises of /f/ and /s/ differ in several ways; most impor-

tantly, /f/ has a flat diffuse spectrum, while /s/ shows prominent energy peaks. The spectra of /f/ and /θ/, and of /s/ and /ʃ/, however, show more similarities; cross splicing within these pairs might well show effects with English listeners. Whalen (1981) found that English listeners' categorization of an ambiguous synthetic fricative noise as either /s/ or /ʃ/ was influenced by formant transitions. In his experiment, a synthetic ten-step noise continuum was combined with coherent or inappropriate natural vocalic portions, including formant transitions. Interestingly, the formant transitions contributed to listeners' decision only at those steps of the noise continuum which modeled noise spectra with energy peaks appropriate for natural /ʃ/ or /s/ spectra. This suggests that for English listeners cross spliced stimuli containing fricative noise with more defined spectral peaks (as /ʃ/) in combination with mismatching formant transitions may lead to a similar effect for /s/, as found mainly for /f/. In our study, however, the difference between the cross-spliced pairs apparently overrode a potential confusion for the English listeners. Furtherresearch could investigate whether mismatching information in formant transitions to /s/ might also mislead English listeners—for example, into classifying an input as post-alveolar.

Importantly, the Polish data suggest that the acoustic make-up of a fricative by itself does not determine the use of formant transitions. Even though /s/ has salient acoustic characteristics (Harris, 1958; Strevens, 1960; Jassem, 1965) which make it perceptually very robust, Polish listeners were affected in particular for this fricative. Thus, the crucial factor in the use of formant transitions appears to be the acoustic make-up of a fricative in relation to all other fricatives in the phoneme inventory.

The present results indicate that listeners integrate cues in a language-specific way. The information conveyed in formant transitions appears to play a crucial role in determining fricative categorization for Spanish, English, and Polish listeners. This language-specific way of selecting cues for attention does not seem to be a strategy that a listener can easily adapt to the requirements of the situation, or to the experimental situation. The stimulus set in our experiments did not contain the dental fricative /θ/. That is, a direct distinction between the two confusable fricatives /f/ and /θ/ was not necessary for efficient performance within the experimental situation. Nonetheless, the Spanish and English listeners were substantially misled by incorrect formant transitions for /f/. Similarly, the Polish listeners were misled by incorrect formant transitions for /s/, even though the palatal fricatives, which in Polish might be confused with /s/, were not present in the experiment. This suggests that for listeners of these languages, formant transitions are part and parcel of the fricative categories.

We have distinguished "attention" from "sensitivity" to formant transitions. Experiment V showed that Dutch listeners perceive an acoustic difference between the identity- and cross-spliced items. They rated cross-spliced /f/ tokens as poorer examples of /f/, though in phoneme monitoring these poorer examples were not responded to significantly differently from the better examples. We assume that the attunement to a native language does not have any consequences on a low auditory level: sensitivity is unaffected. All listeners may perceive acoustic mismatches between formant transitions and noise spectrum, but language experience determines whether this information is attended to in fricative identification. Experiment V shows that the mismatching information in the transitions led Spanish listeners into the percept of a different fricative; the availability of more fricative categories encourages attention to subtle cues such as formant transitions. Where there is no alternative category—as in the case of Dutch—mismatching information in formant transitions may be treated as just allophonic variation. Thus what Spanish listeners in experiment V could identify as a dental fricative or even as a stop, Dutch listeners simply judged to be /f/. The number of possible choices for identifying an ambiguous stimulus has an effect on the distinctiveness of categories, and thus on listeners' response options. Recall that the goodness ratings of the Dutch listeners in experiment V did not suffice to explain the errors made by Spanish listeners, however. Thus the Dutch and Spanish listeners differed in how mismatching information affected fricative identification.

Primary cues are defined by some researchers (e.g., Stevens and Blumstein, 1981) as invariant acoustic properties which are independent of the phonetic context and sufficient to evoke the percept of a given phoneme. Secondary cues, in contrast, are context-dependent cues, exploited by listeners to support primary cues when needed, for instance in difficult listening conditions. We have shown that a context-dependent cue can also make an important and systematic contribution to fricative identification. Spanish listeners missed over 25% of the /f/ tokens which were surrounded by misleading formant transitions. The selection of primary and secondary cues appears to be language and phoneme specific, and depends on the degree to which cues enable listeners to distinguish native phoneme categories accurately and efficiently. Even though other acoustic characteristics, such as the generally higher intensity of the fricative noise, are used by listeners to distinguish sibilants from other fricatives, Polish listeners appear to use cues in the formant transitions, simply because of the number of confusable sibilants in their native phoneme repertoire.

In our experiments, listeners did not categorize or discriminate pairs of fricatives. In phoneme monitoring, participants react as soon as they recognize the target, and they do so only if the acoustic stimulus matches their abstract memory of the target. Reduced or mismatching information—here, the cross-spliced formant transitions—led Spanish, English, and Polish listeners into errors. Most previous studies of fricative perception have used untimed identification tasks. Results showed that Argentinian listeners could use transition information for some fricative contrasts (Borzone de Manrique and Massone, 1981), Dutch listeners apparently did not use it (Klaassen-Don, 1983), while English listeners appeared to use transition information in some studies (Harris, 1958) but not in others (Jongman, 1989). We cannot exclude the possibility that with unlimited response time listeners may be able to extract more information from static cues than they do in a running-speech situation, and that characteristics of particular experiments may have been

more versus less encouraging to such strategies. A task such as categorization (Whalen, 1981)[1], for example, could induce a different listening strategy; in categorization, listeners assign an acoustic signal to one or another category, and it is reasonable to assume that the mental representations of these categories, including the acoustic cues which distinguish between them, are in listeners' focus of attention, and might not need to be retrieved with every stimulus. This could affect both response accuracy and reaction times.

Adult listeners are specialized in identifying their native phonemes. An efficient way of selecting acoustic cues is thus another feature of language-specific processing which children must acquire in the course of their language development. In the same way that children learn to distinguish only native language contrasts (e.g., Werker and Tees, 1999; Sebastián-Gallés and Soto-Faraco, 1999), children must learn to be parsimonious with their attention to the subtle details of the acoustic signal and with the selection of relevant cues. Research by Nittrouer and colleagues (Nittrouer and Miller, 1997a, 1997b; Nittrouer, 2002) shows that there is indeed a developmental shift in the relevance of the cues conveyed by the frication and by the dynamics in the formant transitions for fricative identification. American English speaking children between 4 and 7 years of age show a developmental decrease in their weighting of formant transitions and a developmental increase in their weighting of the noise characteristics for /s/ and /ʃ/. On the other hand, another study by Nittrouer (2002) showed that American English speaking children and adults are more similar in assigning weight to formant transitions for the distinction between the labiodental and the dental fricatives. Thus, the developmental shift is restricted to the contrasts which are sufficiently characterized by the static cues alone. Nittrouer argues that the attention/sensitivity to dynamic cues diminishes when children learn which cues carry "phonetic informativeness" in their native language.

Children's speech perception differs even up to 10 years of age from adults' speech perception (Elliot and Katz, 1980). Nittrouer's developmental weighting shift theory contrasts with, for instance, explanation in terms of auditory cortex maturation (Sussman, 2001). Most of the data relevant to this debate come so far from English, and we suggest that the debate would profit from additional data from other languages, for instance, the five languages of the present study. Our results show that children will reorganize their sensitivity to formant transitions in a language specific way to spectrally similar fricatives. English, Spanish, and Polish children should keep their attention to formant transitions, whereas Dutch and German children will not.

The shift in attention during language development entails that a listener would have to reacquire, or reorganize attention to these cues in order to attain a native-like perception in a second language. Previous research (Repp, 1981; Hazan, Iverson, and Bannister, 2005) suggests that listeners can indeed direct attention to otherwise unused phonetic cues, at least after being exposed to sufficient training. Future research will have to determine how rapidly speakers of a language without perceptually similar fricatives can learn to take advantage of formant transitions to efficiently distinguish between perceptually similar fricatives in a second language.

Are fricatives perceived only on the basis of the static characteristics of their fricative spectrum, or do formant transitions also play a role? A large number of studies have addressed these questions, but the pattern of results, as we demonstrated in the Introduction, has been contradictory. Previous studies have examined the question in different languages; and language-specific phonology may be the key to whether listeners rely solely on spectral cues to fricative identity, or also attend to transition information. Even though all listeners will always make use of information in the fricative spectrum, for listeners of some languages formant transitions also play a crucial role for some of their native fricatives. Mismatching acoustic information in formant transitions may be perceived by all listeners at a low phonetic level, but the use of this information for the identification of a given fricative seems to depend on whether the spectral characteristics of its frication suffice to distinguish this fricative from all other fricatives in the listener's language.

ACKNOWLEDGMENTS

The authors are grateful to Professor Alan Garnham from the University of Sussex and to Dr. Jolanta Tambor from Uniwersytet Śląski in Katowice for their help in conducting the experiments in England and in Poland. They would further like to thank James McQueen for his helpful comments on an earlier version of this text. This research was supported by the NWO SPINOZA project "Native and Non-Native listening."

[1]Note that Whalen (1981) research also showed effects of context vowels on the identification of fricatives. We in fact included the context vowels as factors into our analyses. As these results did not prove to be language specific, however, we do not report them in detail.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants," J. Exp. Psychol. Hum. Percept. Perform. **14**, 45–60.

Borzone de Manrique, A. M., and Massone, M. I. (**1981**). "Acoustic analysis and perception of Spanish fricative consonants," J. Acoust. Soc. Am. **69**(4), 1145–1153.

Connine, C., and Titone, D. (**1996**). "Phoneme monitoring," Lang. Cognit. Processes **11**, 635–645.

Costa, A., Cutler, A., and Sebastián-Gallés, N. (**1998**). "Effects of phoneme repertoire on phoneme decision," Percept. Psychophys. **60**, 1022–1031.

Elliot, L. L., and Katz, D. (**1980**). "Children's pure-tone detection," J. Acoust. Soc. Am. **67**, 343–344.

Fruchter, D., and Sussman, H. (**1997**). "The perceptual relevance of locus equations," J. Acoust. Soc. Am. **102**(5), 2997–3008.

Hamann, S. (**2003**). *The Phonetics and Phonology of Retroflexes*, (LOT, Utrecht).

Harris, K. S. (**1958**). "Cues for the discrimination of American English fricatives in spoken syllables," Lang Speech **1**, 1–7.

Hazan, V., Iverson, P., and Bannister, K. (**2005**). "The effect of acoustic enhancement and variability on phonetic category learning by L2 learners," *Proceedings of the ISCA Workshop on Plasticity in Speech Perception*, London, UK, June 2005.

Heinz, J. M., and Stevens, K. N. (**1961**). "On the properties of fricative consonants," J. Acoust. Soc. Am. **33**, 589–593.

Jassem, W. (**1965**). "Formants of fricative consonants," Lang Speech **8**, 1–16.

Jassem, W. (**1968**). "Acoustic description of voiceless fricatives in terms of

spectral parameters," in *Speech Analysis and Synthesis*, edited by W. Jassem, Panstwowe Wydawnictwo Naukowe, Warsaw, pp. 189–206.
Jassem, W. (**2003**). "Illustrations of the IPA: Polish," J. Int. Phonetic Assoc. **33**(1), 103–107.
Jongman, A. (**1989**). "Duration of fricative noise required for identification of English fricatives," J. Acoust. Soc. Am. **85**, 1718–1725.
Jongman, A., Sereno, J., Wayland, R., and Wong, S. (**1998**). "Acoustic properties of English fricatives," J. Acoust. Soc. Am. **103**, 3086.
Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**(3), 1252–1263.
Klaassen-Don, L. E. O. (**1983**). "The influence of vowels on the perception of consonants," Doctoral dissertation, Leiden University (unpublished).
LaRiviere, C., Winitz, H., and Herriman, E. (**1975**). "The distribution of perceptual cues in English prevocalic fricatives," J. Speech Hear. Res. **18**, 613–622.
Martin, J. G., and Bunnell, H. T. (**1981**). "Perception of anticipatory coarticulation effects," J. Acoust. Soc. Am. **69**, 559–567.
McQueen, J. M., Norris, D., and Cutler, A. (**1999**). "Lexical influence in phonetic decision making: Evidence from subcategorical mismatches," J. Exp. Psychol. Hum. Percept. Perform. **25**, 1363–1389.
Nittrouer, S. (**2002**). "Learning to perceive speech: How fricative perception changes, and how it stays the same," J. Acoust. Soc. Am. **112**(2), 711–719.
Nittrouer, S., and Miller, M. E. (**1997a**). "Developmental weighting shifts for noise components of fricative-vowel syllables," J. Acoust. Soc. Am. **10l**(1), 572–580.
Nittrouer, S., and Miller, M. E. (**1997b**). "Predicting developmental shifts in perceptual weighting schemes," J. Acoust. Soc. Am. **101**(4), 2253–2266.
Otake, T., Yoneyama, K., Cutler, A., and van der Lugt, A. (**1996**). "The representation of Japanese moraic nasals," J. Acoust. Soc. Am. **100**(6), 3831–3842.
Repp, B. (**1981**). "Two strategies in fricative discrimination," Percept. Psychophys. **30**(3), 217–227.
Sebastián-Gallés, N., and Soto-Faraco, S. (**1999**). "Online processing of native and non-native phonemic contrasts in early bilinguals," Cognition **72**, 111–123.
Stevens, K. N. (**1998**). *Acoustic phonetics* (MIT Press, Cambridge, MA).
Stevens, and Blumstein, S. E. (**1981**). "The search for invariant acoustic correlates of phonetic features," in *Perspectives on the Study of Speech*, edited by P. L. Miller (Erlbaum, Hillsdale, NJ) pp. 1–38.
Strevens, P. (**1960**). "Spectra of fricative noise in human speech," Lang Speech **3**, 32–49.
Sussman, H. (**2001**). "Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions?," J. Acoust. Soc. Am. **109**(3), 1173–1180.
Wang, M. D., and Bilger, R. C. (**1973**). "Consonant confusions in noise: A study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.
Weber, A. (**2001**). "Help or hindrance: How violation of different assimilation rules affects spoken-language processing," Lang Speech **44**, 95–118.
Werker, J. F., and Tees, R. C. (**1999**). "Influences on infant speech processing: toward a new synthesis," Annu. Rev. Psychol. **50**, 509–535.
Whalen, D. H. (**1981**). "Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary," J. Acoust. Soc. Am. **69**, 275–282.
Zygis, M., and Hamann, S. (**2003**). "Perceptual and acoustic cues of Polish coronal fricatives," *Proceedings of the 15th ICPhS*, pp. 395–398.

Cross-language sensitivity to phonotactic patterns in infants

Sachiyo Kajikawa[a)]
NTT Communication Science Laboratories, NTT Corporation, 2-4 Hikari-dai, Seika-cho, Souraku-gun, Kyoto 619-0237, Japan and Tamagawa University Research Institute, 6-1-1 Tamagawagakuen, Machida, Tokyo, 194-8610, Japan

Laurel Fais
Infant Studies Centre, University of British Columbia, 2136 West Mall, Vancouver, British Columbia, V6T 1Z4 Canada

Ryoko Mugitani
NTT Communication Science Laboratories, NTT Corporation, 2-4 Hikari-dai, Seika-cho, Souraku-gun, Kyoto 619-0237, Japan

Janet F. Werker
Infant Studies Centre, University of British Columbia, 2136 West Mall, Vancouver, British Columbia, V6T 1Z4 Canada

Shigeaki Amano
NTT Communication Science Laboratories, NTT Corporation, 2-4 Hikari-dai, Seika-cho, Souraku-gun, Kyoto 619-0237, Japan

(Received 23 December 2005; revised 11 July 2006; accepted 25 July 2006)

This study explored sensitivity to word-level phonotactic patterns in English and Japanese monolingual infants. Infants at the ages of 6, 12, and 18 months were tested on their ability to discriminate between test words using a habituation-switch experimental paradigm. All of the test words, *neek, neeks*, and *neekusu*, are phonotactically legitimate for English, whereas the first two words are critically noncanonical in Japanese. The language-specific phonotactical congruence influenced infants' performance in discrimination. English-learning infants could discriminate between *neek* and *neeks* at the age of 18 months, but Japanese infants could not. There was a similar developmental pattern for infants of both language groups for discrimination of *neek* and *neeks*, but Japanese infants showed a different trajectory from English infants for *neekusu/neeks*. These differences reflect the different status of these word patterns with respect to the phonotactics of both languages, and reveal early sensitivity to subtle phonotactic and language input patterns in each language. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2338285]

PACS number(s): 43.71.Ft, 43.71.Hw [MSS] Pages: 2278–2284

I. INTRODUCTION

The speech perception of humans is tuned to their native language. Infants as young as 6 months of age have already started acquiring native vowel categories, and they process their native consonant categories effectively at around the end of the first year of life (review: Kuhl, 2004; Saffran *et al.*, 2006). This has been demonstrated in previous studies using words of a single syllable in which the target sounds are in word-initial position. As a result of this perceptual accommodation, for example, 10- to-12-month-old English-learning infants come to ignore a non-native contrast such as the Hindi retroflex versus dental stop contrast, even though they discriminate the same contrast at 6–8 months (Werker and Tees, 1984). During the same age period, infants show improvement in their discrimination of phonetic contrasts used in the native language (Kuhl *et al.*, 2006). Similar developmental changes in phoneme perception have been observed also in infants who learn other languages, such as Spanish and Swedish (Kuhl *et al.*, 1992; Bosch and Sebastián-Gallés, 2003).

In addition to phoneme categories, each language has its own set of phonotactic rules, that is, acceptable patterns of phoneme sequences. One kind of phonotactic rule involves the possible positions for phonemes or phoneme sequences in words. For example, the /dl/ cluster is not allowed in word-initial position in French (Hallé *et al.*, 1998), and /zd/ sequence is never observed in word-initial position in English (Mattys and Jusczyk, 2001). Another kind of phonotactic rule is concerned with the combination of phonemes. For example, the combination of [k] and [n] in this order within a syllable is illegitimate in English (Friederici, 2005).

Previous studies have shown that infants before one year of age are already sensitive to phonotactic features of their native language. Friederici and Wessels (1993) reported that 9-month-old Dutch infants have started learning the phonotactic structure of word boundaries. Infants presented with nonsense words /muks/ and /ksmu/ preferred listening to /muks/, which is a legal pattern in Dutch, rather than /ksmu/,

a)Author to whom correspondence should be addressed; electronic mail: kajikawa@lab.tamagawa.ac.jp

which is an illegal pattern. Infant knowledge of phonotactic patterns has also been revealed in discrimination between native language words and non-native language words. In another study, English-learning 9-month-old infants were presented with two language word lists, English and Dutch (Jusczyk *et al.*, 1993). There are some phonetic differences in English and Dutch, but only phonemes that are allowed in both languages were used in one of a series of experiments. For example, the word "knevel" is a legal form for Dutch but an illegal pattern for English, even though each phoneme is legal in both Dutch and English. In these experiments, English-learning infants preferred legal word patterns for English to word patterns that were legal only for Dutch. This result demonstrated that infants' ability to discriminate word lists can be based on phonotactic features.

Nine-month-old English-learning infants are also sensitive to the probability of sound combinations even within the native language (Jusczyk *et al.*, 1994). That is, infants preferred words consisting of frequent sound combinations to those consisting of rare sound combinations. Indeed, infants can even learn to prefer a particular syllable structure (e.g., CVCV over CVCCVC) following only a brief pattern induction phase (Saffran and Thiessen, 2003).

The knowledge of phonotactic patterns also plays an important role in speech segmentation. When infants were presented with target words embedded in a sentence with good phonotactic word boundary cues, they could recognize words (Mattys and Jusczyk, 2001). CVC target words were surrounded by consonants, yielding C-CVC-C forms with C-C clusters at word-initial and word-final positions (V: vowel, C: consonant). The C-C clusters used in one condition infrequently occur in word-medial position in a child-directed corpus. Therefore, these C-C clusters would be good cues for word boundaries. In another condition, C-C clusters that frequently appear in word-medial position were placed in word-initial and word-final positions. In this case, infants tended to interpret the C-C combination as a sequence within a word, and they failed to find the target words. Infants can learn these phonotactic regularities by being exposed to speech stimuli for just a few minutes. Saffran, Newport, and Aslin (1996) reported that infants recognized "words" embedded in phoneme sequences without prosodic cues. In their study, only distributional regularities were controlled and were available to be used as a cue for word boundaries.

The knowledge of native phonotactic patterns has an influence on speech perception in adults. One language-specific perception that is based on native phonotactic patterns is Japanese speakers' auditory illusion of an epenthetic vowel between the consonants in a voiced consonant cluster in words of the form VCCV, like "ebzo" (Dupoux *et al.*, 1999). In this case, Japanese speakers hear /u/ between /b/ and /z/, and their representation of "ebzo" becomes "ebuzo." This auditory illusion relates to the fact that CC clusters are noncanonical in Japanese. Consonants are always followed by vowels except for some special morae.[1] That is, most words containing a CC cluster, such as *tabemno* (correctly *tabemono*, food), are not acceptable as Japanese words. Therefore, in Dupoux *et al.*'s study, Japanese speakers automatically perceived a vowel which actually did not exist between the two voiced consonants. In another study, it was shown that French speakers "misperceive" illegal word-initial /dl/ and /tl/ clusters as legal /gl/ and /kl/ (Hallé *et al.*, 1998). This kind of auditory illusion or perceptual assimilation of illegal clusters to legal patterns may promote speech processing in cases in which sounds are omitted or pronounced ambiguously.

Despite the fact that in Japanese, consonant clusters are not canonical, these clusters are actually observed and accepted in some contexts of fluent speech. High vowels (/i/, /u/) tend to be devoiced when they appear between voiceless consonants or between a voiceless consonant and a pause (e.g., *k⟨i⟩sha*, steam train). When speakers of the Tokyo dialect produce words containing such devoicing contexts, vowel devoicing occurs at a rate of 70%–80% (Imaizumi *et al.* 1999). Japanese speakers are able to recognize the acoustic difference between the canonical pattern (e.g., "*wakakusa*" fresh grass) and the devoiced pattern ("*wakaksa*"), and they recognize the canonical pattern as a better exemplar of Japanese words than the devoiced pattern (Fais *et al.*, 2005). Because of the devoicing process, however, Japanese speakers do not find a difference in meaning between the two forms.

This study investigated the influence of native phonotactic characteristics on infants' perception of words. In particular, we focused on C pause and CC clusters in word-final position for Japanese-learning and English-learning infants. CVC and CVCC sequences are violations of canonical Japanese phonotactic rules. However, because of the process of devoicing, CVC and CVCC forms are not only accepted but occur at a high rate in fluent Japanese speech. On the other hand, both CVC and CVCC sequences fit English phonotactic rules. Infants' performance was explored from two points of view in our study. The first involved the discrimination of a CVCC word, *neeks*, and a CVCVCV word, *neekusu*. Because the last two morae of *neekusu* constitute devoicing contexts in Japanese, *neeks* is a possible, fluent speech pronunciation of the word *neekusu*. Therefore, *neeks* and *neekusu* could be perceived as the same word for Japanese infants. This hypothesis is supported by the work done by Dupoux and colleagues on the auditory illusion of an epenthetic vowel in CC environments (Dupoux *et al.*, 1999). However, in a study in which Japanese adults rated the goodness of these words, most of the Japanese adult participants discriminated *neeks* and *neekusu* (Fais *et al.*, 2005). In the study, Japanese speakers rated the canonical and noncanonical forms differently; further, they rated those with devoicing contexts significantly differently from those without devoicing contexts. This demonstrated that adults are sensitive to the possible acceptability of some final C's or CC's. Therefore Japanese-learning infants might discriminate *neeks* and *neekusu* according to their knowledge of sequence acceptability in Japanese. Thus, there are two possible predictions for Japanese infants: they might fail to distinguish *neeks* and *neekusu* based on their recognition that these two forms can be related by vowel devoicing and thus constitute the "same" word, or they might discriminate the two forms on the basis of their surface phonetic differences. English-learning infants, in whose language environment *neeks* and *neekusu* are

TABLE I. The makeup of the Japanese and English groups.

	Gender	Mean age	Age range
Japanese			
6 months	14M[a] 10F[b]	6m[c]19d[d]	6m[c]5d[d]–7m[c]4d[d]
12 months	11M[a] 13F[b]	12m[c]9d[d]	12m[c]0d[d]–12m[c]27d[d]
18 months	11M[a] 13F[b]	17m[c]26d[d]	17m[c]10d[d]–18m[c]17d[d]
English			
6 months	12M[a] 12F[b]	6m[c]16d[d]	6m[c]0d[d]–7m[c]4d[d]
12 months	12M[a] 12F[b]	12m[c]18d[d]	11m[c]24d[d]–13m[c]3d[d]
18 months	12M[a] 12F[b]	17m[c]27d[d]	17m[c]10d[d]–18m[c]14d[d]

[a]M: male.
[b]F: female.
[c]m: month.
[d]d: day.

definitely different words in legal forms, are predicted to discriminate these two forms.

The second point is the discrimination of a CVCC word, *neeks*, and a CVC word, *neek*. It might be the case that these two words will be considered different by Japanese-learning infants because there is no process, like vowel devoicing, that could relate the two words in Japanese. Thus, both Japanese and English-learning infants will discriminate these two forms. On the other hand, these two words could be difficult for Japanese infants to discriminate since both words are non-canonical Japanese forms. As previous studies have shown, infants before one year of age are sensitive to phonotactic patterns, preferring legal word forms to illegal forms. This predicts that Japanese infants would show clear discrimination of *neeks/neekusu* because *neeks* is phonotactically illegal and *neekusu* is legal. However, it makes no prediction in the case of *neek/neeks*, in which both forms are illegal. For English infants, both pairs are legal-legal combinations and they should show good performance in discrimination for both combinations.

II. METHODS

A. Participants

Japanese and Canadian infants at the ages of approximately 6, 12, and 18 months were tested (Table I). All the infants were born after a 37-week gestational period and had had no problems in vision or hearing based on parental report. The parents of Japanese participants were Japanese native speakers living in the Kinki (western) area of Japan and those of Canadian participants were English native speakers living in Vancouver, Canada. Twenty-four infants were assigned to each age group of the two language groups. Most of the Japanese infants had opportunities to hear English speech from TV programs or CDs, but they had never lived with nor met regularly any native English speaker. Canadian infants had had little or no exposure to Japanese language. Data from an additional 25 infants of the Japanese group and 48 of the English group were excluded in the analysis because of infants' inappropriate condition (crying, sleeping, fussing, etc.), failure to reach the criterion for number of habituation trials ($>$6 trials), or technical problems with the experimental equipment.

B. Stimuli

The stimuli were three nonsense words: *neek* (/ní:k/) a CVC word, *neeks* (/ní:ks/) CVCC, and *neekusu* (/ní:kusu/) CVCVCV. Canonically, CVC and CVCC words do not follow Japanese phonotactic rules, but they are possible words in devoicing contexts in fluent speech. On the other hand, CVCVCV words do follow Japanese phonotactic rules. For English, all the words are legitimate word forms.

The acoustic difference between *neeks* and *neek* is a single consonant (/s/) in word-final position and the difference between *neeks* and *neekusu* is the presence of two vowels (/u/ and /u/) in word-medial and -final positions. Both *neeks* and *neek* are possible forms in Japanese fluent speech, derived by vowel devoicing from canonical forms neekusu and neeku, respectively. *Neeks* and *neekusu* could be the same word for Japanese speakers, related in the same way as [kaks] (hide) and /kakusu/ (hide). In terms of English phonology, *neeks*, *neek*, and *neekusu* are all possible forms. For English speakers, *neeks* and *neek* could be related morphologically, as singular/plural, present third person, or possessive.

The stimulus words were produced by a female English-Japanese bilingual speaker 18 years of age, who had grown up in an American English-speaking family and had lived in Japan for 13 years, attending Japanese schools throughout that time. The words were presented to the speaker in English spelling because there is no kanji or katakana for some of the word forms. The speaker was instructed to pronounce the words with Japanese articulation of the consonants, but with no epenthetic vocalic material at the end of *neek* or in the devoiced vowel positions in *neeks*. It was necessary to use a bilingual speaker because it is difficult for native Japanese speakers to refrain from introducing vocalic material in these positions. The speaker produced multiple tokens of each word until the canonical form *neekusu* was judged to be "natural-sounding" Japanese by three native speakers of Japanese (see Fais *et al.* 2005 for Japanese adults' ratings of these forms). The speaker produced words in an infant-directed style. The words were recorded into a computer at a sampling rate of 16 bit, 44.1 kHz. Five intonational variations of each word were selected and the variations were matched across the three words. Speaking rates were almost the same among the stimulus words.

C. Procedure

The infants were tested in a habituation-switch paradigm (Stager and Werker, 1998). The experiments were conducted in a sound-attenuated booth, whose inside wall was covered with black cloth. An infant sat on the parent's lap and faced a 19 in. PC display. Behind the center cloth, a loudspeaker and a video camera were hidden. An experimenter outside the booth controlled the presentation of audio and visual stimuli and recorded the infant's response by monitoring eye direction using the Habit 2002 program (Cohen *et al.*, 2002).

In the beginning of each trial, an attention getter appeared on the display. When the infant looked at the display, the trial was started and a red-black checkerboard appeared. During each 14 s trial, an audio stimulus was presented

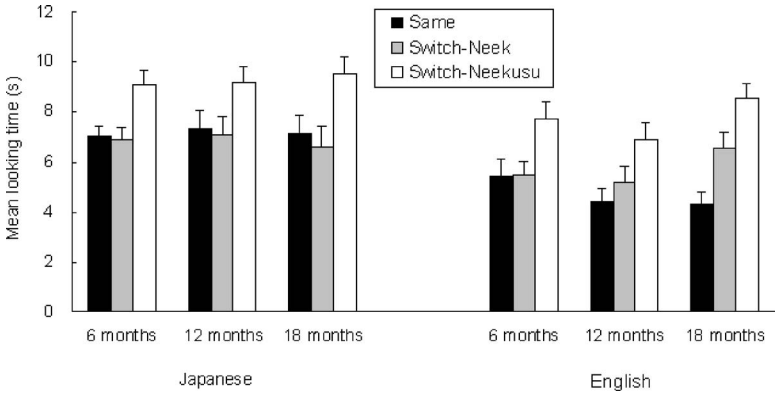


FIG. 1. Mean looking times of Japanese-learning and English-learning groups. Error bars stand for one standard error.

through a loudspeaker. During a trial, a word was presented seven times, using five different intonational variants, with approximately 1 s intervals between words. When the infant looked at the display, the experimenter pushed a key and recorded looking time on the computer. The total looking time was summed as the data for a given trial.

A session consisted of a habituation phase and a test phase. In the habituation phase, a CVCC word (*neeks*) was presented. The habituation phase ended when the total looking time of the last three trials was less than 65% of the total looking time of the block of three trials that had the highest looking time. If the number of habituation trials reached 27 without reaching the 65% criterion, the habituation phase automatically ended and the test phase started. These cases were not included in the analyses. In the subsequent test phase, three stimuli were presented: CVCC (*neeks*), i.e., the same stimulus as that of the habituation phase, CVC (*neek*) as switch 1, and CVCVCV (*neekusu*) as switch 2. The test order was counterbalanced among infants.

After the experiment, infants' eye directions were labeled frame by frame in offline coding as either looking at the monitor or looking away from the monitor. Looking time obtained from this offline coding was used in the analysis. In the habituation-switch paradigm, recovery of looking to a switch stimulus is interpreted as showing that the infant could discriminate the switch stimulus and the habituated stimulus. Therefore, we calculated the recovery of looking during a switch trial and compared it to looking time to the "same" trial.

III. RESULTS

Figure 1 shows the mean recovery of looking time in each age group of Japanese and English infants. We conducted a two-way ANOVA in each language group (word by age mixed design, three word conditions: "same" trial CVCC, switch-*neek* trial CVC, and switch-*neekusu* trial CVCVCV, and three age conditions: 6, 12, and 18 months). For the Japanese group, the effect of word was significant ($F(2,138)=21.26$, $p<0.01$). Neither the effect of age nor the word by age interaction were significant. Post hoc tests (LSD) showed that the infants recovered significantly when they were presented with CVCVCV in a test trial ($p<0.01$, difference of mean looking times between "same" condition and switch-*neekusu* condition, 6 months: diff=−2.05, 12 months: diff=−1.88, 18 months: diff=−2.42). On the other hand, when the infants heard CVC in the switch-*neek* condition, the looking time was not significantly different.

For English infants, the main effect of word was also significant ($F(2,138)=27.19$, $p<0.01$). The effect of age and the word by age interaction was not significant. The differences of looking times between "same" condition and switch-*neekusu* CVCVCV condition were significant in all age groups ($p<0.01$, 6 months: diff=−2.24, 12 months: diff=−2.52, 18 months: diff=−4.21). The differences between "same" condition and switch-*neek* CVC condition were also significant ($p<0.02$). When we tested these differences individually for each age group, only English 18-month-old infants looked longer in the switch-*neek* trial CVC than in the "same" trial (diff=−2.25, $p<0.01$). English 6-month and 12-month infants did not show this tendency, similar to Japanese infants.

The percentage of infants who showed recovery of looking time in the test trials is an index for developmental change in discrimination abilities. In the switch-*neek* condition, eight out of 24 infants showed recovery for *neek* after habituation to *neeks* in the 6-month group, 10 infants did so in the 12-month group, and 13 in the 18-month group. That is, the percentage of infants who discriminated *neek* from *neeks* increased gradually with age (Regression analysis: x =age, y=percentage of infants, $y=22.00+1.75x$, $F(1,2)=147.00$, $p=0.05$), as shown in Figure 2. This tendency is identical to that for the English group. For the English group, 10 infants showed recovery at 6 months, 13 at 12 months, and 16 at 18 months ($y=29.33+2.08x$, $F(1,2)=1875.00$, p

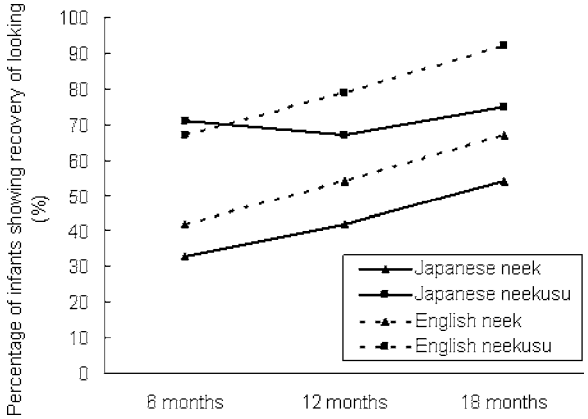


FIG. 2. Percentages of infants who showed recovery of looking in switch trials compared to "same" trial.

<0.02). The percentages of infants who showed recovery to *neek* from *neeks* were higher in the English group than in the Japanese group by 10%–13%. However, the percentages were not significantly different between the two language groups for each age (two-sample test for equality of proportions).

The result was different between the two language groups for the switch-*neekusu* condition than for the switch-*neek* condition, which was described above. The number of infants who showed recovery to *neekusu* in the Japanese group was 17 at 6 months, 16 at 12 months and 18 at 18 months. The percentage of Japanese infants showing recovery did not change with age ($y=67.00-0.33x$, *ns*). On the other hand, the number of infants showing recovery in the English group was 16 at 6 months, 19 at 12 months, and 22 at 18 months. The percentage clearly increased in this language group from 6 to 18 months of age ($y=54.33-2.08x$, $F(1,2)=1875.00$, $p<0.02$). In addition, the percentage of recovery in 6-month infants is almost the same in both language groups, but it is different in the 18-month group: 92% for the English group and 75% for the Japanese group.

Comparing the two switch conditions, *neek* and *neekusu*, in each age group, there was also a cross-linguistic difference. For the Japanese group, the difference between *neek* and *neekusu* in percentage of infants showing recovery became smaller and smaller with age (percentage difference between *neek* and *neekusu*: 38% at 6 months, 25% at 12 months, and 13% at 18 months). In contrast, for the English group, the difference of recovery percentage did not change with age: the differences were all 25%. In other words, the change with age in discrimination of *neek-neeks* and *neekusu-neeks* was parallel in the English group but not in the Japanese group.

IV. DISCUSSION

This study demonstrated some differences between Japanese-learning infants and English-learning infants in their sensitivity to phoneme patterns in words, as well as some common developmental trends. There are three factors determining infant performance of discrimination: phonetics, phonotactics, and surface input in the environment (the vowel devoicing context in this study). We discuss differences and similarities in Japanese and English language groups for two different cases, *neek/neeks* and *neekusu/neek*, according to these three factors.

A. Phonetics

From the viewpoint of phonetic discrimination, Japanese- and English-learning infants should perceive the stimulus words in the same way. In our experimental design, we compared infants' responses to *neekusu* and *neeks*, and to *neek* and *neeks*. Acoustically, the test CVCVCV word, *neekusu*, is greatly different from the habituated CVCC word, *neeks*, especially as compared to the other test CVC word, *neek*. That is, the difference between *neekusu* and *neeks* is two vowels (/u/), which create two additional syllables, and the difference between *neek* and *neeks* is one consonant (/s/), with the number of syllables the same. Therefore, we might predict that *neekusu* is acoustically more discriminable from *neeks* than *neek* is for both language groups of infants. This prediction was supported by the results of our experiment: both Japanese- and English-learning infants could detect the change from one syllable to three in the test words.

In contrast, 6- and 12-month-old infants of both language groups could not discriminate *neek* from *neeks* (18-month-old infants will be discussed below). The percentage of infants who showed discrimination between *neek* and *neeks* increased linearly in both language groups. It is possible that the distinction between the sounds /k/ and /ks/ is a difficult one and may develop late, much as the /th/-/d/ distinction does for English-hearing babies (Polka *et al.*, 2001). Alternatively, phoneme sequences at the ends of words may be less salient and thus more difficult to discriminate than those at the beginning of words (Jusczyk *et al.*, 1999). By both of these accounts, the *neek/neeks* distinction is difficult for 6- and 12 month-old-infants.

B. Phonotactics

Note that the stimulus words have different phonotactic status in Japanese and English, as described in Table I. CVC (*neek*) and CVCC (*neeks*) are phonotactically valid forms for English but not for Japanese. These language-specific phonotactic patterns were reflected in English and Japanese infants' different performances. Only English-learning 18-month infants reached statistical significance testing looking time to *neek* and *neeks*. Despite the fact that the contrast was presented in a low-salience position (Jusczyk *et al.*, 1999), the more language-experienced English-hearing 18-month-old infants were able to discriminate these two legal word forms. On the other hand, even though the percentage of Japanese infants who showed recovery of looking for *neek* increased with age, as might be expected for discrimination of a difficult phonetic contrast, Japanese infants did not show significant discrimination. They were not able to discriminate between two phonotactically illegal word forms.

The difference between *neekusu* and *neeks* should be clearly recognized by English infants since *neekusu* and *neeks* are phonotactically different word forms. Our results show that English-learning infants could discriminate these words from 6 months of age and the percentage of infants who discriminated these increased gradually with age. In Japanese, on the other hand, these two words might be interpreted as the same word, because *neeks* is a possible devoiced form of *neekusu*. However, Japanese infants did discriminate the two forms, and thus did not treat *neekusu* and *neeks* as the same word. This result supports our prediction that infants would be able to discriminate *neekusu/neeks* (legal/illegal contrast) based on their preference for legal over illegal word forms. However, note that this result also reflects salient differences at the phonetic level as discussed above.

C. Surface input

As described above, Japanese infants of all three ages discriminated *neekusu* from *neeks*. This result is consistent

with the adult ratings of these two words: adult Japanese speakers also discriminate these words in terms of goodness of form (Fais *et al.*, 2005). However, as is the case with the adult data, there is a noteworthy tendency in Japanese infants' responses that reveals infants' sensitivity to the devoicing relationship between *neekusu* and *neeks*. The percentage of Japanese infants who showed recovery of looking to *neekusu* did not increase with age, whereas the percentage of English infants who showed recovery increased linearly. This tendency might be related to the surface input of devoicing forms in Japanese. Namely, the speech input that Japanese infants receive includes both CVCVCV and CVCC forms. Rather than getting better and better at differentiating these forms, then, Japanese infants show discrimination of the forms, but do not increase the degree of discrimination as English infants do. At the age of 18 months, Japanese learners may start to process CVCVC and CVCC differently from English learners. This type of response could well underlie the fine-grained appreciation that Japanese adults develop for the acceptability of word forms that contain devoicing contexts.

To understand fully the results pertaining to Japanese infants' understanding of devoicing contexts, it is necessary to know more about vowel devoicing rates and contexts in Japanese infant-directed speech. While no research has yet been done on that topic, other previous work has shown that teachers reduced vowel devoicing to hearing-impaired children (Imaizumi *et al.*, 1995). To provide good exemplars to infants who are learning language, caregivers may in fact avoid vowel devoicing. If this speculation is true, the low frequency of devoicing input from mothers may impede young infants' acquisition of vowel devoicing patterns.[2] In any event, by the age of approximately 18 months, infants do show behavior that seems to be shaped by their understanding of phonotactic patterns and of the surface appearance of consonant clusters.

V. CONCLUSIONS

To explore the influence of native language phonotactic patterns on infant speech perception, we tested infants' ability to discriminate phoneme sequences. The results demonstrated two points of influence from the native language: (1) if two illegitimate word form patterns are contrasted, it is more difficult for infants to discriminate the patterns than if two legitimate patterns are contrasted, (2) surface input of word form patterns, those involving devoicing contexts in this study, has begun to influence infant discrimination by the age of 18 months. The first point is supported by the results of discrimination for *neek-neeks* at 18 months of age. English-learning infants, for whom both *neek* and *neeks* are legitimate phoneme sequences, could discriminate these patterns, whereas Japanese-learning infants, for whom both are illegitimate patterns, could not. The second point is supported by the fact that the number of Japanese infants who discriminated *neekusu* and *neeks* did not increase from 6 to 18 months of age. In contrast, the number of infants who showed discrimination of these patterns did increase for the English-learning group. We interpret this result as reflecting the cross-linguistic difference in the phonotactic status of the surface input of phoneme sequences: in English, *neekusu* and *neeks* are legal, unrelated forms; in Japanese they are potentially related by the legitimate and common process of vowel devoicing. For Japanese infants, it was difficult to discriminate the two forms derived by vowel devoicing, *neek* and *neeks*. However, the number of infants who showed recovery of looking suggested that infants would get better in discrimination of these patterns, even though the number was still fewer than that of English infants. This is the expected pattern for development of discrimination of forms which represent different words in the language. On the other hand, Japanese infants did not show as robust a development in discrimination of *neekusu* and *neeks* as English infants did, although the number of infants who showed recovery at the age of 6 months was almost identical to that in the English group. This, we suggest, is the pattern for development of discrimination of forms that are related phonotactically in the language. Based on this interpretation, we suggest that Japanese infants are beginning to apprehend this relationship by 18 months of age.

ACKNOWLEDGMENTS

The author grateful to Hélène Deacon and Christiane Dietrich for helpful discussion and Tomoko Kawaguchi for her assistance. They also thank all the participants in the experiments.

[1]These exceptions include geminates such as the "tt" in *kitte*, "stamp," and a syllabic nasal that may occur before another consonant. Neither of these patterns are included in the work reported here, and so will not be discussed further.

[2]As for the frequency of vowel devoicing, there is so far no research concerning infant-directed speech. As in teachers' speech to hearing-impaired children, adults may tend to reduce the devoicing rate in speech to infants. We have evidence that Japanese mothers do produce devoiced vowels in infant-directed speech; the research regarding rates of devoicing in Japanese infant-directed speech is ongoing at the time of this writing (Fais et al., 2006).

Bosch, L., and Sebastián-Gallés, N. (**2003**). "Language experience and the perception of a voicing contrast in fricatives: Infant and adult data," in *Proceedings of the 15th International Conference of Phonetic Science*, 1987–1990.

Cohen, L. B., Atkinson, D. J., and Chaput, H. H. (**2000–2002**). "Habit 2002: A new program for obtaining and organizing data in infant perception and cognition studies (version 1.0) [Computer Software]," Austin, TX: The University of Texas.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (**1999**). "Epenthetic vowels in Japanese: A perceptual illusion?" J. Exp. Psychol. **25**, 1568–1578.

Fais, L., Kajikawa, S., Werker, J., and Amano, S. (**2005**). "Japanese listeners' perceptions of phonotacitc violations," Lang Speech **48**, 185–201.

Friederici, A. D. (**2005**). "Neurophysiological markers of early language acquisition: from syllables to sentences," Trends in Cognitive Science **9**, 481–488.

Friederici, A. D., and Wessels, J. M. I. (**1993**). "Phonotactic knowledge of word boundaries and its use in infant speech perception," Percept. Psychophys. **54**, 287–295.

Hallé, P. A., Segui, J., Frauenfelder, U., and Meunier, C. (**1998**). "Processing of illegal consonant clusters: A case of perceptual assimilation?," J. Exp. Psychol. **24**, 592–608.

Imaizumi, S., Fuwa, K., and Hosoi, H. (**1999**). "Development of adaptive phonetic gestures in children: Evidence from vowel devoicing in two different dialects of Japanese," J. Acoust. Soc. Am. **106**, 1033–1044.

Imaizumi, S., Hayashi, A., and Deguchi, T. (**1995**). "Listener adaptive char-

acteristics of vowel devoicing in Japanese dialogue," J. Acoust. Soc. Am. **98**, 768–778.
Jusczyk, P. W., Bauman, A., and Goodman, M. (**1999**). "Sensitivity to sound similarities in different utterances by 9-month-olds," J. Mem. Lang. **40**, 62–82.
Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., and Jusczyk, A. M. (**1993**). "Infants' sensitivity to the sound patterns of native language words," J. Mem. Lang. **32**, 402–420.
Jusczyk, P. W., Luce, P. A., and Charles-Luce, J. (**1994**). "Infants' sensitivity to phonotactic patterns in the native language," J. Mem. Lang. **33**, 630–645.
Kuhl, P. K. (**2004**). "Early language acquisition: Cracking the speech code," Nat. Rev. Neurosci. **5**, 831–843.
Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. (**2006**). "Infants show a facilitation for native language phonetic perception between 6 and 12 months," Dev. Sci. **9**, F1–F9.
Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (**1992**). "Linguistic experience alters phonetic perception in infants by 6 months of age," Science **255**, 606–608.
Mattys, S. L., and Jusczyk, P. W. (**2001**). "Phonotactic cues for segmentation of fluent speech by infants," Cognition **78**, 91–121.
Polka, L., Colantonio, C., and Sundara, M. (**2001**). "A cross-language comparison of /d/-/th/ perception: Evidence for a new developmental pattern," J. Acoust. Soc. Am. **109**, 2190–2201.
Saffran, J. R., Newport, E. L., and Aslin, R. N. (**1996**). "Statistical learning by 8-month-old infants," Science **274**, 1926–1928.
Saffran, J. R., and Thiessen, E. D. (**2003**). "Pattern induction by infant language learners," Dev. Psychol. **39**, 481–494.
Saffran, J. R., Werker, J. F., and Werner, L. A. (**2006**). "The infant's auditory world: Hearing, speech, and the beginnings of language," in *the 6th Edition of the Handbook of Child Psychology*, edited by W. Damon and R. M. Lerner (Wiley, New York), 58–108
Stager, C. L., and Werker, J. F. (**1998**). "Methodological issues in studying the link between speech-perception and word learning," in *Advances in Infancy Research, Volume 12*, edited by C. Rovee-Collier, L. Lipsitt, and H. Hayne (Ablex, Stamford, CT), 237–256.
Werker, J. F., and Tees, R. C. (**1984**). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," Infant Behav. Dev. **7**, 49–63.

Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants

Feng-Ming Tsao[a)]
Department of Psychology, National Taiwan University, Taipei, Taiwan 106 Taiwan, Republic of China

Huei-Mei Liu
Department of Special Education, National Taiwan Normal University, Taiwan, Republic of China

Patricia K. Kuhl
Institute for Learning and Brain Sciences, University of Washington, USA

(Received 19 September 2005; revised 25 July 2006; accepted 26 July 2006)

Previous studies have shown improved sensitivity to native-language contrasts and reduced sensitivity to non-native phonetic contrasts when comparing 6–8 and 10–12-month-old infants. This developmental pattern is interpreted as reflecting the onset of language-specific processing around the first birthday. However, generalization of this finding is limited by the fact that studies have yielded inconsistent results and that insufficient numbers of phonetic contrasts have been tested developmentally; this is especially true for native-language phonetic contrasts. Three experiments assessed the effects of language experience on affricate-fricative contrasts in a cross-language study of English and Mandarin adults and infants. Experiment 1 showed that English-speaking adults score lower than Mandarin-speaking adults on Mandarin alveolo-palatal affricate-fricative discrimination. Experiment 2 examined developmental change in the discrimination of this contrast in English- and Mandarin-leaning infants between 6 and 12 months of age. The results demonstrated that native-language performance significantly improved with age while performance on the non-native contrast decreased. Experiment 3 replicated the perceptual improvement for a native contrast: 6–8 and 10–12-month-old English-learning infants showed a performance increase at the older age. The results add to our knowledge of the developmental patterns of native and non-native phonetic perception. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2338290]

PACS number(s): 43.71.Hw, 43.71.Ft, 43.71.Es, 43.70.Fq [ALF] Pages: 2285–2294

I. INTRODUCTION

The initial state of the infant's mind and the mechanisms for developmental change have stood at the heart of the nature-nurture debate regarding phonetic perception (Kuhl, 2000; 2004; Werker and Curtin, 2005; Best, 1995; Nittrouer, 2001; Aslin *et al.*, 2002). Young infants are able to discriminate phonetic contrasts from both their native (e.g., Eimas, Siqueland, Jusczyk, and Vigorito, 1971) and from foreign languages (e.g., Lasky *et al.*, 1975; Streeter, 1976; Trehub, 1973), while adult listeners generally find non-native discrimination difficult (e.g., Miyawaki *et al.*, 1975; Strange and Jenkins, 1978). Studies suggest that listening to native language speech alters infants' phonetic discrimination to produce a language-specific bias during the first year of life (Best *et al.*, 1995; Kuhl *et al.*, 1992; Werker and Tees, 1984).

Many studies have demonstrated a decline in infants' discrimination of non-native phonetic contrasts between 6 and 12 months of life (Best and McRoberts, 2003; Best, McRoberts, LaFleur, and Silver-Isentadt, 1995; Kuhl *et al.*, 2006; Werker and Tees, 1984). For example, Werker and Tees (1984) tested speech discrimination for non-native contrasts in 6- to 12-month-old infants from English-speaking families with two non-native place contrasts: the Hindi retroflex/dental stops [ʈa] vs [ta] and the velar/uvular glottalized voiceless stops [k̀i] vs [q̀i] in the Thompson language of the Salish Indians. The results demonstrated that infants aged 6–8 months were sensitive to differences between these non-native contrasts but that this perceptual sensitivity to foreign contrasts was significantly reduced in 10–12-month-old infants. However, not all studies on non-native contrasts show a decline between 6 and 12 months of age (Best *et al.*, 1988; Polka *et al.*, 2001).

Fewer tests have examined developmental change for native-language contrasts between 6 and 12 months (Eilers, Wilson, and Moore, 1977; Kuhl *et al.*, 2006; Polka *et al.*, 2001). Kuhl *et al.* (2006), in a recent cross-language study using American English /r-l/ in tests on American and Japanese 6–8 and 10–12-month-old infants, showed that native-language phonetic perception improves between 6 and 12 months, while non-native phonetic perception showed the typically observed decline. Previous studies on native language phonetic perception only suggested a developmental change *during* the first year of life (Eilers *et al.*, 1977) or have shown improvement *after* the first year of life (Polka *et al.*, 2001; Sundara *et al.*, 2006). Studies using a neural measure of discrimination have shown the pattern of facilitation in native language perception and the decline in non-native

[a)]Author to whom correspondence should be addressed; Department of Psychology, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, TAIWAN 106, Phone: +886–2–3366–4467, Fax: +886–2–2363–1463. Electronic mail: tsaosph@mail2000.com.tw

perception in the first year of life—event-related potential studies of phonetic perception show larger amplitude differences for native phonetic discrimination in 10–12 month olds when compared to 6–8-month-old infants (Cheour, 1998; Rivera-Gaxiola *et al.*, 2005), and a decline for non-native discrimination.

It is not clear why there is an age variation in the patterns of developmental change for native and non-native speech. It has been suggested that timing differences of the developmental change in perception could be attributable either to the acoustic fragility of certain phonetic contrasts, such as fricatives, or the frequency of occurrences of the phonetic unit in unstressed contexts, which would also diminish the acoustic cues (Polka *et al.*, 2001; Sundara *et al.*, 2006). Without tests on additional phonetic contrasts, these questions cannot be answered. This study examined infants' perceptual sensitivity to affricate-fricative contrasts, which have not been previously studied, using a cross-language design. Infants in two age ranges, 6–8 months of age and 10–12 months of age, were tested in Taiwan and the United States to investigate the patterns of developmental change for these consonants during the second-half of the first year of life.

Because cross-cultural studies on adults have not been done using these Mandarin contrasts, our studies began with an examination of the acoustic events that signal the affricate-fricative distinction contained in Mandarin, and perceptual studies on Mandarin and English adult speakers. Many studies on adult speakers show that the phonetic contrasts of foreign languages can be difficult to discriminate (Goto, 1971, Miyawaki *et al.*, 1975; Sheldon and Strange, 1982; Trehub, 1976; Werker, Gilbert, Humphery, and Tees, 1981). The classic case is the difficulty that Japanese speakers have in discriminating English /r-l/, a phonemic contrast that is not utilized in Japanese (Miyawaki *et al.*, 1975). In the present study, it is important to note that speakers of both languages (Mandarin and English) utilize an affricate-fricative manner distinction phonemically in their language; however, the distinction in Mandarin utilizes a different place of articulation, and there are subtle acoustic differences in the way the affricate-fricative distinction is realized in the two languages. We therefore hypothesized that adult speakers of the two languages would emphasize different acoustic cues in perception, and that adult speakers of English would therefore find it more difficult to discriminate the Mandarin contrast.

To summarize, this study's goals were twofold: (a) Experiment 1 examined adult perception of affricate-fricative consonants in a cross-language test on Mandarin and English speakers, and (2) Experiments 2 and 3 examined cross-language patterns of developmental change in native (Exp 2 and 3) and non-native (Exp 2) phonetic perception using affricate-fricative consonants during the first year of life. We hypothesized that (a) adult Mandarin speakers would outperform adult English speakers on Mandarin contrasts in Experiment 1, (b) that infants tested in Experiments 2 and 3 on native-language affricate-fricative contrasts would show significant improvement between 6 and 12 months of age, and (c) that infants for whom the affricate-fricative contrast was non-native in Experiment 2 would show a decline between 6 and 12 months of age.

II. EXPERIMENT 1: ENGLISH- AND MANDARIN-SPEAKING ADULTS ON MANDARIN FRICATIVE VS AFFRICATE DISCRIMINATION

A reasonable starting point to explore phonetic discrimination difficulty for non-native contrasts is to identify and test contrasts that are phonemic in one language but not in another language. The alveolo-palatal affricate vs fricative contrasts of Mandarin Chinese, e.g., /tɕ/ vs /ɕ/ and /tɕʰ/ vs /ɕ/, are appropriate for this purpose. First, the manner difference between affricate and fricative is phonemic in English (e.g., /tʃ/ vs /ʃ/ and /dʒ/ vs /ʒ/), but the place of articulation of the Mandarin contrasts, alveolo-palatal, does not occur in English. The English palato-alveolar consonants have a constriction in the vocal tract that is forward of alveolo-palatal sounds (Ladefoged and Maddieson, 1995). This articulation difference results in different acoustic features between Mandarin and English. For example, the spectral peaks of Mandarin alveolo-palatal consonants (around 4900 Hz, Liu, 1996) are located between English palato-alveolar (3800 Hz) and alveolar consonants (6839 Hz) (Jongman *et al.*, 2000).

Second, the exact phonetic features that distinguish affricate and fricative consonants differ in the two languages. Mandarin has three voiceless alveolo-palatal sounds, including two affricates /tɕ/, /tɕʰ/ and one fricative /ɕ/. The phonetic feature of aspiration distinguishes the affricate /tɕ/ [−aspirated] from its counterpart /tɕʰ/ [+aspirated]. In contrast, the English palato-alveolar consonants (e.g., /tʃ/ vs /ʃ/ and /dʒ/ vs /ʒ/) are distinguished with the phonetic features of articulation manner (affricate vs fricative) and voicing (voiced vs voiceless) (Chomsky and Halle, 1968). The phonetic differences between English and Mandarin result in different acoustic correlates in the two languages. Amplitude rise time and frication duration are the two relevant acoustic cues distinguishing the affricate vs fricative contrasts in English (Cutting and Rosner, 1974; Howell and Rosen, 1983; Kluender and Walsh, 1992; Hedrick, 1997). However, frication duration is perceptually more prominent than amplitude rise time for identification of the affricate and fricative (Kluender and Walsh, 1992). Studies on the acoustic correlates of Mandarin affricate and fricative contrasts show that affricates have shorter frication duration and higher occurrence of an initial burst (Liu *et al.*, 2000). However, no study has measured the amplitude rise time of this Mandarin phonetic contrast. In addition, it is unclear whether frication duration and amplitude rise time are both perceptually relevant for differentiating affricates and fricatives in Mandarin Chinese. In brief, both the phonetic and acoustic differences between Mandarin and English provide a reasonable basis for hypothesized performance differences between English and Mandarin adult speakers on the perception of Mandarin alveolo-palatal contrasts.

Experiment 1 tested the hypothesis that English adult speakers will show significantly poorer performance when compared to adult Mandarin speakers on the discrimination of Mandarin alveolo-palatal affricate and fricative contrasts.

TABLE I. Acoustic features of Mandarin affricate versus pricative contrasts in experiment 1.

Phonetic contrasts	Phonetic feature differences	Stimulus pairs	Amplitude rise time (+3 dB SPL from start in ms)	Frication noise duration (in ms)
Set 1:	Aspirated	1	30 vs 100	Same (130 ms)
/tɕʰ/ vs /ɕ/	Affricate vs	2	30 vs 60	Same (130 ms)
	Fricative	3	30 vs 60	Same (160 ms)
Set 2:	Aspirated	4	Same (30 ms)	130 vs 80
/tɕʰ/ vs /ɕ/	Affricate vs	5	Same (30 ms)	100 vs 50
	Unaspirated	6	(30 ms)	100 vs 80
Set 3:	Unaspirated	7	Same (50 ms)	55 vs 95
/tɕ/ vs /ɕ/	Affricate vs	8	Same (50 ms)	55 vs 85
	Fricative	9	(50 ms)	65 vs 95

Performance differences between adult speakers of the two languages were also expected to guide the selection of the specific speech stimuli that were used in Experiment 2.

A. Method

1. Participants

Thirty-six undergraduate and graduate students of the University of Washington without history of severe language or hearing impairments participated in this study. One group consisted of 18 native English speakers (ten females, eight males). Another group included 18 native Mandarin Chinese speakers (nine females, nine males) from either Taiwan or China. All participants were tested in Seattle. English speakers were recruited from the Psychology Subject Pool and received class credits for their participation. Mandarin speakers were recruited from a solicitation on web pages of international student organizations and received $20 for their participation.

2. Stimuli

Nine synthesized speech pairs were created for /tɕi/, /tɕʰi/, and /ɕi/ tokens using the HLsyn speech synthesizer 2.2 (1996); they were sampled at 11025 Hz. These pairs tested three sets of phonetic contrasts: /tɕi/ vs /tɕʰi/; /tɕi/ vs /ɕi/; and /tɕʰi/ vs /ɕi/. The synthetic stimuli varied in either amplitude rise time or frication noise duration. The acoustic differences for each stimulus pair are shown in Table I. The values of frication duration in each pair were chosen based on an acoustic analysis of Mandarin alveolo-palatal affricate and fricative consonants (Liu, 1996). However, no acoustic data are available for amplitude rise time in Mandarin affricate and fricative consonants. For amplitude rise time in the aspirated affricate /tɕʰ/ vs fricative /ɕ/ pair, perceptual studies in English (Cutting and Rosner, 1974; Howell and Rosen, 1983) suggested that phonetic boundaries between palato-alveolar affricate and fricative pairs were located around 50 ms. Therefore, the values of amplitude rise time and frication duration were varied across phonetic boundaries to generate separate phonetic categories for Mandarin-speaking adults. Amplitude rise time is defined as the time to reach the maximum amplitude (+3 dB SPL) of frication noise from the onset of the syllable.

The spectral peak of frication noise is around 4700 Hz, and the frication noise energy mainly occurs above 2500 Hz, typical values for Mandarin speakers (Liu, 1996). During the 245 ms vowel portion of the syllables, the formant frequencies were: 293, 2274, 3186, and 3755 Hz, respectively, for F1 through F4. The bandwidths of F1–F4 were 80, 90, 150, and 350 Hz, respectively. The fundamental frequency (pitch) of the syllable was 120 Hz, a typical value for male Mandarin speakers (Huang, 1996). Finally, tokens were equalized in rms amplitude. The appropriateness of these speech stimuli were judged by native Mandarin speakers from Taiwan and China in a pilot study.

3. Procedures

A computer presented pairs of these stimuli in an AX discrimination task. On each trial, the participant heard two tokens separated by 350 ms through earphones at a comfortable listening level of approximately 65 dBA in a sound-attenuated booth. Participants were asked to decide whether the pair of stimuli were the same or different. The probability of stimulus presentation order, i.e., A-A, A-B, B-A, and B-B, was equal and stimulus pairs were randomly presented across subjects. Prior to the test, participants practiced with synthetic stop-vowel (/ba/ vs /pa/) stimuli that were easy for both language groups and received feedback on their responses. During the test stage, 180 test pairs (=9 pairs $\times$4 presentation orders$\times$5 repetitions) were presented with no feedback. Every subject completed this experiment in 30–40 min.

B. Results and discussion

A bias-free measure of sensitivity (d') was calculated for the two language groups for each stimulus pair and the values are illustrated in Fig. 1. As shown, the Mandarin group shows better discrimination across all pairs (average $d'=1.74$, percent correct=77.62%) when compared with the

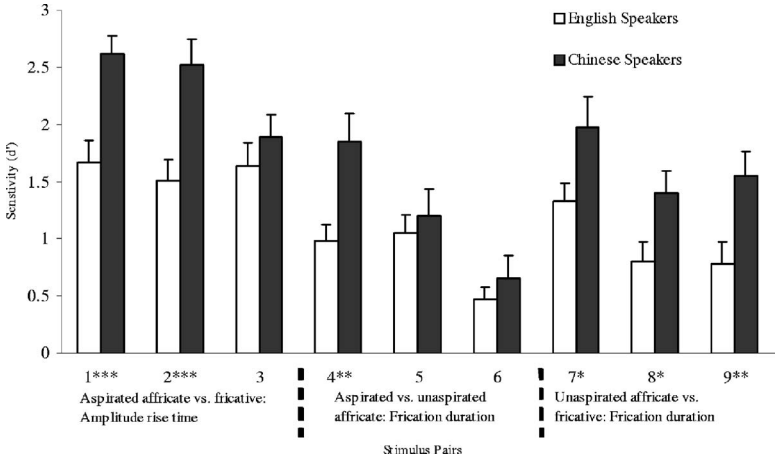


FIG. 1. Sensitivity (d') in discriminating Mandarin contrasts (*SE* in error bars, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$).

American English group (average $d'=1.13$, percent correct $=67.72\%$). A mixed-design two-way ANOVA (between subject factor: Language group $\times$ within subject factor: Stimulus pair) of d' reveals that both language $[F(1,34)=17.132]$ and stimulus factors $[F(8,272)=17.377]$ are significant at $p<0.001$. The nonsignificant interaction $[F(8,272)=1.842, p>0.1]$ of these two factors reveals that Mandarin speakers consistently performed better than American English speakers in discriminating these Mandarin contrasts.

The results of Experiment 1 indicate that amplitude rise time and frication duration contribute significantly to Mandarin speakers' discrimination of the three sets of phonetic contrasts. Amplitude rise time is effective for discriminating the aspirated affricate /tɕʰ/ vs fricative /ɕ/ contrast, $F(2,34)=5.445, p<0.01$. Frication duration is utilized by Mandarin speakers to distinguish the aspirated /tɕʰ/ vs unaspirated /tɕ/ affricate, $F(2,34)=14.563, p<0.001$, and unaspirated affricate /tɕ/ vs fricative /ɕ/ contrasts, $F(2,34)=3.478, p<0.05$.

An interesting question is the relative contribution of amplitude rise time and frication duration in the discrimination of Mandarin contrasts. Studies have shown that both rise time and frication duration covary in the English affricate and fricative distinction although frication duration is perceptually more effective (Kluender and Walsh, 1992). A different weighting of acoustic cues to the affricate and fricative distinction between English and Mandarin Chinese might be one of the key factors in the performance difference between the two language groups. This experiment examined the performance differences between the two language groups by utilizing only a small set of acoustic values. Further studies that covary rise time and frication duration will be needed to investigate the relative contribution of both acoustic cues in the discrimination of Mandarin affricative vs fricative consonants.

To summarize, the results of Experiment 1 show that English speakers experience more difficulty overall when compared to Mandarin speakers in the discrimination of Mandarin alveolo-palatal fricative vs affricate contrasts, indicating that language experience affects perceptual discrimination for non-native contrasts. In addition, the results suggest that amplitude rise time and frication duration are important for discriminating Mandarin alveolo-palatal contrasts. These data were used to guide selection of a contrast to use in tests on Mandarin- and English-learning infants in Experiment 2.

III. EXPERIMENT 2: TESTING MANDARIN- AND ENGLISH-LEARNING INFANTS ON THE PERCEPTION OF A MANDARIN AFFRICATE-FRICATIVE CONTRAST

The goal of this experiment was to explore the time course of developing native and non-native phonetic perception using a cross-language design. In Experiment 1, English-speaking adults performed significantly poorer than Mandarin-speaking adults in the discrimination of Mandarin Chinese alveolo-palatal fricative vs affricate tokens. The largest difference between English- and Mandarin-speaking adults was shown in the aspirated-affricate vs fricative contrast when the amplitude rise time was varied, and so this contrast was utilized in tests on infants.

A. Method

1. Participants

The participants were 69 infants, 37 American and 32 Taiwanese. Of the 37 American infants, 19 were in the age range of 6–8 months (mean age at test=7.2 months; range =6.8–7.7 months; boys=nine, girls=ten), and 18 were in the age range of 10–12 months (mean age at test=10.9 months; range=10.8–11.1 months; boys=nine, girls=nine). Of 32 Taiwanese babies, half of them were 6–8 months (mean age at test=7.4 months; range=6.9–8.4 months; boys=eight, girls=eight), and half were in the age range of 10–12 months (mean age at test=11.3 months; range =10.7–12.2 months; boys=11, girls=five). An additional 19 infants failed to complete testing due to an inability to pass the training (17), an equipment failure (1), or a failure to return for all of the required sessions (1). Infants who failed to pass the training did not differ by age or language: five 6–8-month-old American infants (drop-out rate=20.8%), four 6–8-month-old Taiwanese infants (drop-out rate =20.0%), two 10–12-month-old American infants (drop-out rate=10.0%), and six 10–12-month-old Taiwanese infants (drop-out rate=27.3%) failed to meet the criterion. Results of Fisher's exact probability test on the drop-out rate indi-

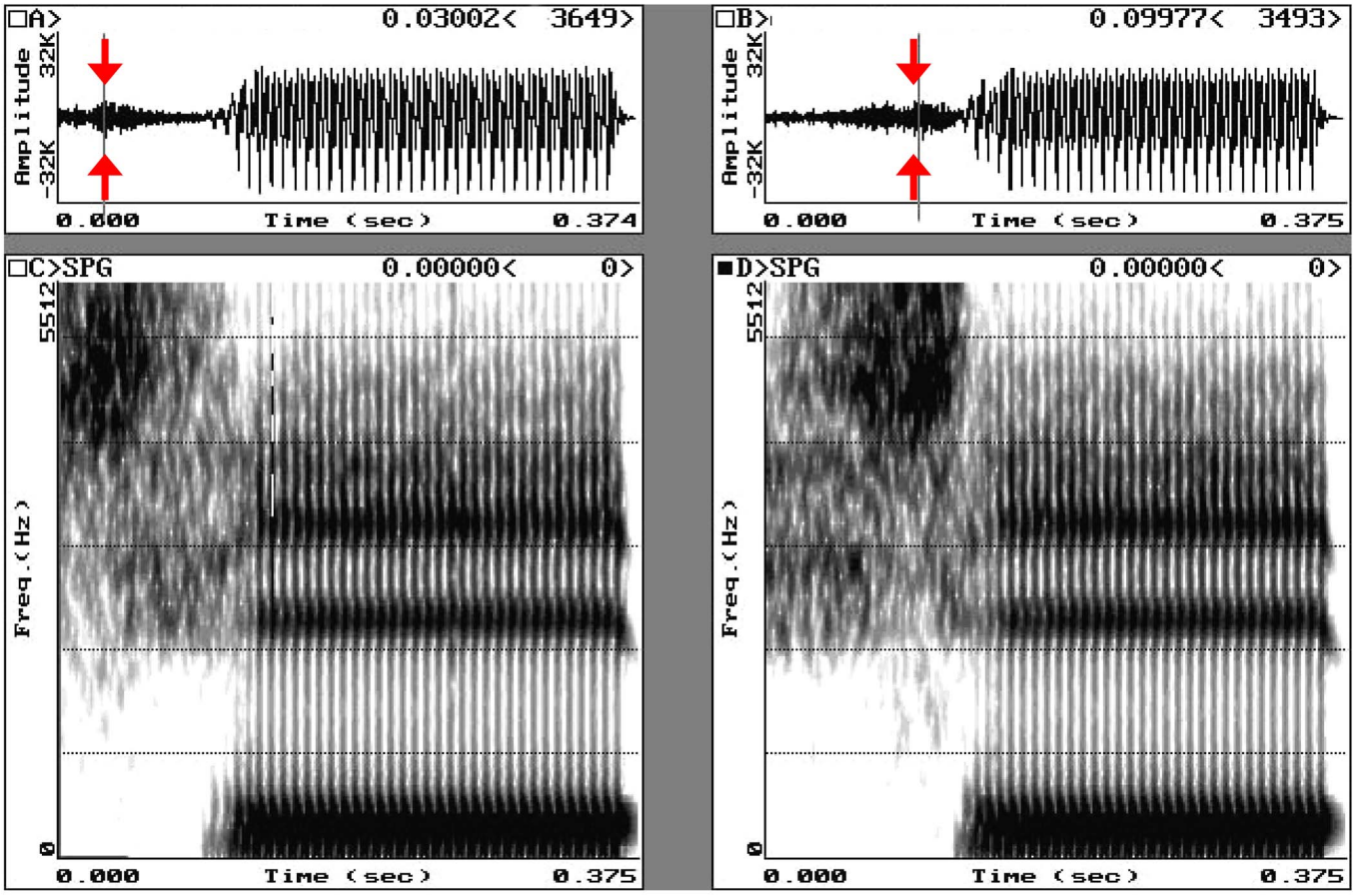


FIG. 2. (Color online) Wave forms (top panel) and spectrograms (bottom panel) of affricate (left) and fricative (right) stimuli in Exp. 2. Arrows indicate the location of the maximum amplitude in frication noise.

cated neither the age nor language effect reached significance. Pre-established criteria for inclusion in the study were that infants had no known visual or auditory deficits, were full term (born ± 14 days from due date), had uncomplicated deliveries, were normal birth weight (6–10 lbs), were developing normally, and that members of their immediate families had no history of hearing loss. Parents were paid $30 for completing the experiment.

American infants were recruited through the database of names contained in the Infant Studies Subject Pool (ISSP) at the University of Washington. Taiwanese infants were recruited either through listings of names on the House Registry of the Lin-Ya Area, Kaohsuing City, Taiwan, or from notices soliciting families' participation which were displayed at the Kaohsiung Chung-Kung Children's Hospital. Although Taiwan is a multi-lingual society, Mandarin Chinese is the most dominant language in homes. A Mandarin-dominant (or only) language environment was verified for the Taiwanese infants through a language background questionnaire in Chinese that was administered to the caregiver before the study began. The criteria of judging Mandarin the dominant language were: (1) parents and infant's caretakers used Mandarin when addressing their infants, and (2) infants heard Mandarin since birth.

2. Stimuli

The stimuli were pair 1 from Experiment 1, which consisted of computer synthesized male tokens of Mandarin Chinese alveolo-palatal /tɕʰi/ and /ɕi/ syllables. Figure 2 illustrates wave forms and spectrograms for these stimuli (see Table I for acoustic differences). Tokens were played to infants at a comfortable listening level of approximately 65 dBA. Mandarin-speaking adults (n=14) tested in the same conditioned head-turn procedure as infants discriminated this contrast at 95.21% (SD=5.84).

3. Apparatus

Stimuli were presented using a digital signal processor (CAC Bullet III) controlled by a portable computer (Dell Inspiron 3500). The computer was also used to record infant head-turn responses. The sounds were reproduced with 11.025 k 16 bit samples per second, and were low-pass filtered at a 5.5 kHz cutoff frequency. Stimuli were amplified (Shure FP42) and delivered to subjects in an adjoining sound-treated test room via a loudspeaker (Boston Acoustics CR7). Parents and experimenters wore headphones (David Clark H3050) and listened to masking music during the tests so they could not distinguish between the stimuli presented to infants. Infants' responses were monitored in the control room via use of a closed-circuit camera (RCA TC7011) and a video monitor. Both American and Taiwanese infants were tested with the same apparatus, experimenters, and test protocols to carefully control the experimental task between the two different countries.

4. Test suite

The test suite consisted of two rooms. In the test room, an infant was held on its parent's lap, facing forward while the assistant was seated at a 60° angle to the infant's right side. An Assistant maintained the infant's attention by manipulating a series of engaging, silent toys to bring the child's gaze to midline (straight ahead of the infant). A bank of two visual reinforcers, located at a 60° angle to the infant's left side, each consisted of a dark Plexiglas box (13 in. × 13 in. × 13 in.) containing a commercially available mechanical toy (e.g., a bear pounding a drum). The toys were not visible until they were activated and lights mounted inside the box were illuminated. The visual reinforcers were placed on either side of the loudspeaker, and were at eye level for the infant. A camera, located in front of the infant, but hidden from view by a curtain with a hole cut for the lens, fed an image of the test room to the adjoining control room, where the Experimenter observed the infant's behavior. In all phases of training and testing, trials are initiated by the Assistant. The Experimenter, who cannot hear the stimuli presented during trials, and who is unaware of the type of trial selected automatically by the compute, indicates infants' head turns by pressing a computer key.

5. Procedure

The conditioned head-turn technique was used to assess infants' discrimination abilities (Kuhl, 1985; Werker *et al.*, 1997). The "background" speech sound, /ɕi/, was repeated once every 2 s (Interstimulus interval, ISI=1625 ms). Infants first were trained to produce a head turn for visual reinforcement whenever the background speech sound was changed to the "target" speech sound, /tɕʰi/. Only one direction of stimulus change was tested because potential directional effects in infant testing (e.g., Kuhl *et al.*, 2006; Polka and Bohn, 1996, 2003; Polka *et al.*, 2001) might complicate the developmental pattern. The experimental protocol required a two-step Conditioning Phase: (1) an intensity cue was initially added to assist infants in detecting the sound change and (2) the intensity cue was eliminated to establish infants' abilities to discriminate the contrast in the absence of a loudness difference. During Conditioning, all trials involve a change in the stimulus from the background to the target stimulus (Change Trials). After this two-step Conditioning Phase, a Test Phase was initiated; during the Test Phase, an equal number of Change and Control trials are run in random order. All phases of the experiment were under computer control. The same basic procedure has been used in previous infant studies in this laboratory (Liu, Kuhl, and Tsao, 2003; Kuhl, Tsao, and Liu, 2003; Tsao, Liu, and Kuhl, 2004).

During the first step of Conditioning, infants were trained to associate presentation of the target speech sound with the activation of the visual reinforcers. The Assistant initiated a trial when infants appear ready (focused on the toys held by the assistant). Then, the target sound interrupted the repetitive presentation of the background speech sound, and was presented at a level 4 dBA higher than the background speech sound. The target stimulus was presented three times, and infants had to produce a head turn in response to the sound change within this 6 s period. The louder target speech sound facilitated infant learning to turn their heads and associate the sound change with the presentation of the reinforcer. When the infant produced a head turn on two consecutive trials, the infant proceeded to the second Conditioning phase, during which the target sound is presented at the same level as the background sound; infants can only use the phonetic difference between sounds as a cue. Infants must produce three consecutive head turns within 30 training trials to advance to the Test Phase. During the Conditioning phase, infants were cued to produce a head turn by the activation of the reinforcers near the end of the trial if they had not produced a head turn.

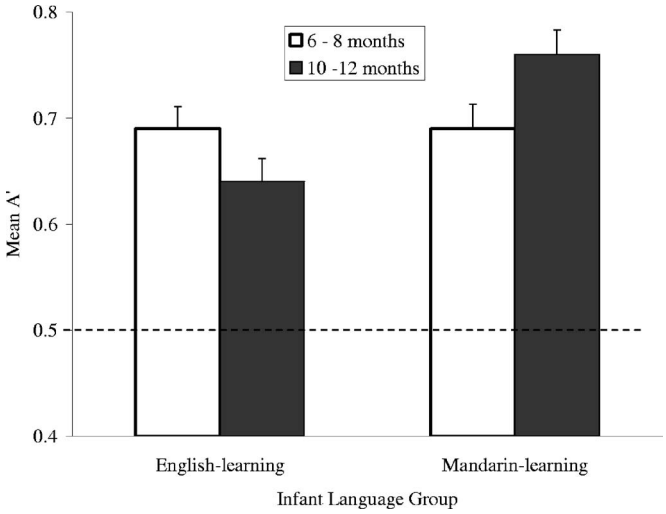


FIG. 3. Phonetic discrimination of 6–8 and 10–12-month-old English- and Mandarin-learning infants on Mandarin contrast (dotted line shows the chance level in Exp 2; *SE* in error bars).

The Test Phase consisted of 30 trials, an equal number of Change and Control (no-change) Trials, presented in random order. Infants were tested in 20 min sessions on consecutive days, when possible, but all completed testing within one week. Most infants were tested in two days (Conditioning on Day 1 and then test on Day 2), but up to three sessions were allowed to complete the test. When infants returned on the second day for testing, three Conditioning trials (+ intensity cue) were used to refresh the infants on the experimental procedure before the test trails began. Infants who failed to pass the two-phase Conditioning in two sessions were eliminated from the experiment.

B. Results and Discussion

The results provided information about the time course of development of native- and non-native phonetic perception for affricate-fricative consonants. Figure 3 provides the A′ scores as a function of age and language group for the Taiwanese and American infants. A′ is a distribution-free sensitivity measure that takes both "hits" and "false alarms" into consideration to provide an estimate of an infant's accuracy in detecting the sound change. This measure (range=0 −1, chance level=0.5) is similar to d' (Grier, 1971) and has been used in infant speech discrimination studies (e.g., Anderson, Morgan, and White, 2003; Polka *et al.*, 2001).

As shown in Fig. 3, increasing age affects discrimination capacity very differently in the two language groups. Mandarin-learning infants show an increase in performance over time while English-learning infants show a decline. The difference in the effect of age on discrimination performance in the two groups was tested by analysis of variance. Taiwanese infants demonstrated a substantial improvement in their performance on the discrimination task, 6–8 month olds ($M=0.69$, $SD=0.08$) and 10–12 month olds ($M=0.76$, $SD=0.08$), $F(1,30)=5.41$, $p=0.027$. In contrast, American infants showed a trend toward declining sensitivity, younger infants ($M=0.69$, $SD=0.10$) and older infants ($M=0.64$, $SD=0.10$), $F(1,35)=3.09$, $p=0.088$. Performance of the two language groups at the younger age was nonsignificant, $F(1,33)<1$, and both above chance level at $p<0.001$, one-sample t test, Taiwanese infants, $t(15)=9.81$; American infants, $t(18)=8.37$. In contrast, the performance of older infants was significantly different for the two language groups, $F(1,32)=14.33$, $p<0.001$; performance of both older infant groups is significantly above chance level at $p<0.001$, one-sample t test, Taiwanese infants, $t(15)=12.41$; American infants, $t(17)=6.21$. A two-way ANOVA (Language group X Age) on the A′ was conducted. The main effect of age was not significant, $F(1,65)<1$, though that of language did reach significance, $F(1,65)=6.28$, $p=0.015$. The age x language interaction was highly significant, $F(1,65)=7.88$, $p=0.007$, and indicated a divergent trend for developing native and non-native language perception during the first year of life.

The results show a divergent trend in the development of native and non-native phonetic perception. Infants listening to a native language contrast show a significant improvement between 6 and 12 months, as observed recently by Kuhl *et al.* (2006), while those listening to a non-native contrast show a pattern of decline in the ability to discern differences between speech sounds of the non-native language, though one that does not reach significance. The pattern of facilitation in the first year, seen for /r-l/ in American infants (Kuhl *et al.*, 2006) and in the present study for affricate-fricative contrasts in Taiwanese infants, suggests that experience with native language leads infants to develop increased sensitivity to native contrasts.

IV. EXPERIMENT 3: PHONETIC DISCRIMINATION OF ENGLISH-LEARNING INFANTS ON THE ENGLISH AFFRICATE VS FRICATIVE CONTRAST

Experiment 2 demonstrated that the Mandarin affricate-fricative contrast shows an increase between 6–8 and 10–12 months for Mandarin-learning infants. Experiment 3 was designed to extend these findings to American infants listening to their native-language affricate-fricative (palato-alveolar) contrast.

A. Method

1. Participants

The participants were 17 American infants aged 6–8 months (mean age at test=6.9 months; range =6.8–7.7 months; boys=ten, girls=seven) and 19 American infants aged 10–12 months (mean age at test=10.7 months; range=10.5–11.1 months; boys=ten, girls=nine). Infants were recruited through the database of the Infant Studies Subject Pool (ISSP) at the University of Washington. An additional 17 infants failed to complete testing due to inability to pass the Conditioning Phase (11), or failure to return for all of the required sessions (six). Of the 11 infants who failed to pass Conditioning, the drop-out rate was not significantly different between age groups, younger infants ($n=5$, drop-out rate=22.7%) and older infants ($n=6$, drop-out rate=24.0%), Fisher's exact test, $p>0.1$. The inclusion and exclusion criteria of subject selection were the same as in Experiment 2. Parents were paid $30 when their infants completed the experiment.

2. Stimuli

The stimuli were computer synthesized tokens of English palato-alveolar affricate /tʃi/ and fractive /ʃi/ syllables created using a male voice. They were matched in all acoustic details other than the temporal features during the initial portion of the consonants. Previous studies showed that frication duration is the primary acoustic cue for the affricate-fricative distinction in English (Kluender and Walsh, 1992).

Frication duration of the affricate and fricative tokens were 80 and 180 ms, respectively. The amplitude rise time was 30 ms shorter than the frication duration to generate more natural-sounding speech tokens for English speakers. Therefore, both amplitude rise time and frication duration differed in the two tokens. The spectral peak frequency of this pair of tokens was 2800 Hz. Acoustic parameters of the vowel /i/ were exactly the same as in the previous experiments. The duration of the vowel was 245 ms. The two stimuli were judged to be good instances of English native categories and were easily discriminated by English-speaking adults in a pilot study. Tokens were equalized in rms amplitude and were played to infants at a comfortable listening level of approximately 65 dBA.

3. Apparatus and Procedure

The procedure and apparatus were identical to that used to test perceptual development of infants' speech discrimination on native and non-native contrasts in Experiment 2. The background syllable was fricative /ʃi/ and the target syllable was affricate /tʃi/.

B. Results and Discussion

Experiment 3 tested two specific predictions. First, based on the results of Experiment 2, English-learning infants' performance on the native palato-alveolar affricate-fricative distinction was hypothesized to show an increase with age. Second, performance of 11-month-old English-learning infants was expected to exceed that shown for the nonnative affricate-fricative contrast tested in Experiment 2.

The results of this experiment support these two predictions. Figure 4 illustrates the A′ scores of the two English-learning groups on the discrimination of English /tʃ/ vs /ʃ/. Older English-learning infants were more sensitive ($M=0.78$, $SD=0.05$) than younger infants ($M=0.70$, $SD=0.11$)

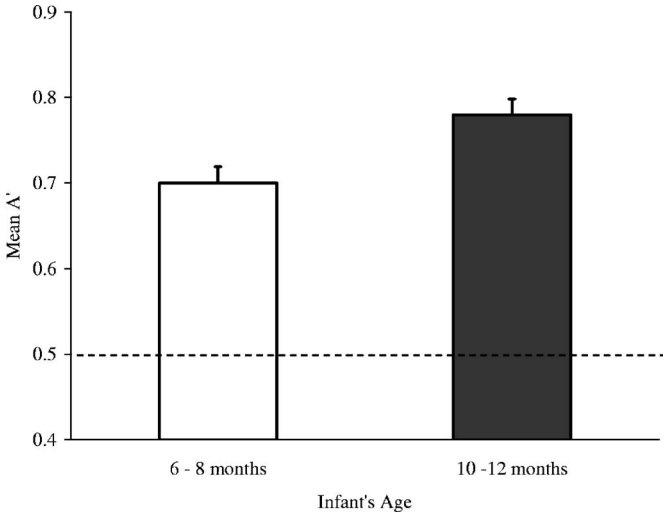


FIG. 4. Phonetic discrimination of 6–8 and 10–12-month-old English-learning infants on English contrast (dotted line shows the chance level in Exp 3; *SE* in error bars).

to the differences in the two phonemes, $F(1,34)=6.39$, $p=0.016$. In addition, English-learning 11-month-old infants were more accurate in detecting the acoustic differences between the native affricate-fricative contrast than they were in detecting the affricate-fricative non-native contrast in Experiment 2, $F(1,35)=30.35$, $p<0.001$. The younger English-learning infants performed similarly for both English and Mandarin contrasts, $F(1,34)<1$. Combining the results of young and old English-learning infants on the discrimination of both native (Experiment 3, English palato-alveolar affricate-fricative) and non-native (Experiment 2, Mandarin alveolo-palatal affricate-fricative) contrasts, a two-way ANOVA (Age × Phonetic contrast) reveals a significant contrast effect, $F(1,69)=11.64$, $p<0.001$, and nonsignificant age effect, $F(1,69)<1$. The interaction of Age and Phonetic contrast is significant, $F(1,69)=8.67$, $p=0.004$. Therefore, the results demonstrate the divergent trend of perceptual development for native and non-native contrasts. In addition, 11-month-old English-learning infants performed at the same level on the discrimination of native affricate vs fricative contrast as Mandarin-learning infants at the same age discriminating their native affricate-fricative contrast. No significant difference was evident in a one-way ANOVA on infant language group, $F(1,33)<1$. This suggests that the developmental pace of perceiving affricate and fricative contrasts is similar for infants raised in two very different language environments.

V. GENERAL DISCUSSION

The effects of language experience on developmental change in speech perception were examined using a cross-language design with both adults and infants on the perception of native and non-native speech sounds. The results of Experiment 1 demonstrate the impact of language experience on phonetic discrimination in adults. Mandarin-speaking adults are more accurate than English-speaking adults in distinguishing the acoustic differences between Mandarin alveolo palatal fricative and affricate consonants.

English and Mandarin-learning infants were compared in Experiment 2 to examine sensitivity change to native and non-native contrasts during the second half of the first year of life using a Mandarin affricate vs fricative contrast from Experiment 1. The results of Experiment 2 revealed a divergent trend in perceptual development for native and non-native contrasts and are consistent with the view that infants develop language-specific processing around their first birthday. The results of Experiment 2 provide support for the idea that native contrasts show a pattern of facilitation over time; perceptual sensitivity of 10–12-month-old infants improved over that seen in 6–8-month-old infants. This improvement for native contrasts is not mirrored in performance on non-native contrasts. Older infants perform less accurately than younger infants on the discrimination of a non-native contrast. Furthermore, Experiment 3 buttresses these findings on facilitation for native-language contrasts by showing that 10–12-month-old English-learning infants are significantly more sensitive than 6–8-month-old English-learning infants in discriminating their native English affricate-fricative distinction. This pattern of facilitation for native-language phonetic learning in the first year has been suggested by neural studies (Cheour, 1998; Rivera-Gaxiola *et al.*, 2005), and clearly shown in a behavioral test for the American English /r-l/ contrast (Kuhl *et al.*, 2006). Thus the facilitation pattern is seen for liquids (Kuhl *et al.*, 2006) and the two affricate-fricatives tested in the present experiments; a third contrast shows facilitation, though later in development (Polka *et al.*, 2001; Sundara *et al.*, 2006). One study observed a decline in a difficult fricative (/s-z/) native-language contrast (Best and McRoberts, 2003).

What accounts for the variance across phonetic contrasts in the pattern of facilitation or decline seen developmentally? The timing of developmental change could be due to the amount of experience with specific phonetic units in the native language. The fact that the frequency of occurrence of native consonants is not equally distributed in language input to infants suggests that different patterns of development may exist for different consonants. It is estimated that the coronal stops are more frequent than the dorsal stops in the English infant-directed speech (Anderson *et al.*, 2003). For non-native perception, one recent study demonstrated that 8.5-month-old English-learning infants performed less accurately distinguishing the non-native coronal stops than the dorsal stops (Anderson *et al.*, 2003). Infants utilize the distributional probabilities of phonetic features in the native language to perceive speech sounds. For example, Maye, Werker, and Gerken (2002) examined the impact of distributional properties of speech sounds in infants by varying these properties in 2 min exposures to a series of eight stimuli from a voice-onset time continuum; infants' discrimination abilities were improved by "bimodally" distributed experience (see Maye and Weiss, 2003, for discussion). Kuhl *et al.* (1992) examined the impact of distributional properties by testing perception in 6-month-old infants from two countries whose experience with natural language provided them with vowels whose distributional properties differed in language input; infants' vowel categorization abilities were enhanced for native-language vowel categories (see Kuhl, 2004 for dis-

cussion). These studies suggest a statistical basis of phonetic learning and predict that listening experience with native language enhances the perception of native-language phonetic categories while reducing sensitivity to non-native contrasts. If infants are sensitive to distributional properties of phonetic categories when listening to native language, as suggested by previous research, the possibility exists that the frequency of consonants and their distribution in ambient speech, as well as their acoustic properties, may be shown to affect perceptual development of different native contrasts.

The results of infant experiments and previous studies clearly demonstrate that infants are born with language-general processing abilities (e.g., Best and McRoberts, 2003; Kuhl, 2004; Werker and Tees, 1984), and the results of this study further demonstrate the divergence in development of the perception of native and non-native phonetic contrasts during the first year of life. What is needed is a hypothesis that explains all the patterns seen to date in developmental phonetic perception data. What determines when facilitation for native contrasts will occur; and what determines which non-native contrasts will decline and which do not? Various hypotheses have been developed: Best and McRoberts (2003) hypothesize knowledge of articulatory organs, Maye, Werker and Gerken (2002) argue that a distributional frequency hypothesis accounts for the data, and Kuhl and her colleagues argue that a combination of motherese and distributional frequency accounts for native-language learning with social and cognitive factors playing a critical role (Liu *et al.*, 2003; Kuhl *et al.*, 2003; Kuhl *et al.*, in press). Additional contrasts will need to be tested to examine why some contrasts show developmental change prior to 12 months, while others do not; such studies will also allow comparisons among theories.

Anderson, J. L., Morgan, J. L., and White, K. S. (**2003**). "A statistical basis for speech sound discrimination," Lang Speech **46**, 155–182.

Aslin, R. N., Werker, J. F., and Morgan, J. L. (**2002**). "Innate phonetic boundaries revisited (L)," J. Acoust. Soc. Am. **112**, 1257–1260.

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Best, C. T., and McRoberts, G. W. (**2003**). "Infant perception of non-native consonant contrasts that adults assimilate in different ways," Lang Speech **46**, 183–216.

Best, C. T., McRoberts, G. W., and Sithole, N. N. (**1988**). "The phonological basis of perceptual loss for nonnative contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants," J. Exp. Psychol. Hum. Percept. Perform. **14**, 345–360.

Best, C. T., McRoberts, G. W., LaFleur, R., and Silver-Isenstadt, J. (**1995**). "Divergent developmental patterns for infants' perception of two nonnative consonant contrasts," Infant Behav. Dev. **18**, 339–350.

Cheour, M. (**1998**). "Development of language-specific phoneme representations in the infant brain," Nat. Neurosci. **1**, 351–353.

Chomsky, N., and Halle, M. (**1968**). *Sound Pattern of English* (Harper and Row, New York).

Cutting, J. E., and Rosner, B. S. (**1974**). "Categories and boundaries in speech and music," Percept. Psychophys. **16**, 564–570.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (**1971**). "Speech perception in infants," Science **171**, 303–306.

Eilers, R. E., Wilson, W. R., and Moore, J. M. (**1977**). "Developmental changes in speech discrimination in infants," J. Speech Hear. Res. **20**, 766–780.

Goto, H. (**1971**). "Auditory perception by normal Japanese adults of the sounds "L" and "R"," Neuropsychologia **9**, 317–323.

Grier, J. B. (**1971**). "Nonparametric indexes for sensitivity and bias: Computing formulas," Psychol. Bull. **75**, 424–429.

Hedrick, M. (**1997**). "Effect of acoustic cues on labeling fricatives and affricates," J. Speech Lang. Hear. Res. **40**, 925–938.

Howell, P., and Rosen, S. (**1983**). "Perception of rise time and explanations of the affricate/fricative contrast," Speech Commun. **2**, 164–166.

Huang, K. I. (**1996**). "Articulation, acoustic features, and perception of vowels," in *Fundamentals of Speech Pathology Vol. 2*, edited by C. H. Tseng (Psychology, Taipei, Taiwan), pp. 1–31.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Kluender, K. R., and Walsh, M. A. (**1992**). "Amplitude rise time and the perception of the voiceless affricate/fricative distinction," Percept. Psychophys. **51**, 328–333.

Kuhl, P. K. (**1985**). "Methods in the study of infant speech perception," in *Measurement of Audition and Vision in the First Year of Life: A Methodological Overview*, edited by G. Gottlieb and N. A. Krasnegor (Ablex, Norwood, NJ), pp. 223–251.

Kuhl, P. K. (**2004**). "Early language acquisition: Cracking the speech code," Nat. Rev. Neurosci. **5**, 831–843.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (**in press**). "Developmental phonetic perception: Native language magnet theory expanded (NLM-e)," Philos. Trans. R. Soc. London, Ser. B

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. (**2006**). "Infants show a facilitation effect for native language phonetic perception between 6 and 12 months," Dev. Sci. **9**, F13–F21.

Kuhl, P. K., Tsao, F. M., and Liu, H. M. (**2003**). "Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning," Proc. Natl. Acad. Sci. U.S.A. **100**, 9096–9101.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (**1992**). "Linguistic experience alters phonetic perception in infants by 6 months of age," Science **255**, 606–608.

Ladefoged, P., and Maddieson, I. (**1995**). *The Sounds of the World's Languages* (Blackwell, Malden, MA).

Lasky, R. E., Syrdal-Lasky, A., and Klein, R. E. (**1975**). "VOT discrimination by four to six and a half month old infants from Spanish environments," J. Exp. Child Psychol. **20**, 215–225.

Liu, H. M., Kuhl, P. K., and Tsao, F. M. (**2003**). "An association between mothers' speech clarity and infants' speech discrimination skills," Dev. Sci. **6**, F1–F10.

Liu, H. M., Tseng, C. H., and Tsao, F. M. (**2000**). "Perceptual and acoustic analysis of speech intelligibility in Mandarin-speaking young adults with cerebral palsy," Clin. Linguist. Phon. **14**, 447–464.

Liu, H. M. (**1996**). "Perceptual and acoustic analysis of speech intelligibility in young adults with cerebral palsy," Unpublished Master's Thesis, National Kaoshuing Normal University, Taiwan.

Maye, J., and Weiss, D. J. (**2003**). "Statistical cues facilitate infants' discrimination of difficult phonetic contrasts," *Proceedings of the 27th Annual Boston University Conference on Language Development*, pp. 508–518.

Maye, J., Werker, J. F., and Gerken, L. (**2002**). "Infant sensitivity to distributional information can affect phonetic discrimination," Cognition **82**, B101–B111.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., and Fujimura, O. (**1975**). "An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English," Percept. Psychophys. **18**, 331–340.

Nittrouer, S. (**2001**). "Challenging the notion of innate phonetic boundaries," J. Acoust. Soc. Am. **110**, 1598–1605.

Polka, L., and Bohn, O. S. (**1996**). "Cross-language comparison of vowel perception in English-learning and German-learning infants," J. Acoust. Soc. Am. **100**, 577–592.

Polka, L., and Bohn, O. S. (**2003**). "Asymmetries in vowel perception," Speech Commun. **41**, 221–231.

Polka, L., Colantonio, C., and Sundara, M. (**2001**). "A cross-language comparison of /d/-/ð/ perception: Evidence for a new developmental pattern," J. Acoust. Soc. Am. **109**, 2190–2201.

Rivera-Gaxiola, M., Silva-Pereyra, J., and Kuhl, P. K. (**2005**). "Brain potentials to native- and non-native speech contrasts in seven and eleven-month-old American infants," Dev. Sci. **8**, 167–172.

Sheldon, A., and Strange, W. (**1982**). "The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception," Appl. Psycholinguist. **3**, 234–261.

Strange, W., and Jenkins, J. (**1978**). "Role of linguistic experience in the perception of speech," in *Perception and Experience*, edited by R. D. Walk and H. L. Pick (Plenum, New York), 125–169.
Streeter, L. A. (**1976**). "Language perception of 2-month-old infants shows effects of both innate mechanisms and experience," Nature (London) **259**, 39–41.
Sundara, M., Polka, L., and Genesee, F. (**2006**). "Language-experience facilitates discrimination of /d-ð/; in monolingual and bilingual acquisition of English," Cognition **100**, 369–388.
Trehub, S. E. (**1973**). "Infants' sensitivity to vowel and tonal contrasts," Dev. Psychol. **9**, 91–96.
Trehub, S. E. (**1976**). "The discrimination of foreign speech contrasts by infants and adults," Child Dev. **47**, 466–472.
Tsao, F. M., Liu, H. M., and Kuhl, P. K. (**2004**). "Speech perception in infancy predicts language development in the second year of life: a longitudinal study," Child Dev. **75**, 1067–1084.
Werker, J. F., and Tees, R. C. (**1984**). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," Infant Behav. Dev. **7**, 49–63.
Werker, J. F., Gilbert, J. H. V., Humphery, K., and Tees, R. C. (**1981**). "Developmental aspects of cross-language speech perception," Child Dev. **52**, 349–355.
Werker, J. F., Polka, L., and Pegg, J. E. (**1997**). "The conditioned head turn procedure as a method for testing infant speech perception," Early Dev. Parenting **6**, 171–178.
Werker, J. F., and Curtin, S. (**2005**). "PRIMIR: A developmental framework of infant speech processing," Lang. Learn. Dev. **1**, 197–234.

Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners

Erwin L. J. George,[a] Joost M. Festen, and Tammo Houtgast
ENT/Audiology, VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

(Received 7 July 2005; revised 26 June 2006; accepted 12 July 2006)

The Speech Reception Threshold for sentences in stationary noise and in several amplitude-modulated noises was measured for 8 normal-hearing listeners, 29 sensorineural hearing-impaired listeners, and 16 normal-hearing listeners with simulated hearing loss. This approach makes it possible to determine whether the reduced benefit from masker modulations, as often observed for hearing-impaired listeners, is due to a loss of signal audibility, or due to suprathreshold deficits, such as reduced spectral and temporal resolution, which were measured in four separate psychophysical tasks. Results show that the reduced masking release can only partly be accounted for by reduced audibility, and that, when considering suprathreshold deficits, the normal effects associated with a raised presentation level should be taken into account. In this perspective, reduced spectral resolution does not appear to qualify as an actual suprathreshold deficit, while reduced temporal resolution does. Temporal resolution and age are shown to be the main factors governing masking release for speech in modulated noise, accounting for more than half of the intersubject variance. Their influence appears to be related to the processing of mainly the higher stimulus frequencies. Results based on calculations of the Speech Intelligibility Index in modulated noise confirm these conclusions. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2266530]

PACS number(s): 43.71.Ky, 43.66.Sr, 43.71.Gv, 43.66.Mk [JHG] Pages: 2295–2311

I. INTRODUCTION

A common complaint among hearing-impaired listeners is that speech, although audible, may not be understood, especially in conditions where background noise is involved. When masker levels fluctuate over time, the difference between normal and hearing-impaired listeners is remarkable. Hearing-impaired listeners appear to benefit less from the relatively silent periods or gaps in this type of masker (Festen and Plomp, 1990; Bronkhorst and Plomp, 1992; Takahashi and Bacon, 1992; Hygge *et al.*, 1992; Gustafsson and Arlinger, 1993; Stuart and Phillips, 1996; Peters *et al.*, 1998; Snell *et al.*, 2002; Nelson *et al.*, 2003; Wagener and Brand, 2005). Understanding the processes behind this effect is important, since fluctuating backgrounds are very common in everyday situations.

A. Factors influencing masking release

It is still unclear to what extent audiometric differences can account for the above-mentioned differences in masking release. In most cases, merely a lack of audibility, due to audiometric hearing loss or masking noise, is not enough to explain the problems experienced when listening to speech in fluctuating noise (Eisenberg *et al.*, 1995; Bacon *et al.*, 1998; Summers and Molis, 2004). As already observed by Plomp (1978), many hearing-impaired listeners have difficulties understanding speech in noise, even if both speech and noise are well above threshold. This motivated a model description for the Speech Reception Threshold (SRT) based on two parameters: hearing loss due to attenuation and hearing loss due to distortion. When speech and noise have similar overall spectra and are above threshold for all frequencies, the distortion component is considered to be the reflection of suprathreshold deficits in hearing. These deficits are considered to be caused by deterioration of the functioning of the inner ear, like reduced temporal and spectral resolution and a loss of normal auditory compression. The interrelationship between these deficits and their relation with the hearing threshold is still under discussion (Ludvigsen, 1985; Moore *et al.*, 1999; Oxenham and Bacon, 2003).

Reduced temporal resolution is known to adversely affect masking release. A loss of temporal resolution gives rise to more forward masking, i.e., the masker will decay more slowly after termination of the masking sound, thus decreasing perceived gap size. Therefore, hearing-impaired listeners with reduced temporal resolution are expected to experience a smaller masking release in speech (Glasberg *et al.*, 1987; Festen and Plomp, 1990; Glasberg and Moore, 1992; Festen, 1993; Dubno *et al.*, 2003). In these studies, however, the amount of variance in masking release accounted for by measures of temporal resolution remains unspecified.

The influence of reduced spectral resolution on masking release is less clear, although spectral resolution plays a central role for speech intelligibility in noise (Celmer and Bienvenue, 1987; Healy and Bacon, 2006). Results from two studies by Baer and Moore (1993, 1994) show that loss of spectral resolution is related to reduced masking release. Measurements simulating cochlear implants also confirm the importance of spectral resolution to masking release (Nelson and Jin, 2004; Xu *et al.*, 2005). In contrast, however, ter Keurs *et al.* (1993a, 1993b) show that reduced spectral reso-

[a]Electronic mail: elj.george@vumc.nl

lution in hearing-impaired listeners is only loosely associated with speech intelligibility in noise, although a significant influence of spectral smearing on masking release was found. They conclude that even listeners with significantly broadened filters still have sufficient spectral resolution to resolve spectral cues important for speech intelligibility.

Besides audibility and suprathreshold deficits, a third factor that may affect masking release, mentioned by, e.g., Füllgrabe *et al.* (2006), is a listener's general ability to reconstruct the spectro-temporal structure of speech from incomplete information. As demonstrated by Warren (1970), human listeners can perceive missing phonemes by using the redundancies in speech at the acoustic, phonetic, phonological, and/or lexical level. More recently, Howard-Jones and Rosen (1993) showed that normal-hearing listeners are able to integrate information across various spectral bands at different times, a process which they called "uncomodulated glimpsing." This ability is likely to be adversely affected in listeners with deteriorated spectral or temporal resolution. Other effects that may affect masking release are comodulation across spectral bands (Hall *et al.*, 1984) and informational masking (Summers and Molis, 2004). In particular, comodulation across bands (comodulation masking release or CMR) may also be related to deteriorated spectral or temporal resolution (Hall *et al.*, 1988), although it is expected to only slightly influence masking release (Festen, 1993).

B. Accounting for audibility

A common approach to distinguish between problems in speech reception related to limited audibility and to suprathreshold deficits is to include not only normal-hearing (NHR) and hearing-impaired (HI) listeners in an experiment, but also a third group of subjects with only threshold-related problems. In the current experiment, this group consisted of normal-hearing listeners who received an additional masking noise, such that their masked pure-tone thresholds were equal to the average hearing threshold in the hearing-impaired group at 1/3-octave frequencies (cf. Fabry and Van Tasell, 1986; Humes *et al.* 1987; Zurek and Delhorne, 1987; Dubno and Schaefer, 1992). This group with simulated hearing loss will be indicated as the SIM group.

When listening to speech, hearing-impaired listeners may suffer from reduced audibility and suprathreshold problems. However, listeners with a simulated hearing loss, by definition, only suffer from a threshold-related problem. Therefore, comparison of the differences in the SRT between these groups makes it possible to distinguish between threshold-related and suprathreshold problems in understanding speech. Suprathreshold problems in speech understanding are defined here as the specific part of individual deterioration in speech intelligibility performance, that cannot be accounted for by a loss of audibility. The difference in masking release between the NHR group and the SIM group is then regarded as an estimate for the threshold-related component of the speech hearing loss, while the difference between the HI group and the SIM group is considered to be an estimate of the component due to suprathreshold deficits.

This approach was pursued earlier by Bacon *et al.* (1998), who measured masking release in temporally complex backgrounds for normal-hearing, hearing-impaired, and noise-masked normal-hearing listeners. They concluded that reduced masking release in speech for hearing-impaired listeners can only sometimes be accounted for entirely by reduced audibility. Similar conclusions can be drawn from other studies (Zurek and Delhorne, 1987; Dubno and Schaefer, 1992; Eisenberg *et al.*, 1995; Dubno *et al.*, 2002, 2003), although the contribution of audibility and specific suprathreshold deficits to speech understanding still remains unclear.

An alternative method to account for the effects of audibility is to spectrally adapt the auditory stimuli, to assure audibility of the signal at all frequencies. This method was also applied in the current study: all measurements were performed in two spectral modes, with the masker and the speech spectrally shaped according to the long-term average of natural speech, or optimized with respect to individual hearing thresholds. Differences in results between the two modes will be investigated to assess the influence of spectrum shape on speech recognition.

A drawback of spectrally adapting the signal with respect to individual hearing thresholds is, however, that the overall level of the signal is different for each listener. Effects of presentation level will thus have to be considered. It is known for normal-hearing listeners that masking release, spectral resolution, and temporal resolution are level-dependent. Spectral resolution deteriorates with increasing presentation level for normal-hearing listeners (Dubno and Schaefer, 1992; Sommers and Humes, 1993a, b), while temporal resolution is enhanced at higher levels (Jesteadt *et al.*, 1982; Fitzgibbons, 1983; Fitzgibbons and Gordon-Salant, 1987). Masking release is expected to be positively affected by the better temporal resolution at higher levels (Festen, 1993). However, Summers and Molis (2004) showed some evidence that benefit of masker fluctuations decreased for levels above 60 dB SPL. To adequately distinguish between threshold-related and suprathreshold deficits, the effect of level on masking release and on spectral and temporal resolution as found in the current experiment will be taken into account before investigating the influence of the actual suprathreshold deficits on masking release.

Finally, the effect of audibility can be accounted for by applying the Speech Intelligibility Index or SII (ANSI S3.5-1997), a measure of speech intelligibility performance which is able to handle intersubject audiogram and spectrum differences. The SII has been extensively validated for stationary masking noise and recently, Rhebergen and Versfeld (2005) proposed an extension to the model, which makes it also applicable to fluctuating background maskers. A slightly modified version of their model will be used in the current study to translate the measured SRT values to SII values. The SII will be introduced further in Sec. III D. Validation and details of the SII model can be found in the Appendix.

C. Objectives

The objectives of this paper are (1) to determine the magnitude of the differences between normal-hearing and hearing-impaired listeners in masking release for speech

across various conditions of modulated noise, and (2) to investigate the extent to which differences in masking release between groups can be accounted for by reduced audibility and by suprathreshold deficits. Thus, although results in the various conditions will be reported, this paper will focus on differences in masking release between the groups, instead of differences between conditions. Therefore, the results will be collapsed across the various modulation characteristics, to obtain an overall measure of masking release for speech due to masker modulation.[1]

It is expected, consistent with the above-mentioned literature, that differences in hearing thresholds between groups will not be enough to explain differences in masking release. Therefore, correlations will be studied between differences in masking release and four measures of spectral and temporal acuity, as measured for all individual listeners. Two of these measures [spectral resolution (F) and temporal resolution (T)] are derived from psychophysical detection tasks and are regarded to estimate the individual auditory-filter and temporal-window width around the central frequency of 1 kHz. The other two [speech reception bandwidth threshold (SRBT) and speech reception timewidth threshold (SRTT)] use a procedure similar to the SRT test to estimate the amount of speech information needed in a limited frequency range and in short time intervals, respectively. The reason for measuring the SRBT and the SRTT is that they reflect a listener's ability to extract speech information from a spectral or a temporal gap in noise (Noordhoek *et al.*, 1999, 2000). They are considered to be related to spectral and temporal resolution (F and T), but also to the earlier mentioned ability to reconstruct the spectro-temporal structure of speech from incomplete information by using redundancies at the acoustic, phonetic, phonological, and/or lexical level (cf. Warren, 1970). As such, they may involve capacities that are relevant for masking release in modulated noise.

II. EXPERIMENT AND METHOD

A. Apparatus

The experiment was run on a Dell personal computer, with a Creative Labs Audigy external sound device and Beyer Dynamic DT48 headphones. To be able to reach high stimulus levels, an additional Shure FP22 stereo headphone amplifier was used. All measurements were performed while listener and investigator were seated in a sound-insulated room. Interfering signals were generated by multiplying white noise with the appropriate amplitude modulation function, after which the speech spectrum was imposed by filtering with a 2048-point finite impulse response (FIR) filter. Final spectral shaping of both speech and masker was performed via a 1024-point windowed FIR filter, by using individual thresholds as inputs. This filter also corrected the frequency response of the headphones and restricted the bandwidth of both noise and speech signal to frequencies between 125 Hz and 8 kHz.

B. Speech material

A set of short meaningful everyday sentences was used, as developed and evaluated by Versfeld *et al.* (2000). The first 32 lists of this set, read by a male speaker, were used, each list containing 13 sentences of eight or nine syllables. The set was developed to enable efficient measurement of the SRT in stationary speech-shaped noise and can be considered as being equivalent to the smaller sentence set of Plomp and Mimpen (1979), giving rise to a standard error in SRT of about 1 dB for normal-hearing listeners. Under all masker conditions, the long-term spectra of the speech and the masker were similar in shape.

TABLE I. Details on the temporal characteristics of the masking noises. For the block-modulated maskers, nondefault values are displayed in bold-italics. The temporal wave forms of the various background maskers are shown in Fig. 1. Not mentioned is the SRBT, which was measured for a subset of listeners. For more details on the SRTT and the SRBT, see the text.

No.	Masker description	Duty cycle (dc) (%)	Mod depth (md) (dB)	Mod freq (f_{MOD}) (Hz)
1	Silence	…	…	…
2	Stationary	100	0	0
3	Speech modulation	Undefined	Speechlike	Speechlike
4	Block, default	50	∞	16
5	Block, dc=75%	***75***	∞	16
6	Block, md=15 dB	50	***15***	16
7	Block, f_{MOD}=32 Hz	50	∞	***32***
8	SRTT (fixed SNR)	Variable	∞	16

C. Interfering signals

SRT measurements were performed using a variety of background maskers, differing both in temporal and spectral characteristics. Eight different temporal masking conditions were applied in two modes, with the masker and the speech spectrally shaped according to the long-term average of natural speech, or optimized with respect to individual hearing thresholds. The long-term spectral shape of the masker was equal to that of the speech, in all conditions. All tests were conducted both in test and retest, so, a total of 32 (8 backgrounds × 2 spectral modes × 2 tests) SRT measurements were performed for each participant. In addition to the set of 32 SRT measurements, the SRBT (Noordhoek *et al.*, 1999, 2000) was measured for a subset of listeners.

1. Temporal characteristics

SRT measurements were performed in background noise with temporal characteristics ranging from stationary noise to fast block-wave modulated noise. Masker characteristics were chosen to investigate the different aspects of relatively silent periods to speech intelligibility as adequately as possible. Detailed information about the used conditions is given in Table I. The temporal wave forms of the various background maskers are shown in Fig. 1.

The silence (1) and stationary-masker (2) conditions were included as reference conditions. In condition (3), a masker was used that specifically mimics the intensity fluctuations of speech. This masker was generated using the method described by Festen and Plomp (1990), which splits up a steady-state masker in a low- and high-frequency part with 1000-Hz crossover frequency. Both parts are then modulated separately with the envelope of speech from the

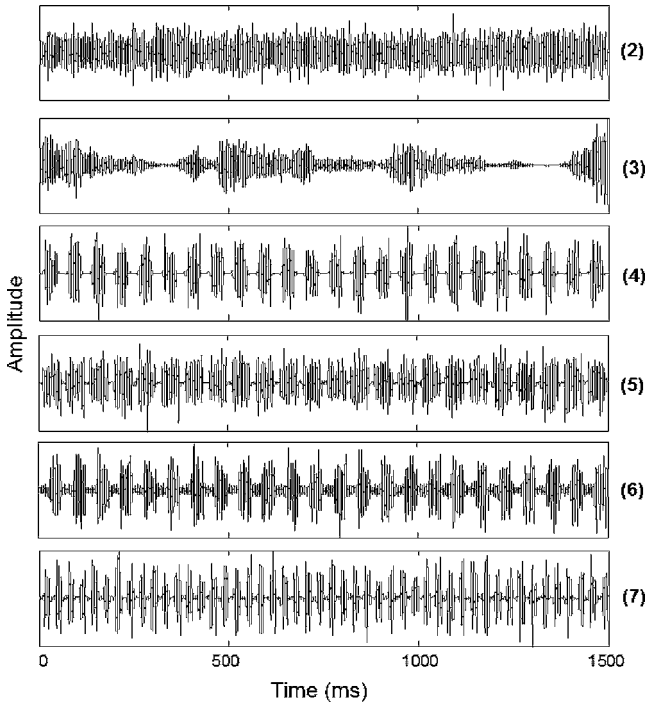


FIG. 1. The temporal wave forms of the various background maskers: steady-state noise (2), speech-modulated noise (3), and four different forms of block-modulated noise (4–7). The temporal wave form of the SRTT condition (8, not shown) is similar to the wave form of condition (4), only with an adaptively changing duty cycle.

corresponding frequency region, after which they are added while restoring the original level ratio between the two bands.

Conditions (4)–(7) use block-modulated noise. To investigate the masking effects of modulated noise, condition (4) was chosen as the default condition, whereas in each of the conditions (5)–(7) one masker-parameter is varied, respectively the duty cycle (dc), the modulation depth (md), and the modulation frequency (f_{MOD}) of the masker. In condition (4), the masker contains a fairly long silent period, which may possibly lead to a better SRT, when the listener uses the gaps in the noise optimally. Conditions (5)–(7) decrease, each in its own way, the available amount of speech information in the gap, giving rise to a deterioration of the SRT. This approach enables a comparison of SRTs between these conditions to determine the effect of different masker-parameters.

Finally, condition (8) was included to determine the time window width of clear speech, that a listener needs to correctly reproduce 50% of the sentences. In this condition, the masker and the speech are of equal level, and block modulated (chopped) in such a way that they alternate. This means that speech is only present when the masker is absent and vice versa. The difference with a standard SRT measurement is that the duty cycle of both masker and speech is varied adaptively, rather than the signal-to-noise-ratio. The speech duty cycle at which 50% of the sentences is reproduced correctly will be referred to as the SRTT.

Unmentioned in Table I is the spectral equivalent of the SRTT, the SRBT (Noordhoek *et al.*, 1999, 2000), which was measured for a subset of listeners. It is not listed in Table I since it was not included in the original design of the experiment, but was added considering its similarity to the SRTT. Like the SRTT, it measures the listener's ability to reconstruct speech from fragments, but now in the spectral domain. The adaptive measurement procedure used is the same as in the SRTT, but the available speech bandwidth is varied instead of the speech duty cycle. The SRBT is defined as the speech bandwidth at which 50% of the sentences can be reproduced correctly. Speech and noise were presented at the half-way point in the listener's dynamic range, similar to the SRT in adapted spectral mode.

2. Spectral characteristics

All temporal masker conditions were presented in two modes, either with a spectrum shaped according to the long-term average of natural speech or with a spectrum that was optimized using the individual's absolute thresholds. These two spectral measurement modes will be referred to as SRT-normal (SRT-n) and SRT-adapted (SRT-a), respectively. The long-term spectral shape of the speech was always equal to that of the masker.

In the SRT-n mode, the masker level was set such that the overall level between 125 Hz and 1 kHz was equal to that level in the SRT-a mode, giving rise to only small level differences between the two modes at low frequencies.

In the SRT-a mode, individual hearing thresholds were used to adapt the spectrum of the masker to reach 1/3-octave masker levels equal to the estimated middle of the dynamic range for each listener. The lower limit of the dynamic range was chosen to be the individual pure-tone threshold at each 1/3 octave, while the upper limit was the uncomfortable loudness level (UCL), here chosen at 110 dB SPL for all listeners. The masker level was the mean of these two. Since pure-tone thresholds were only measured at 1/3-octave frequencies from 125 Hz to 8 kHz, intermediate threshold levels were calculated by interpolation. The speech signal was shaped accordingly and was varied in level to set the signal-to-noise ratio in the adaptive measurement procedure.

Figure 2 gives an overview of masker spectra in the SRT-n and SRT-a modes, for a normal-hearing listener and two imaginary hearing-impaired listeners with either a typical flat or a typical sloping hearing loss. From this figure, it can be seen that differences between the two spectral modes mainly occur at higher frequencies (above 1 kHz). The SRT-a masker spectrum is above threshold at all frequencies, thus assuring audibility for both normal-hearing and hearing-impaired listeners. This means that in this mode the possible effect of the individual threshold on speech intelligibility is minimized. The SRT-n masker spectrum, on the other hand, approaches the hearing threshold as frequency increases, giving rise to a possible loss of available speech information, mainly in the higher frequencies. This effect occurs especially for listeners with a sloping hearing threshold.

Both spectral modes have their advantages and their drawbacks. Measurement results from the SRT-n mode are bound to give a good impression of the speech intelligibility in real life, since natural spectra are applied. However, inter-individual differences in audibility might influence intelligibility in this mode, making it less powerful in determining

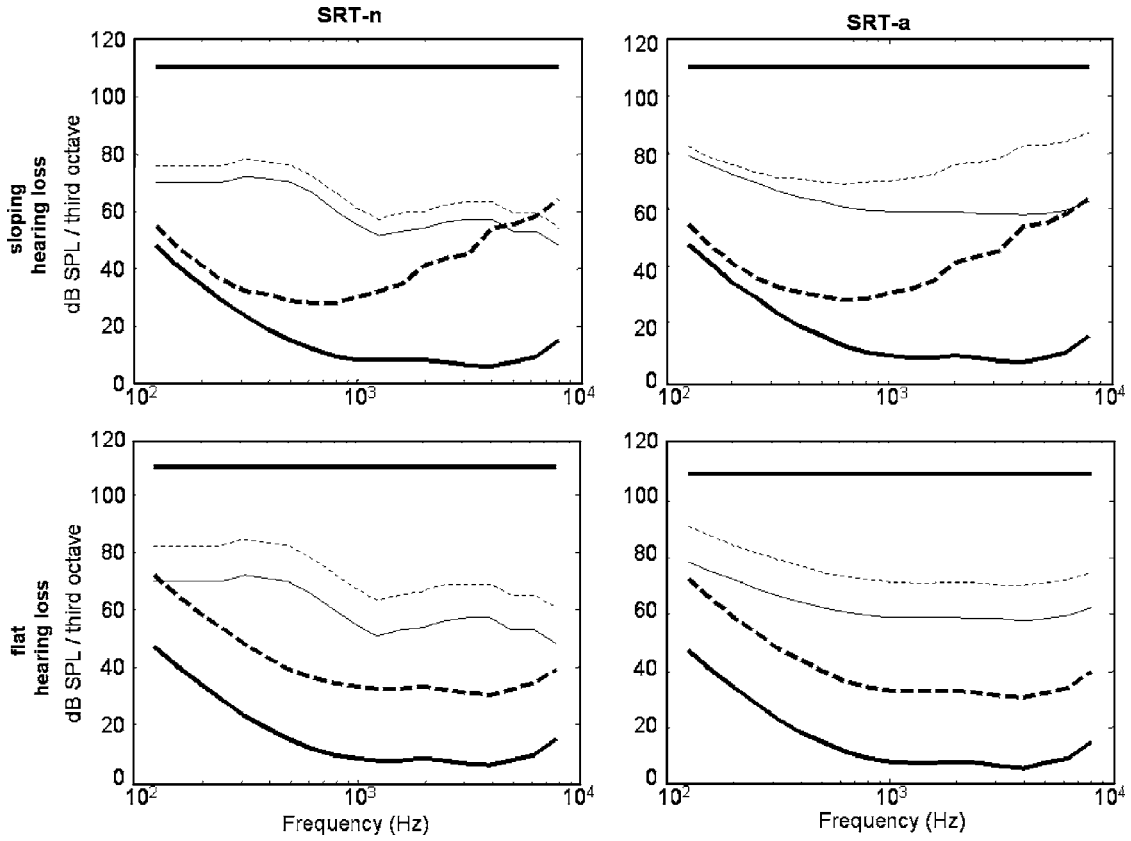


FIG. 2. Hearing threshold, uncomfortable loudness level (UCL), and the masker spectrum in both normal (left panel) and adapted (right panel) spectral mode. *Solid* lines show the pure-tone hearing threshold (solid bold) and the masker spectrum (solid normal) for a listener without hearing loss. *Dashed* lines show the pure-tone hearing threshold (dashed bold) and the masker spectrum (dashed normal) for a typical listener with sloping hearing loss (upper panel) and flat hearing loss (lower panel). The UCL was set at 110 dB SPL at all frequencies for all listeners and is marked by a straight solid bold line.

effects of suprathreshold deficits. The SRT-a mode minimizes the differences in audibility, and will therefore be more powerful to examine suprathreshold deficits. However, it has the disadvantage that listeners may not be used to listening to sounds in the middle of their dynamic range. In particular, the presence of high-frequency boosted sounds might introduce an unwanted additional disadvantage in understanding speech, especially for listeners with a sloping hearing loss.

D. Procedure

Each experimental session started with the measurement of the listener's pure-tone hearing threshold at all 19 1/3-octave frequencies between 125 and 8000 Hz, using the same apparatus used during all other measurements. The audiogram was used later as an input to shape the spectrum of the stimuli. The experiment included a total of 32 SRT-measurements (including the SRTT), plus the measurement of spectral and temporal acuity and the SRBT, defined in the following. Each measurement was performed twice, following a test-retest design, which makes it possible to estimate reliability for all data points. Measurements were conducted monaurally, using the listener's best ear, which was chosen according to his or her audiogram, or, when the hearing loss was symmetrical, personal preference in telephone conversation.

1. SRT

SRT measurements were performed using a simple adaptive up-down procedure as described by Plomp and Mimpen (1979). In each condition, the appropriate masker and a list of 13 sentences, unknown to the listener, were presented. The masker level was kept fixed, while the speech level was varied adaptively to estimate the SRT, defined as the estimate of the speech-to-noise ratio at which 50% of the sentences could be reproduced without error. In each condition, the first sentence was presented at a level below threshold and repeated, at 4-dB higher levels with each repetition, until the listener was able to reproduce it correctly. The remaining 12 sentences were presented only once, following an one-up-one-down adaptive procedure, with a 2-dB step size. An errorless reproduction of the entire sentence was required for a correct response. The SRT was estimated as the average presentation level of sentences 4–13.

Sentences were presented to the listener over the headset and to the investigator visually on a computer screen. To avoid the confounding of both measurement order and sentence lists with condition effects, the order of conditions was counterbalanced across subjects according to an eight-by-eight digram-balanced Latin square, while sentence order was kept fixed. This approach ensures that each of the eight possible temporal conditions, and also pairs of succeeding conditions, occurred only once within each subgroup of eight

listeners. For the same reason, half of the participants in each subgroup started with SRT-a conditions, while the other half started with SRT-n conditions.

2. SRTT and SRBT

In the SRTT condition, a comparable adaptive procedure was used. However, in the SRTT condition the speech-to-noise ratio was fixed, such that the rms level of signal and masker were equal. The duty cycle of alternating block-chopped speech and noise was varied in a complementary way to estimate the Speech Reception Timewidth Threshold or SRTT, defined as the speech duty cycle (i.e., available timewidth) at which 50% of the sentences was reproduced correctly. Speech was only present when the masker was absent and vice versa, so there was no simultaneous masking. Masker presence was necessary, however, to provide comfortable listening, since chopped speech on its own would be perceived as annoying.

Step sizes of the adaptive SRTT procedure were chosen to fit the step sizes of the standard SRT procedure (i.e., 4 dB for the first sentence and 2 dB for all other sentences). Classical speech intelligibility prediction models, like the Articulation Index as introduced by French and Steinberg (1947) and Kryter (1962), assume that, when presenting speech to a listener, speech information grows linearly to its maximum over a 30 dB range. Thus, changing the speech level with 4 or 2 dB steps corresponds to a change of available information of 13.3% or 6.7%, respectively. Therefore, duty cycle changes of, respectively, 12% for the first sentence and 6% for all other sentences seemed appropriate. As in all other conditions, presentation of the first sentence started at a duty-cycle below reception threshold and was repeated, with increasing speech duty cycle, until the listener was able to reproduce it correctly. All other sentences were presented only once. The SRTT was estimated as the average speech duty-cycle while presenting sentences 4 to 13.

Just as the SRTT gives an estimate of the amount of speech information needed in short time intervals, the speech reception bandwidth threshold or SRBT, as introduced by Noordhoek *et al.* (1999, 2000), gives a good estimate of the amount of speech information a listener needs in a limited frequency range. The SRBT measurement used is the same as Noordhoek's, following the same adaptive procedure as the SRTT, but varying the available speech bandwidth instead of the speech duty cycle. Speech and noise were presented at a level half-way up the listener's dynamic range, similar to the SRT in adapted spectral mode.

Unfortunately, normal-hearing participants were already tested before it was decided to include SRBT measurements in the test battery. This means that they were also not included in the counterbalanced Latin square conditional order. Instead, the SRBT was measured after a participant had completed all other measurements.

3. Spectral and temporal resolution

Each listener's spectral and temporal acuities were determined by employing an adaptive measurement procedure as introduced and validated by Hilkhuysen *et al.* (2005). Validation was performed by measuring 18 normal-hearing listeners. They showed reliable auditory-filter and time-window widths which were free of noteworthy learning effects, and which varied with presentation level and frequency and corresponded to values as commonly found in the literature. An advantage of the chosen test procedure is that it requires no prior training and can therefore be performed relatively fast (within 15 min) compared to more classical measurement procedures (gap detection, tones in bands of noise). Moreover, the task to perform is relatively simple ("count the number of sweeps") and easy to explain to naïve listeners. In the current experiment, both temporal and spectral acuity were determined in the frequency region around 1 kHz, at a level half-way up the listener's dynamic range (as in adapted spectral mode, to assure audibility).

The measurement procedure included three tests, in which listeners were asked to report the number of tone sweeps (0, 1, 2, or 3) they were able to detect in: (i) steady-state noise without grid; (ii) noise containing a spectral grid with a 50% duty cycle on a log-frequency scale; (iii) noise containing a temporal grid with a 50% duty cycle. In all three noises, the tone sweeps to detect were sinusoids with a duration of 200 ms, sweeping upward over a range of 1.6 octaves centered around 1 kHz (0.57–1.74 kHz) at a speed of 8 octaves/s. Thus, the sweep reached its center frequency after 100 ms. The masker duration was 2.2 s, and the possible tone sweeps could start at 0.6, 1.0, or 1.4 s after masker onset. The phase of the temporal noise grid varied randomly over trials, while the phase of the spectral noise grid always provided a spectral gap logarithmically centered around 1 kHz. Third-octave levels of the masker were set half-way up the listener's dynamic range, and sweep and masker had equal spectrum shape in the frequency range of the sweep.

The level of the tone sweeps (for the steady-state noise) or the gap width of the noise-grid maskers was varied adaptively in a one-up-one-down 4-AFC-procedure (Levitt, 1971), starting above detection threshold for all listeners. In the two noise grids, the level of the tone sweeps was fixed at a signal-to-noise ratio of −6 dB. In steady-state noise, the initial step size was 4 dB, while, in the grid conditions, the gap width was initially changed with a factor $\sqrt{2}$. After four transitions of an incorrect response following a correct response, the actual test started, using step sizes of 2 dB or a factor 2^(1/4), respectively. The actual test consisted of a random-order set of 24 stimuli, in which each number of sweeps (zero to three) was presented six times. Responses to the zero-sweep trials had no consequences for the adaptive procedure. The detection threshold was defined by the average level or gap width of the last 18 presentations in which one or more sweeps were present. On the basis of the three outcome measures (threshold level in steady-state noise, width of the spectral grid, width of the temporal grid), auditory-filter and time-window widths were estimated by a fitting procedure, assuming a symmetrical one-parameter RoEx-function for the spectral filter shape (Patterson *et al.*, 1982) and a decaying exponential for the time window shape (Duifhuis, 1973; Festen and Plomp, 1981).

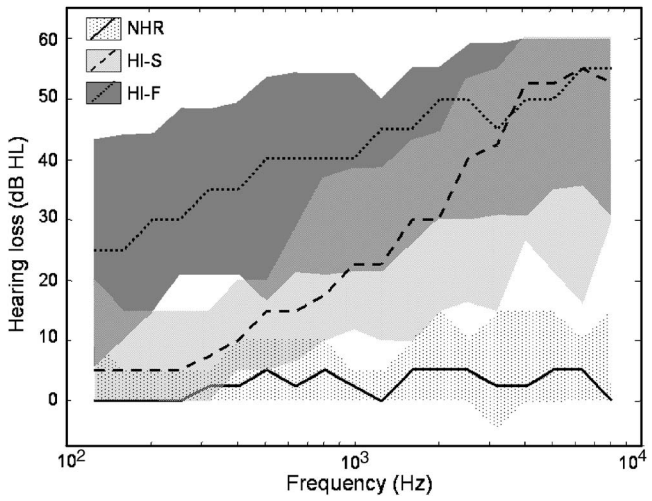


FIG. 3. Average pure-tone hearing threshold (*re:* ISO-389-1991) and the range between the 5th and 95th percentiles for normal-hearing listeners (NHR) and two groups of hearing-impaired listeners (HI-S and HI-F). The overlap between HI-S and HI-F is displayed as a medium gray region.

E. Participants

1. Normal-hearing listeners (NHR)

The reference group consists of 8 normal-hearing listeners, selected to have pure-tone hearing thresholds better than 10 dB HL at 0.25, 0.5, and 1 kHz and better than 15 dB at 2 and 4 kHz. Six of the eight participants included in this group were university students. Group age ranged from 19 to 56 years, with an average of 28.5 years.

2. Hearing-impaired listeners (HI)

A group of 32 listeners with a sensorineural hearing impairment were selected from the patient database of the audiology department of the VU University Medical Center. Sixteen listeners having a flat hearing threshold (HI-F) and sixteen listeners with a sloping hearing threshold (HI-S) were selected. To be included in the HI-F group, pure-tone hearing thresholds on octave frequencies between 0.25 and 4 kHz were required to be larger than 25 dB HL, not varying more than 10 dB around their average. The HI-S group was selected to have a pure-tone hearing threshold better than 10 dB HL at 0.25 kHz, and between 30 and 60 dB at 4 kHz. These groups were included to investigate the effect of hearing-loss shape on speech intelligibility.

The final HI-F group included only 13 of the original 16 participants identified. One participant did not fulfill the inclusion criteria and two participants in this group showed very large inconsistencies between results on test and retest and were therefore excluded from further analysis. The final 13 participants in the HI-F group were aged between 27 and 80 years, with an average of 64.5 years. The age of the 16 participants in the HI-S group ranged from 45 to 76 years, with an average of 60.8 years.

Figure 3 displays the hearing threshold for the normal-hearing and for the two groups of hearing-impaired participants, including the regions between the 5th and 95th percentiles. As can be seen, the overlap between both groups of hearing-impaired is fairly large, especially in the higher frequencies. Instead of representing two groups of hearing-impaired with clearly different audiograms, the hearing thresholds seem to be more adequately described as forming a continuum. Therefore, results of the two groups of hearing-impaired will not be dealt with separately, although the distinction is preserved in displaying the results, to enable easy comparison with the two SIM groups.

3. Simulated hearing loss listeners (SIM)

A group of participants with simulated hearing loss was also included, consisting of 16 normal-hearing listeners not included in the NHR group. These listeners' thresholds were elevated by adding noise to the stimuli. The criteria used to select participants in this group were the same as for the normal-hearing reference group.

Similar to the group of hearing-impaired, this group of simulated hearing loss listeners was split in two. In 8 listeners, a rather flat hearing threshold was simulated, equal to the average hearing threshold of the 13 hearing-impaired listeners in the HI-F group. A more sloping hearing threshold, equal to the average hearing threshold of the 16 hearing-impaired listeners in the HI-S group, was simulated in the other 8 listeners. These two subgroups will hereafter be referred to as the SIM-F and the SIM-S group, respectively. The participants in the SIM-F group were aged between 19 and 29 years, with an average of 24.5 years. The age of the participants in the SIM-S group ranged from 19 to 24 years, with an average of 21.8 years.

The hearing threshold simulation was performed by presenting an additional noise to the listeners in the SIM groups. The goal of this additional noise was to mask the low intensity parts of the signal (cf. Fletcher, 1940; Hawkins and Stevens, 1950), such that the listeners would experience audibility problems when listening to speech. The introduction of an additional broadband noise seems the most appropriate control condition for comparison to sensorineural hearing impairment (Humes *et al.*, 1988). Broadband noise simulates the reduced dynamic range and closely approximates the loudness-growth function of hearing-impaired listeners (Lochner and Burger, 1961; Stevens, 1966). In addition, it has the advantage that performance in the SIM group can be measured at presentation levels comparable to that of hearing-impaired listeners, although the high level of neuronal activity produced by the noise may not be equivalent to a reduction in activity due to cochlear hearing loss (Fabry and Van Tasell, 1986).

The spectrum level of the additional external noise was chosen to be equal to

$$X = Q - R, \tag{1}$$

where X is the spectrum level of the internal noise representing the elevated threshold, at all 1/3-octave frequencies, Q is the pure-tone threshold level of the audiogram to be simulated (in this case, the average hearing threshold of the appropriate subgroup of hearing-impaired), and R is the critical ratio in dB, as reported by Pavlovic (1987). Since the audiogram measurement at the start of each experimental session was performed in the presence of this background noise, the correct simulation of the desired hearing threshold could be

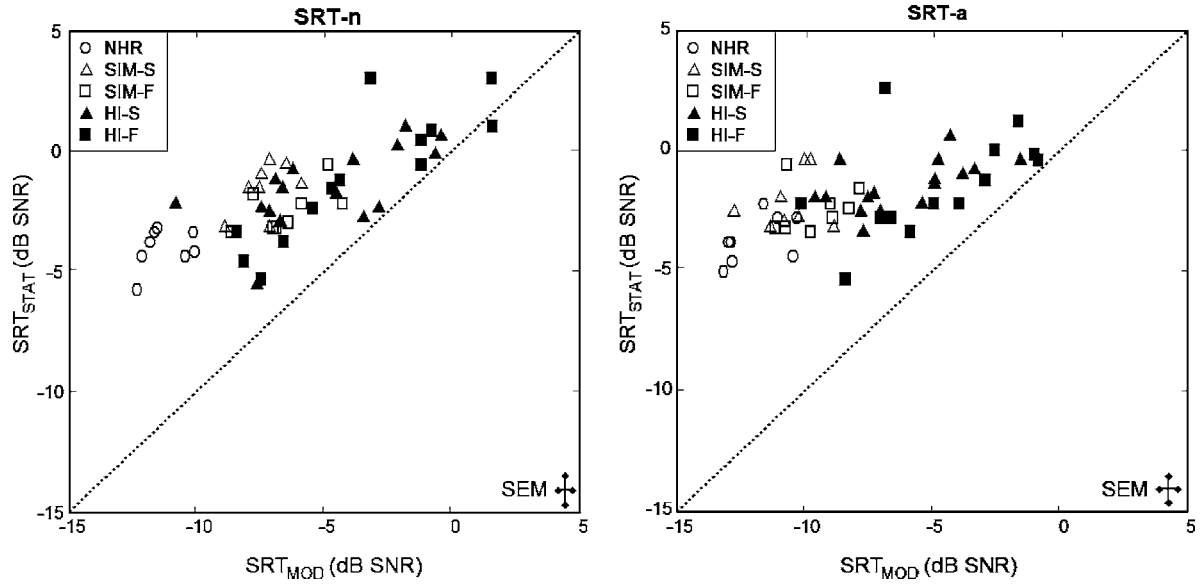


FIG. 4. SRT in stationary noise vs SRT in modulated noise, defined as the mean SRT from four modulated noise conditions, in both normal (left panel) and adapted (right panel) spectral mode. Data are displayed for normal-hearing listeners (NHR), two groups of hearing-impaired listeners (HI-S and HI-F), and two groups of listeners with simulated hearing loss (SIM-S and SIM-F). The crosses in the lower right corners show the standard errors of measurement (SEM).

checked. All participants in the SIM groups showed thresholds within a few dB of the desired threshold, at all 19 1/3-octave frequencies.

III. RESULTS AND DISCUSSION

The following gives an overview of the results. In Sec. III A, results from SRT measurements are discussed broadly, to clarify our observation that presentation level is an important factor which should be taken into account when drawing conclusions regarding the effect of suprathreshold deficits on masking release for speech. Sec. III B presents a discussion on the effect of presentation level on tests that are considered to be linked to suprathreshold deficits. These level effects are used to relate the actual suprathreshold deficits to masking release for speech in modulated noise in Sec. III C. In Sec. III D, the Speech Intelligibility Index (SII) is introduced. The SII is applied to transform SRT to SII values, the results of which are discussed in Sec. III E. Finally, Sec. III F addresses the difference in results between both spectral modes of signal presentation.

A. Effect of presentation level on masking release (SRT)

The first goal of our work was to determine the differences in SRT between modulated and stationary noise for normal-hearing and hearing-impaired listeners. Figure 4 gives an overview of results from the SRT measurements in both spectral modes. For each participant, SRT_{MOD} is defined as the average SRT over the four block-modulated masker conditions (conditions 4–7). As mentioned in Sec. I, the shown results are thus collapsed over the various modulation characteristics, to obtain an overall measure of masking release. SRT_{STAT} is the SRT measured in stationary noise (condition 2). Detailed results are presented in Table II.

The data displayed in Fig. 4 show rather small spread among listeners for speech intelligibility in stationary noise, while interindividual differences in SRT are more apparent in modulated noise, as noted earlier by, e.g., Bacon *et al.* (1998) and Versfeld and Dreschler (2002). The difference between SRT_{MOD} and SRT_{STAT} can be regarded as the extent to which an individual listener benefits from the relatively silent periods in modulated noise, i.e., the release from masking. In Fig. 4, each individual's vertical deviation from the diagonal is the graphical representation of this benefit. As expected, normal-hearing listeners, denoted by the open circles in Fig. 4, clearly benefit from the relatively silent periods in modulated maskers, improving their mean SRT from −4.1 dB in stationary noise to −11.3 dB in modulated noise. In contrast, hearing-impaired listeners, as denoted by the closed symbols in Fig. 4, benefit less when going from stationary to modulated noise; some even obtain no benefit at all.

Listeners with a simulated hearing loss (SIM-S and SIM-F groups, denoted by open triangles and squares) are already distributed closer to the diagonal than normal-hearing listeners, i.e., already have a smaller release from masking than normal-hearing listeners. This indicates that a listener's ability to make use of the relatively silent periods in modulated noise deteriorates, even when no suprathreshold deficits are involved. To put it differently, even when a listener's problems are purely threshold-related, the benefit from the gaps in modulated noise seems to be reduced. This means that a reduced benefit from masker modulations is at least partly linked to threshold-related factors.

To explore this observation further, the difference between SRT_{MOD} and SRT_{STAT} (i.e., the masking release or benefit) is plotted against presentation level in Fig. 5. More specifically, the negative value of the experienced benefit is plotted along the vertical axis to preserve the analogy to the SRT, so, a more negative value indicates a better performance. Presentation level has been defined here as the long-term A-weighted level of the masker.

To estimate benefit as a function of presentation level, a

TABLE II. Group results on speech reception thresholds (SRT) for normal-hearing listeners (NHR), two groups of hearing-impaired listeners (HI-S and HI-F), and two groups of listeners with simulated hearing loss (SIM-S and SIM-F). SRT_{MOD} is the average SRT over the four block-modulated masker conditions (conditions 4–7). Benefit is the difference between SRT_{MOD} and the SRT measured in stationary noise (condition 2).

			Normal spectral mode (SRT-n)					Adapted spectral mode (SRT-a)				
No.	Masker description	Unit	NHR	HI-S	HI-F	SIM-S	SIM-F	NHR	HI-S	HI-F	SIM-S	SIM-F
1	Silence	dB (A)	24.7	39.6	56.0	45.1	63.8	23.0	48.4	57.5	52.4	63.8
2	Stationary	dB SNR	−4.1	−1.6	−1.1	−1.6	−2.5	−3.7	−1.5	−1.5	−2.2	−2.4
3	Speech modulation	dB SNR	−11.7	−3.3	−2.7	−6.1	−5.5	−11.8	−5.2	−3.5	−9.2	−7.7
4	Block, default	dB SNR	−17.3	−7.8	−5.1	−11.1	−9.1	−17.9	−9.7	−6.9	−15.9	−13.5
5	Block, dc = 75%	dB SNR	−6.4	−2.3	−1.5	−3.9	−4.2	−6.8	−2.9	−2.4	−6.2	−5.8
6	Block, md=15 dB	dB SNR	−8.8	−4.2	−3.6	−5.6	−5.1	−8.8	−5.2	−4.7	−7.0	−6.8
7	Block, f_{MOD}=32 Hz	dB SNR	−12.7	−5.6	−4.7	−8.8	−7.5	−14.3	−6.6	−5.2	−13.3	−12.2
8	SRTT (fixed SNR)	%	33.3	45.5	48.7	39.0	43.8	31.4	42.5	45.7	34.9	36.0
	SRT_{MOD}	dB SNR	−11.3	−5.0	−3.7	−7.3	−6.5	−11.9	−6.1	−4.8	−10.6	−9.5
	Benefit	dB	−7.2	−3.4	−2.6	−5.7	−4.0	−8.2	−4.6	−3.3	−8.4	−7.1

simple linear regression was performed on data from the normal-hearing and the two SIM groups, as denoted by all open symbols in Fig. 5. It was assumed that listeners in these three groups experience no problems related to suprathreshold deficits. Therefore, only the deviation from this regression line represents a reduced masking release possibly due to suprathreshold deficits for a specific listener. This deviation will be referred to as "Δbenefit," the capital Greek delta indicating that the effect of presentation level has been accounted for. Data from normal-hearing and simulated hearing loss listeners are distributed close to the regression line, so Δbenefit will be small for these listeners. In contrast, data from most hearing-impaired listeners deviate from the regression line, so their Δbenefit will be larger, indicating a reduced masking release due to suprathreshold deficits.

It should be noted that the effects of presentation level and hearing threshold cannot be fully distinguished in the current experiment: listeners in the simulated hearing-impaired group were measured at higher presentation levels, but also had an (artificially) raised hearing threshold. So, the regression line in Fig. 5 shows the combined effects of presentation level and masking noise. In any case, the combined effect needs to be taken into account before considering the effects of suprathreshold deficits on masking release.

In summary, the data show that even listeners with only an audibility problem (i.e., an artificially raised hearing threshold) experience less benefit than normal-hearing when it comes to understanding speech in modulated noise. Therefore, to study the effects of suprathreshold deficits, it is necessary to eliminate this threshold-related effect.

B. Effect of presentation level on spectral and temporal resolution

The second goal of this study was to investigate the extent to which differences between normal-hearing and

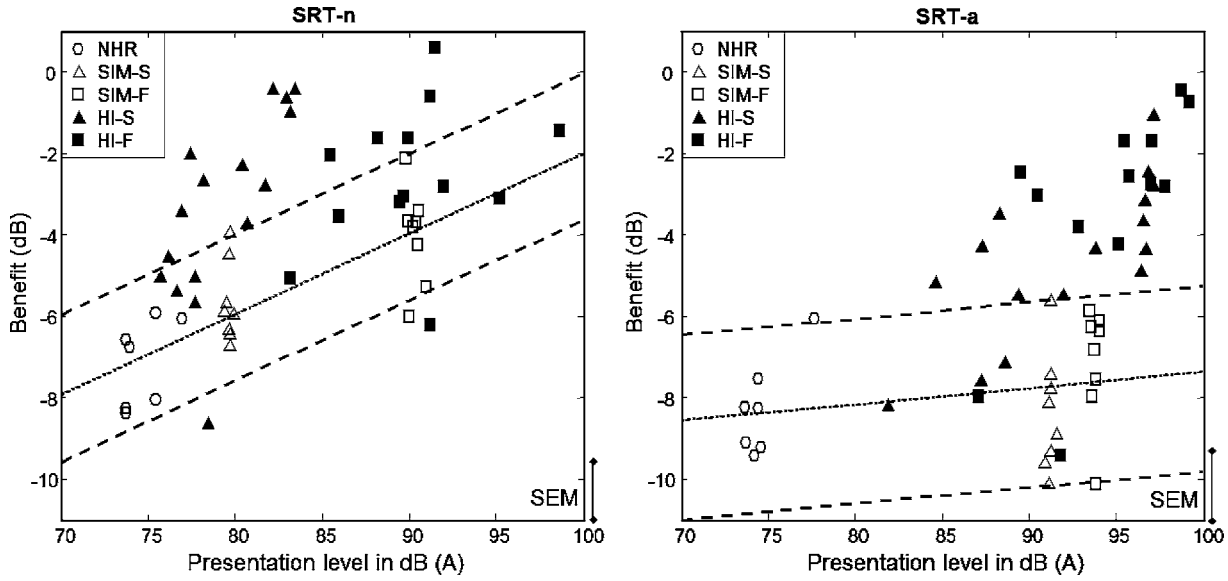


FIG. 5. Difference between SRT in modulated noise and SRT in stationary noise (i.e., benefit) as a function of masker presentation level, in both normal (left panel) and adapted (right panel) spectral mode. Data are displayed for normal-hearing listeners (NHR), two groups of hearing-impaired listeners (HI-S and HI-F), and two groups of listeners with simulated hearing loss (SIM-S and SIM-F). The dotted lines are regression lines through all open symbols; the dashed lines represents the 5th and 95th percentiles. The bars in the lower right corners show the standard errors of measurement (SEM).

TABLE III. Group averages of measured spectral resolution F, temporal resolution T, SRBT, SRTT, age, and pure-tone average (PTA), the latter defined as the average pure-tone hearing threshold at octave frequencies 0.5, 1.0, and 2.0 kHz. Averages were calculated for normal-hearing listeners (NHR), two groups of hearing-impaired (HI-S and HI-F), and two groups of listeners with simulated hearing loss (SIM-S and SIM-F). The average result on SRBT for normal-hearing listeners was taken from Noordhoek (1999).

Variable	Unit	NHR	HI-S	HI-F	SIM-S	SIM-F
F	ERB (Hz)	193.5	234.0	292.7	232.3	291.6
SRBT	octaves	(1.4)	1.64	1.87	1.49	1.57
T	ERD (ms)	4.36	4.52	6.68	2.74	2.48
SRTT	%	31.4	42.5	45.7	34.9	36.0
Age	Years	28.5	60.8	64.5	21.8	24.5
PTA	dB HL	4.4	22.8	40.0	23.2	40.9

hearing-impaired listeners can be accounted for by suprathreshold deficits. Manifestations of individual suprathreshold deficits are the measured spectral resolution (F) and temporal resolution (T), but also the SRBT and the SRTT (in adapted spectral mode), which are measures of how much speech information is needed to reach 50% intelligibility.

Our second goal can thus be achieved by relating the measured spectral and temporal resolutions, and the SRTT and SRBT to the experienced unmasking of speech in modulated noise (i.e., Δbenefit). However, as for masking release, the measured resolutions are not independent of presentation level either. This makes it necessary to first investigate these effects. The discussion to follow will be restricted to spectral and temporal resolution, but also applies to SRTT and SRBT values. Group averages of the raw data can be found in Table III.

Figure 6 shows the measured spectral and temporal resolution as a function of presentation level. Since both spectral and temporal resolution measurements were performed around 1 kHz in the middle of the dynamic range, presentation level has in this case been defined as the masker level in dB SPL per 1/3 octave in the 1-kHz band. Again, a linear regression was performed on data from normal-hearing and simulated hearing loss listeners, who are assumed to experience no problems related to suprathreshold deficits. As before, presentation level is confounded with the presence of masking noise for these listeners. However, since both the tone sweeps and the noise grids in the measurements were presented strictly at a level half-way up the dynamic range of a listener, a raised threshold is expected to play only a minor role. Therefore, the observed effect may be considered solely due to increased presentation level.

Figure 6(a) shows that, for normal-hearing and simulated hearing loss listeners, spectral resolution deteriorates with presentation level. Hearing-impaired listeners, denoted by closed symbols, seem to follow that trend: the spectral resolution of the present group of mild to moderate hearing-impaired listeners does not deviate more from the regression line than results from listeners in the SIM groups. This indicates that deteriorating spectral resolution can be explained from increased presentation levels. Since noise-masked normal-hearing and hearing-impaired listeners show comparable results, spectral resolution does not appear to qualify as a suprathreshold deficit for the present group of hearing-impaired. So, the fact that hearing-impaired listeners show less finely tuned auditory filters does not appear to be a consequence of damage in the inner ear, but the result of increased presentation levels, as suggested earlier by, e.g., Dubno and Schaefer (1992) or Sommers and Humes (1993a, 1993b).

On the other hand, Fig. 6(b) shows that the deteriorating temporal resolution in hearing-impaired listeners cannot be entirely understood in terms of increased presentation level.

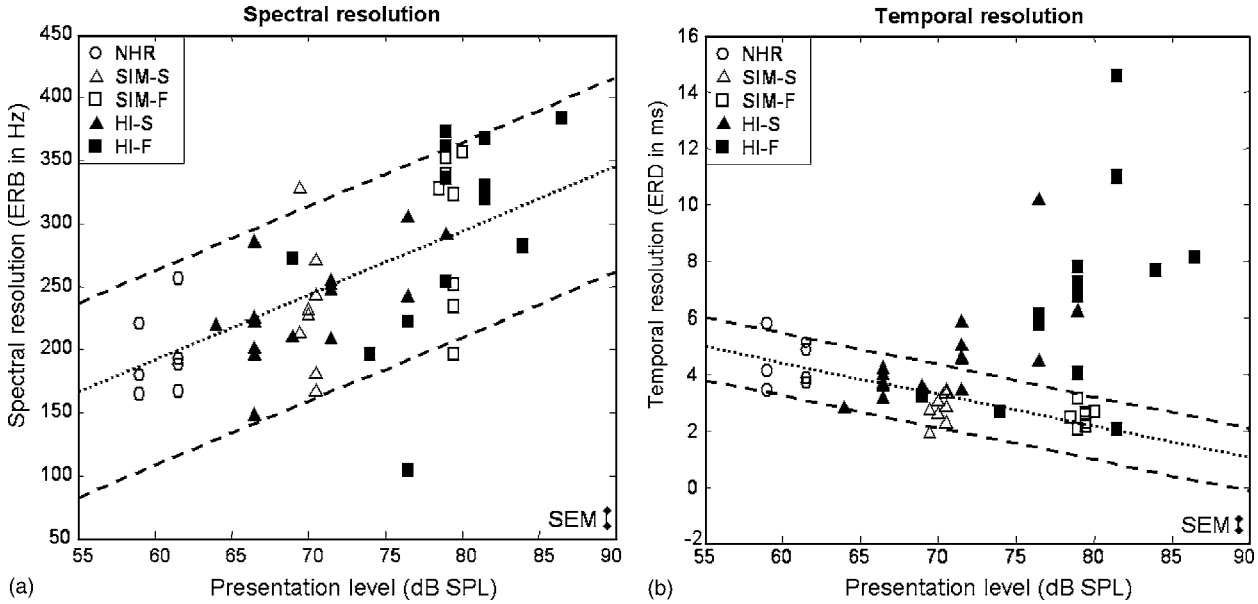


FIG. 6. Spectral resolution F (a) and temporal resolution T (b) as a function of presentation level, for normal-hearing listeners (NHR), two groups of hearing-impaired listeners (HI-S and HI-F), and two groups of listeners with simulated hearing loss (SIM-S and SIM-F). The dotted lines are regression lines through all open symbols; the dashed lines represent the 5th and 95th percentiles. The bars in the lower right corners show the standard errors of measurement (SEM).

Results for the normal-hearing and simulated hearing loss listeners (open symbols), with only an audibility problem, indicate that temporal resolution improves with presentation level. However, the temporal resolution of hearing-impaired listeners (closed symbols) largely deviates from this trend, indicating that their reduced temporal resolution does qualify as an actual suprathreshold deficit (cf. Florentine and Buus, 1984).

From these results, it is concluded that the deviation of measured temporal and spectral resolution from the regression line is the most appropriate measure for the actual amount of suprathreshold deficits of a specific listener. These deviations for spectral and temporal resolution will be referred to as "ΔF" and "ΔT," respectively. In an analogous way, the effect of presentation level on SRTT and SRBT can be taken into account. Linear regressions were performed on SRTT and SRBT data from normal-hearing and simulated hearing loss listeners. Individual deviations from these regression lines will be referred to by "ΔSRTT" and "ΔSRBT," respectively.

In summary, it can be concluded that the effect of presentation level needs to be taken into account to get a clear view on the influence of suprathreshold deficits on masking release for speech in modulated maskers. For the present group of mild to moderate hearing-impaired listeners, the deterioration of spectral resolution does not qualify as an actual suprathreshold deficit, since it can be accounted for fully by increased presentation level. In contrast, temporal resolution does qualify as an actual suprathreshold deficit.

C. Relations between suprathreshold deficits and masking release (SRT)

In the previous section, it was shown that presentation level has a large effect on possible manifestations of suprathreshold deficits and that this effect should be taken into account before the relation with masking release can be studied. After applying the necessary corrections, it will now be investigated to which extent the intersubject differences in masking release (Δbenefit) can be accounted for by actual suprathreshold deficits.

Table IV presents correlations between Δbenefit and the four predictor variables ΔF, ΔSRBT, ΔT, and ΔSRTT, both in normal and adapted spectral mode. A log-transformation was necessary to provide for normality of ΔT. Age and pure-tone average (PTA) are included as additional predictor variables, the latter being defined as the average pure-tone hearing threshold at octave frequencies 0.5, 1.0, and 2.0 kHz.

In normal spectral mode, the data in Table IV show that temporal resolution and SRTT are highly correlated with masking release from modulated noise (or Δbenefit). Age seems to have an effect too, but, since the normal-hearing listeners as a group were considerably younger than the hearing-impaired group, this correlation may be artificial. It is therefore better to consider only data from hearing-impaired listeners, where none of the mentioned correlations are significant anymore. This might have been brought about by interindividual differences in the audibility of parts of the signal in normal spectral mode. Therefore, correlations with data from the adapted spectral mode are expected to give a better estimation of relevant suprathreshold effects, since in the SRT-a, differences in audibility were minimized by amplifying the signal to levels well above threshold for all frequencies.

TABLE IV. Product-moment correlations between Δbenefit and ΔF, ΔSRBT, ΔT, ΔSRTT, age, and pure-tone average (PTA), in both normal and adapted spectral mode, calculated for all participants (left, $N=53$) and for the hearing-impaired listeners only (right, $N=29$). Since no normal-hearing SRBT data are available, the ΔSRBT correlations displayed in italics are calculated including only SIM and HI data ($N=45$). Correlations indicated with an asterisk are significant. $(*)p<0.05$ $(**)p<0.01$.

	All		Only HI	
	Δbenefit (normal)	Δbenefit (adapted)	Δbenefit (normal)	Δbenefit (adapted)
ΔF	0.09	0.10	0.05	0.00
ΔSRBT	*0.20*	*0.38*$^{(**)}$	−0.02	0.22
log(ΔT)	0.44$^{(**)}$	0.76$^{(**)}$	0.23	0.70$^{(**)}$
ΔSRTT	0.50$^{(**)}$	0.66$^{(**)}$	0.16	0.41$^{(*)}$
Age	0.57$^{(**)}$	0.81$^{(**)}$	0.29	0.63$^{(**)}$
PTA	0.19	0.46$^{(**)}$	0.03	0.59$^{(**)}$

Table IV shows that in adapted spectral mode, temporal resolution, SRTT, and age are significantly related to Δbenefit. Spectral resolution is not significantly correlated, and SRBT is only significantly correlated when all listeners are included, the correlation disappearing when looking at data from hearing-impaired listeners only. This means that temporal resolution and SRTT appear to be the key factors governing speech unmasking in modulated noise, while spectral resolution and SRBT have little effect. Furthermore, the correlation with age, even when considering only data from hearing-impaired listeners, indicates that understanding speech in modulated noise deteriorates with age. Finally, PTA appears to be significantly correlated to speech intelligibility in modulated noise.

However, the correlation analysis performed on speech unmasking by modulations, as presented earlier, is not ideal. It gives a distorted picture, because the chosen predictor variables are cross-correlated, as can be seen in Table V. This may lead to induced correlations between some predictor variables and speech unmasking. PTA, for instance, is highly correlated with temporal acuity ΔT and might therefore only contribute in predicting speech unmasking via ΔT.

To control for these cross-correlations, a stepwise multiple regression analysis was performed on the data from hearing-impaired listeners ($N=29$). Reported in the following are the successive contributing predictor variables and the coefficient of determination R^2, corrected for the available degrees of freedom.

The results from the stepwise regression analysis show that in normal spectral mode, as seen earlier, none of the tests contributes significantly to explaining the variance in Δbenefit, possibly due to large interindividual differences in the audibility of the signal. In adapted spectral mode, temporal acuity is the most significant term ($p<0.0001$), accounting for 46% of the variance in Δbenefit, while age is the second most significant ($p=0.0003$), explaining 35% of

TABLE V. Product-moment cross-correlations between ΔF, ΔSRBT, ΔT, ΔSRTT, age, and pure-tone average (PTA). Bold values on the diagonal are autocorrelations between test and retest values. Only HI data were included in the calculations ($N=29$). Correlations indicated with an asterisk are significant. $(*)p<0.05$ $(**)p<0.01$.

	ΔF	ΔSRBT	$\log(\Delta T)$	ΔSRTT	Age
ΔF	**0.81**				
ΔSRBT	$0.68^{(**)}$	**0.85**			
$\log(\Delta T)$	0.24	$0.41^{(*)}$	**0.93**		
ΔSRTT	0.35	$0.54^{(**)}$	0.30	**0.61**	
Age	0.12	0.11	0.32	$0.43^{(*)}$	⋯
PTA	0.32	$0.57^{(**)}$	$0.81^{(**)}$	0.32	0.30

the variance on its own. When both age and $\log(\Delta T)$ are included in the regression model, together they account for 64% of the variance in Δbenefit, while the other predictor variables do not contribute significantly ($p \geqslant 0.08$). Note that both ΔSRTT and PTA are not included anymore, indicating that, indeed, their correlation with Δbenefit was mainly induced by cross-correlations with ΔT and age. Finally, it is worth mentioning that the 64% of explained variance may be an underestimate, considering the nonunity test-retest correlations of the predictor variables.

The large contribution of temporal acuity to explaining variance in masking release is in line with our expectations and with literature (Glasberg *et al.*, 1987; Festen and Plomp, 1990; Glasberg and Moore, 1992; Festen, 1993; Dubno *et al.*, 2003). The fact that, apart from hearing threshold, age is a significant factor, is also consistent with earlier findings (e.g., Gustafsson and Arlinger, 1993; Snell *et al.*, 2002). Increasing evidence exists in literature that age can affect the temporal processing of sounds, independent of hearing loss (Snell, 1997; Grose *et al.*, 2001; Dubno *et al.*, 2002; Gordon-Salant and Fitzgibbons, 2004; Gifford and Bacon, 2005). This effect might be related to the influence of cognitive effects on hearing ability, as suggested by, e.g., Pichora-Fuller *et al.* (1995), Watson *et al.* (1996), Gordon-Salant and Fitzgibbons (1997), and Gatehouse *et al.* (2003), or to other nonperipheral effects like, for instance, the suppression of neuronal envelope locking, as suggested by Las *et al.* (2005). These effects may also play a role in the remaining unexplained part of the variance. At present, a new experiment is being conducted to further investigate this possibility.

The fact that the SRBT and the SRTT do not contribute significantly to explaining variance in masking release may indicate that the ability to use the redundancy of speech (at the acoustic, phonetic or lexical level) only plays a minor role when it comes to obtaining benefit from modulated maskers, even though the correlation of SRTT with Δbenefit ($r=0.41$) is significant. The small test-retest reliability of the SRTT ($r=0.61$) might be another reason why the SRTT does not significantly contribute. Moreover, the ability to restore incomplete speech, as measured by the SRTT and the SRBT, may be cognitive in nature, and may already have been included in the effect of age, see the significant relation between SRTT and age ($r=0.43$) in Table V.

The above noted findings constitute the main result from this experiment: temporal resolution and age are the essential contributors to explain the interindividual differences in speech unmasking in modulated noise. Combined, they account for 64% of the variance in Δbenefit.

D. From SRT to SII

In the previous section, each listener's benefit in SRT when listening to speech in modulated noise as compared to stationary noise was calculated. After taking the effect of presentation level into account, Δbenefit has been used as a measure for the unmasking of speech in modulated noise. A major drawback of this approach lies in the fact that each listener has a different audiogram, giving rise to intersubject differences in audibility. In the adapted spectral mode, it was attempted to overcome this drawback by shaping the spectrum of both signal and masker according to the shape of the individual hearing threshold. However, this in turn means that each participant listened to a different spectrum. Using the SRT to quantify the subjects' ability to perceive speech in noise does not take these intersubject audiogram and spectrum differences into account.

A measure of speech intelligibility performance which is able to handle intersubject audiogram and spectrum differences is the Speech Intelligibility Index or SII (ANSI S3.5-, 1997), which gives an estimate of the amount of speech information available in a certain condition, using the individual's audiogram and the signal and masker spectrum levels as inputs. Since the SII model already accounts for audiogram and spectrum differences among listeners, the calculated SII value needs no correction for presentation level anymore, as will be shown in the following.

A SII of about 0.30–0.35 is commonly considered to be enough to reach 50% speech intelligibility for normal-hearing listeners. Hearing-impaired listeners generally need more speech information, which is supposed to be caused by less efficient handling of the information due to suprathreshold deficits. Therefore, the amount that the SII is raised may be a good measure of the effect of suprathreshold deficits on speech intelligibility.

The SII has been extensively validated for stationary masking noise and recently, Rhebergen and Versfeld (2005) proposed an extension to the model, which makes it also applicable to fluctuating background maskers. Their approach gives a good account for most existing data, producing SIIs around 0.35 for normal-hearing listeners. A slightly modified version of their model has been used here to translate the measured SRT values in SII values. Validation and details of this model can be found in the Appendix.

E. Relations between suprathreshold deficits and masking release (SII)

Figure 7 gives an overview of the calculated SII values for individual listeners in both spectral modes. SII_{MOD} is the average SII value over three 16-Hz modulated masker conditions (see the Appendix), while SII_{STAT} is the SII calculated for stationary noise. The figure shows that normal-hearing participants, as expected, display SII values of around 0.3 in both stationary and modulated noise. Listeners with simulated hearing loss, experiencing only audibility

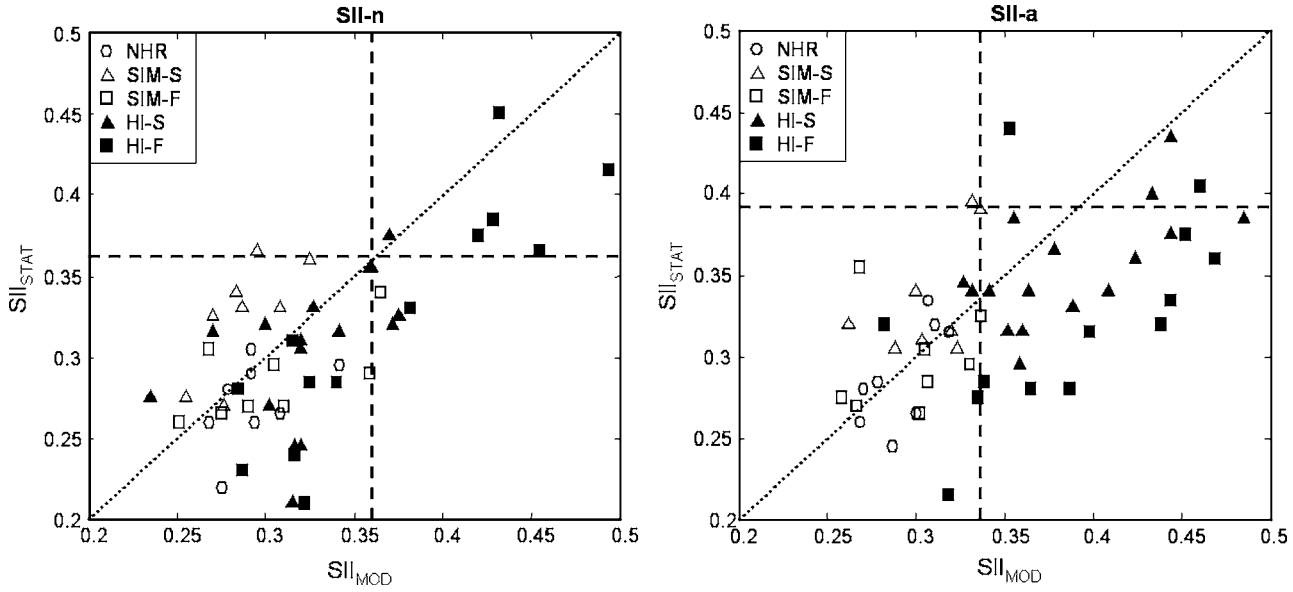


FIG. 7. SII in stationary noise vs SII in modulated noise, the latter defined as the mean SII in three 16-Hz modulated noise conditions, in both normal (left panel) and adapted (right panel) spectral mode. Data are displayed for normal-hearing listeners (NHR), two groups of hearing-impaired listeners (HI-S and HI-F), and two groups of listeners with simulated hearing loss (SIM-S and SIM-F). Dashed lines represent the one-tailed 95th percentiles for all open symbols.

problems due to their artificially elevated threshold, also display a SII around 0.3 in both cases. Although this was to be expected, it forms a strong contrast with the obtained elevated SRT in this group. This demonstrates the power of applying the SII method in fluctuating noise, which appears to adequately incorporate problems due to only an elevated threshold.

Many hearing-impaired listeners show suprathreshold problems in stationary noise and modulated noise, as their SIIs are raised compared to normal-hearing listeners. The difference between SII_{MOD} and SII_{STAT} can then be regarded as the extent to which an individual listener experiences additional disadvantages when listening to speech in modulated noise compared to stationary noise. Since audibility is already incorporated in the SII, the difference in SII between stationary and modulated maskers (i.e., SII_{DIFF}) seems an appropriate measure for the experienced problem in speech unmasking in modulated noise due to suprathreshold deficits.

Next, it will be investigated to which extent these differences in SII can be accounted for by actual suprathreshold deficits. Table VI gives correlations between SII_{DIFF} and the predictor variables ΔF, ΔSRBT, ΔT, ΔSRTT, age, and pure-tone average (PTA), in both normal and adapted spectral mode. These results are shown to be comparable to the correlations found between Δbenefit and the six predictor variables, as displayed in Table IV. When, as before to prevent artificial correlations, only data from hearing-impaired listeners are considered, temporal resolution, age, and PTA have the largest effects on speech unmasking, being significantly correlated to SII_{DIFF} in adapted spectral mode. In normal spectral mode, only PTA is significantly relevant.

These results can be regarded as a confirmation of our earlier conclusion that temporal resolution and age are the key factors governing speech unmasking in modulated noise. As before, stepwise multiple regression analyses were performed on the data from hearing-impaired listeners to control for the cross-correlations between the predictor variables (see Table V).

Results from these analyses show that in normal spectral mode, only PTA significantly contributes ($p=0.02$), accounting for 14% of the variance in SII_{DIFF}. In adapted spectral mode, temporal resolution is the most significant term ($p=0.0002$), accounting for 36% of the variance in SII_{DIFF}, while age is second most significant ($p=0.0006$), accounting for 31% of the variance on its own. When including both predictor variables, they account for 53% of the variance in SII_{DIFF}. As before with SRT, ΔF does not contribute on its own ($p=0.66$), but, rather surprisingly, becomes significant ($p=0.03$) when both temporal resolution and age are included first. When ΔF is included together with ΔT and age, a total of 59% of the variance in SII_{DIFF} can be explained. As

TABLE VI. Product-moment correlations between SII_{DIFF} and ΔF, ΔSRBT, ΔT, ΔSRTT, age, and PTA, in both normal and adapted spectral mode, calculated for all participants (left, $N=53$) and for hearing-impaired listeners only (right, $N=29$). Since no normal-hearing SRBT data are available, the ΔSRBT correlations displayed in italics are calculated including only SIM and HI data ($N=45$). Correlations indicated with an asterisk are significant. $(*)p<0.05$ $(**)p<0.01$.

	All		Only HI	
	SII_{DIFF} (normal)	SII_{DIFF} (adapted)	SII_{DIFF} (normal)	SII_{DIFF} (adapted)
ΔF	0.09	0.07	0.00	−0.09
ΔSRBT	*0.19*	*0.26*	0.08	0.15
$\log(\Delta T)$	$0.43^{(**)}$	$0.67^{(**)}$	0.31	$0.64^{(**)}$
ΔSRTT	0.26	$0.50^{(**)}$	−0.03	0.28
Age	$0.44^{(**)}$	$0.68^{(**)}$	0.25	$0.60^{(**)}$
PTA	$0.28^{(*)}$	$0.37^{(**)}$	$0.44^{(*)}$	$0.54^{(**)}$

in the SRT case, the other predictor variables do not significantly contribute anymore to the multiple regression ($p \geqslant 0.08$).

These findings are consistent with the earlier SRT results, although the small but significant contribution of spectral resolution is surprising. These SII results, however, corroborate our earlier conclusion that temporal resolution and age are the essential contributors to explaining unmasking of speech in modulated background maskers.

F. Difference between normal and adapted spectral mode

In the previous section on SII results, it is striking that, in normal spectral mode, no predictor variables, except for PTA, have a significant effect on speech unmasking in modulated noise, while they do in adapted spectral mode. In the earlier discussion on the SRT results, this was explained as a consequence of large interindividual differences in audibility of part of the signal in normal spectral mode. However, this argument is not valid in the present discussion on the SII results, since audibility differences are already accounted for by the SII calculations.

An alternative explanation for the large difference between SII-n and SII-a results can be found when looking back to Fig. 7. In normal spectral mode, SII values for most hearing-impaired listeners are distributed fairly close to normal-hearing data, indicating only small differences between SII_{STAT} and SII_{MOD}. Only 6 hearing-impaired listeners display a higher-than-normal SII in stationary noise, indicated by a position above the horizontal dashed line. A total of 10 hearing-impaired listeners, including these 6, display a raised SII in modulated noise. In adapted spectral mode, however, most hearing-impaired SII values deviate from normal-hearing data. Moreover, SII_{MOD} is larger than SII_{STAT} for most hearing-impaired listeners. Only 4 hearing-impaired listeners display a raised SII in stationary noise, but no less than 24 out of 29 hearing-impaired listeners display a higher-than-normal SII_{MOD}, indicating that they suffer from suprathreshold deficits when listening to speech in modulated noise.

First, these observations indicate that problems in speech intelligibility due to suprathreshold deficits are more prominent in modulated maskers than in stationary maskers. That is, a listener suffering from suprathreshold deficits will experience more problems understanding speech in modulated maskers than in stationary noise.

Second, these observations indicate that differences in speech intelligibility due to suprathreshold deficits are more prominent in adapted spectral mode than in normal spectral mode. Since differences between stimuli of the two spectral modes mainly occur at the higher frequencies, this finding suggests that mainly the processing of high-frequency stimulus components is affected by suprathreshold deficits, whereas the processing of lower-frequency signals remains relatively unaffected. These findings are consistent with earlier results by Apoux and Bacon (2004), who show that normal-hearing listeners rely more on the higher frequency regions when it comes to the processing of temporal information of speech in noise. Moreover, this result is in agreement with indications by Hogan and Turner (1998), who found that hearing-impaired listeners used the information in higher frequencies less efficiently than normal-hearing listeners, dependent on their degree of hearing loss. Turner and Brus (2001) later suggested that this effect may be linked to the types of speech cues residing in the different regions of the speech spectrum and the consequences of hearing loss upon the transmission of these cues.

In summary, it can be concluded that intersubject differences in speech understanding due to suprathreshold deficits are more pronounced in modulated noise than in stationary noise. Moreover, these differences are more prominent in adapted spectral mode than in normal spectral mode, indicating that they are mainly related to the processing of higher stimulus frequencies.

IV. CONCLUSIONS

The main results arrived at in the previous sections can be summarized as follows:

I. Normal-hearing listeners with only an audibility-problem (i.e., an artificially raised hearing threshold) experience a reduced masking release for speech in modulated noise. Therefore, to study the effect of suprathreshold deficits, it is necessary to take the threshold-related effect of presentation level on masking release into account (Sec. III A).

II. To get a clear view on the influence of suprathreshold deficits, the effect of presentation level on their manifestations needs to be taken into account. For the present group of mild to moderate hearing-impaired listeners, the observed deteriorated spectral resolution does not qualify as a suprathreshold deficit, since it can be accounted for fully by increased presentation level. In contrast, the observed deteriorated temporal resolution does qualify as an actual suprathreshold deficit (Sec. III B).

III. Temporal resolution and age are the essential contributors to explaining the interindividual differences in masking release for speech in modulated background maskers. Combined, they account for more than half of the intersubject variance. Results based on SII calculations confirm this conclusion (Secs. III C and III E).

IV. Applying the SII method in modulated noise (cf. Rhebergen and Versfeld, 2005) adequately incorporates problems due to an elevated threshold, giving rise to values around 0.3 for normal-hearing listeners with and without an elevated threshold, and larger values for hearing-impaired listeners with suprathreshold deficits (Sec. III D / Appendix).

V. Intersubject differences in speech intelligibility due to suprathreshold deficits are more pronounced in modulated noise than in stationary noise (Sec. III F).

VI. Intersubject differences in speech intelligibility due to suprathreshold deficits are more pronounced in adapted spectral mode than in normal spectral mode, indicating that they are mainly related to the processing of higher stimulus frequencies (Sec. III F).

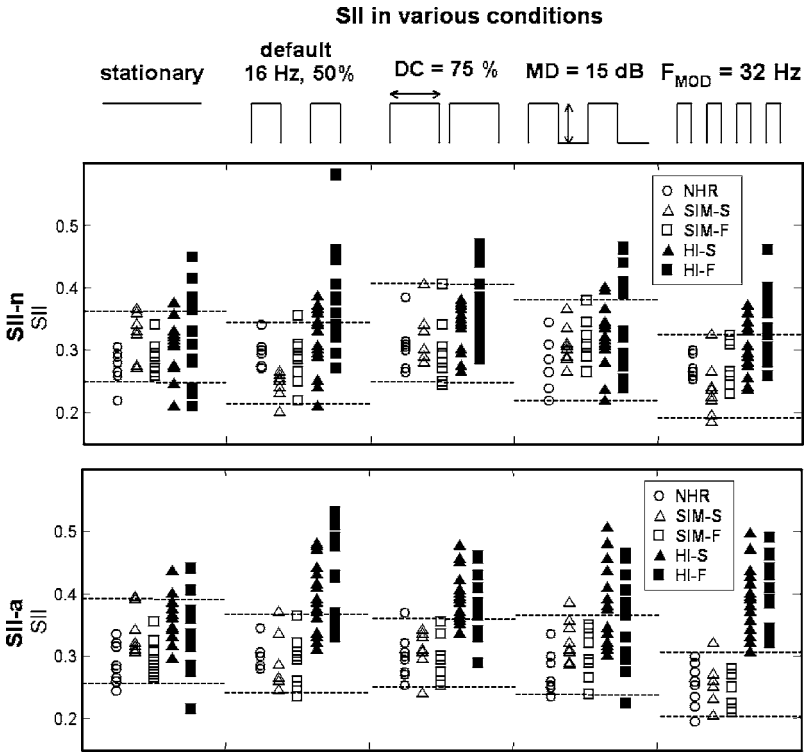


FIG. 8. SII-values in stationary noise and in four modulated background noise conditions, in both normal (upper panel) and adapted (lower panel) spectral mode. Data are displayed for normal-hearing listeners (NHR), two groups of hearing-impaired listeners (HI-S and HI-F), and two groups of listeners with simulated hearing loss (SIM-S and SIM-F). Dashed lines represent the 5th and 95th percentiles of the open symbols. Participants in each of the five different groups have been assigned a different displacement along the *x* axis for visual clarity.

ACKNOWLEDGMENTS

This research was supported by the Heinsius-Houbolt Foundation, The Netherlands. Thanks are due to Gaston Hilkhuysen for providing the tests of temporal and spectral acuities and to two anonymous reviewers for their comments on a previous version of this manuscript.

APPENDIX: SII IN FLUCTUATING NOISE

To be able to predict speech intelligibility in the presence of background noise, the Speech Intelligibility Index or SII was introduced in 1997 (ANSI S3.5-1997), as a revision of the Articulation Index as introduced by French and Steinberg (1947) and Kryter (1962). It calculates the total amount of speech information available to the listener, using the speech spectrum, the noise spectrum, and the listener's hearing threshold as inputs. The SII has been extensively validated for stationary masking noises, but fails to predict speech intelligibility in fluctuating background.

Recently, however, Rhebergen and Versfeld (2005) proposed an extension to the model, which takes the fluctuations in background noise into account. The basic principle of their approach is the partitioning of both speech and noise in small time frames. In each time frame, speech and noise are filtered into 21 critical bands to determine the "local" spectra, which serve as inputs to determine the "local" conventional SII. Next, the SII values of all time frames are averaged to result in the overall SII for that particular speech in noise condition. Their approach gives a good account for most existing data, producing SIIs around 0.35 for normal-hearing listeners.

Using their model, SII values were calculated for the various speech-in-noise-conditions, using the speech and noise signals which were employed in our experiment. A slight modification was made concerning the filtering of the signals in critical bands, using FIR filters with order $N_i = 0.05^* f_i$, where f_i is the center frequency of the *i*th critical band in hertz, thus assuring enough resolution in the frequency domain to provide accurate filtering.

This approach, however, yielded SII values of only between 0.04 and 0.20 for normal-hearing listeners in fluctuating backgrounds, much lower than was expected. Therefore, the width of the time frames was adapted to range from 5.8 ms in the lowest band (150 Hz) to 1.9 ms in the highest band (8000 Hz), instead of the values between 35 and 9.4 ms as mentioned by Rhebergen and Versfeld (2005). These window lengths were taken from Moore *et al.* (1993), but scaled (with a ratio of 0.5) to be in accordance with the average temporal resolution for normal-hearing at 1000 Hz, which was measured to have a value of 4.36 ms.

Figure 8 shows the calculated SII values in both adapted and normal spectral mode, in stationary noise and in four modulated background maskers. The figure shows that the calculated SII appears to be a stable measure, since listeners in both the normal-hearing and the SIM groups display a SII of about 0.3 in most conditions. This demonstrates the power of applying the proposed SII method in fluctuating noise, which appears to adequately incorporate problems due to only an elevated threshold.

The calculated SII for normal-hearing does, however, decrease slightly when the modulation frequency of the masker increases from the default 16 to 32 Hz. Therefore, SII values obtained in the 32-Hz condition were not included in discussing our experimental results. Further development of the SII model will have to aim at solving the small deviations for higher modulation frequencies.

SII_{MOD}, as used in presenting our experimental results, is thus defined as each participant's average SII over the three

16-Hz modulated masker conditions, reaching a value of around 0.3 for all normal-hearing and simulated hearing loss listeners, represented by the open symbols. As can be seen in Fig. 8, hearing-impaired listeners generally display a larger SII, especially in adapted spectral mode. It can therefore be concluded that a SII value of about 0.3 indicates that the listener has no auditory deficits, apart from possibly an elevated threshold, while an increased SII can serve as an indication of a suprathreshold deficit in a particular condition.

[1]Prior to collapsing, the individual conditions were investigated by performing a repeated-measures ANOVA on the data from all hearing-impaired participants. This analysis showed that the effects of none of the independent variables on masking release change over the various conditions, indicating that collapsing over conditions when investigating these effects may be considered reasonable.

ANSI (**1997**). ANSI S3.5-1997, "American national standard methods for the calculation of the Speech Intelligibility Index," American National Standards Institute, New York.

Apoux, F., and Bacon, S. P. (**2004**). "Relative importance of temporal information in various frequency regions for consonant identification in quiet and in noise," J. Acoust. Soc. Am. **116**, 1671–1680.

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.

Baer, T., and Moore, B. C. J. (**1993**). "Effects of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am. **94**, 1229–1241.

Baer, T., and Moore, B. C. J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.

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.

Celmer, R. D., and Bienvenue, G. R. (**1987**). "Critical bands in the perception of speech signals by normal and sensorineural hearing loss listeners," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Dordrecht).

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2003**). "Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal-hearing," J. Acoust. Soc. Am. **113**, 2084–2094.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2002**). "Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing," J. Acoust. Soc. Am. **111**, 2897–2907.

Dubno, J. R., and Schaefer, A. B. (**1992**). "Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners," J. Acoust. Soc. Am. **91**, 2110–2121.

Duifhuis, H. (**1973**). "Consequences of peripheral frequency selectivity for nonsimultaneous masking," J. Acoust. Soc. Am. **54**, 1471–1488.

Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," J. Speech Hear. Res. **38**, 222–233.

Fabry, D. A., and Van Tasell, D. J. (**1986**). "Masked and filtered simulation of hearing loss: Effects on consonant recognition," J. Speech Hear. Res. **29**, 170–178.

Festen, J. M., and Plomp, R. (**1981**). "Relations between auditory functions in normal hearing," J. Acoust. Soc. Am. **70**, 356–369.

Festen, J. M. (**1993**). "Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice," J. Acoust. Soc. Am. **94**, 1295–1300.

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.

Fitzgibbons, P. J. (**1983**). "Temporal gap detection in noise as a function of frequency, bandwidth, and level," J. Acoust. Soc. Am. **74**, 67–72.

Fitzgibbons, P. J., and Gordon-Salant, S. (**1987**). "Temporal gap resolution in listeners with high-frequency sensorineural hearing loss," J. Acoust. Soc. Am. **81**, 133–137.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–66.

Florentine, M., and Buus, S. (**1984**). "Temporal gap detection in sensorineural and simulated hearing impairments," J. Speech Hear. Res. **27**, 449–455.

Füllgrabe, C., Berthommier, F., and Lorenzi, C. (**2006**). "Masking release for consonant features in temporally fluctuating background noise," Hear. Res. **211**, 74–84.

French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.

Gatehouse, S., Naylor, G., and Elberling, C. (**2003**). "Benefits from hearing aids in relation to the interaction between the user and the environment," Int. J. Audiol. **42**, S77–85.

Gifford, R. H., and Bacon, S. P. (**2005**). "Psychophysical estimates of nonlinear cochlear processing in younger and older listeners," J. Acoust. Soc. Am. **118**, 3823–3833.

Glasberg, B. R., and Moore, B. C. J. (**1992**). "Effects of envelope fluctuations on gap detection," Hear. Res. **64**, 81–92.

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.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1997**). "Selected cognitive factors and speech recognition performance among young and elderly listeners," J. Speech Lang. Hear. Res. **40**, 423–431.

Gordon-Salant, S., and Fitzgibbons, P. J. (**2004**). "Effects of stimulus and noise rate variability on speech perception by younger and older adults," J. Acoust. Soc. Am. **115**, 1808–1817.

Grose, J. H., Hall, J. W., and Buss, E. (**2001**). "Gap duration discrimination in listeners with cochlear hearing loss: Effects of gap and masker duration, frequency separation, and mode of presentation," J. Assoc. Res. Otolaryngol. **2**, 388–398.

Gustafsson, H. A., and Arlinger, S. D. (**1993**). "Masking of speech by amplitude-modulated noise," J. Acoust. Soc. Am. **95**, 518–529.

Hall, J. W., Davis, A. C., Haggard, M. P., and Pillsbury, H. C. (**1988**). "Spectro-temporal analysis in normal-hearing and cochlear-impaired listeners," J. Acoust. Soc. Am. **84**, 1325–1331.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.

Hawkins, J. E., and Stevens, S. S. (**1950**). "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Am. **22**, 6–13.

Healy, E. W., and Bacon, S. P. (**2006**). "Measuring the critical band for speech," J. Acoust. Soc. Am. **119**, 1083–1091.

Hilkhuysen, G. L. M., Houtgast, T., and Lyzenga, J. (**2005**). "Estimating cochlear-filter shapes, temporal-window width and compression from tone-sweep detection in spectral and temporal noise gaps," J. Acoust. Soc. Am. **117**, 2598–2599.

Hogan, C. A., and Turner, C. W. (**1998**). "High-frequency audibility: Benefits for hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 432–441.

Howard-Jones, P. A., and Rosen, S. (**1993**). "Uncomodulated glimpsing in "checkerboard" noise," J. Acoust. Soc. Am. **93**, 2915–2922.

Humes, L. E., Dirks, D. D., Bell, T. S., and Kincaid, G. E. (**1987**). "Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers," J. Acoust. Soc. Am. **81**, 765–773.

Humes, L. E., Espinoza-Varas, B., and Watson, C. S. (**1988**). "Modeling sensorineural hearing loss. I. Model and retrospective evaluation," J. Acoust. Soc. Am. **83**, 188–202.

Hygge, S., Ronnberg, J., Larsby, B., and Arlinger, S. (**1992**). "Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds," J. Speech Hear. Res. **35**, 208–215.

International Organization for Standardization (**1991**). ISO 389:1991(E), "Acoustics–Standard reference zero for the calibration of pure-tone air conduction audiometers," (available from the American National Standards Institute, New York).

Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). "Forward masking as a function of frequency, masker level, and signal delay," J. Acoust. Soc. Am. **71**, 950–962.

Kryter, K. R. (**1962**). "Methods for the calculation and use of the Articulation Index," J. Acoust. Soc. Am. **34**, 1689–1697.

Las, L., Stern, E. A., and Nelken, I. (**2005**). "Representation of tone in fluctuating maskers in the ascending auditory system," J. Neurosci. **25**, 1503–1513.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lochner, J. P. A., and Burger, J. F. (**1961**). "Form of the loudness function in the presence of masking noise," J. Acoust. Soc. Am. **33**, 1705–1707.

Ludvigsen, C. (**1985**). "Relations among some psychoacoustic parameters in normal and cochlearly impaired listeners," J. Acoust. Soc. Am. **78**, 1271–1280.
Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1993**). "Detection of temporal gaps in sinusoids: Effects of frequency and level," J. Acoust. Soc. Am. **93**, 1563–1570.
Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). "Inter-relationship between different psycho-acoustic measures assumed to be related to the cochlear active mechanism," J. Acoust. Soc. Am. **106**, 2761–2778.
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**, 962–968.
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.
Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**1999**). "Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. I. Normal-hearing listeners," J. Acoust. Soc. Am. **105**, 2895–2902.
Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**2000**). "Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. II. Hearing-impaired listeners," J. Acoust. Soc. Am. **107**, 1685–1696.
Oxenham, A. J., and Bacon, S. P. (**2003**). "Cochlear compression: Perceptual measures and implications for normal and impaired hearing," Ear Hear. **24**, 352-366.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788–1803.
Pavlovic, C. V. (**1987**). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," J. Acoust. Soc. Am. **82**, 413–422.
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.
Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). "How young and old adults listen to and remember speech in noise," J. Acoust. Soc. Am. **97**, 593–608.
Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.
Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.
Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.
Snell, K. B. (**1997**). "Age-related changes in temporal gap detection," J. Acoust. Soc. Am. **101**, 2214–2220.
Snell, K. B., Mapes, F. M., Hickman, E. D., and Frisina, D. R. (**2002**). "Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity," J. Acoust. Soc. Am. **112**, 720–727.
Sommers, M. S., and Humes, L. E. (**1993a**). "Auditory filter shapes in normal-hearing, noise-masked normal and elderly listeners," J. Acoust. Soc. Am. **93**, 2903–2914.
Sommers, M. S., and Humes, L. E. (**1993b**). "Erratum: Auditory filter shapes in normal-hearing, noise-masked normal and elderly listeners [J. Acoust. Soc. Am. **93**, 2903–2914 (1993)]," J. Acoust. Soc. Am. **94**, 2449–2450.
Stevens, S. S. (**1966**). "Power-group transformations under glare, masking and recruitment," J. Acoust. Soc. Am. **39**, 725–735.
Stuart, A., and Phillips, D. P. (**1996**). "Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing and presbyacusic listeners," Ear Hear. **17**, 478–489.
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.
ter Keurs, M., Festen, J. M., and Plomp, R. (**1993a**). "Effect of spectral envelope smearing on speech reception. II," J. Acoust. Soc. Am. **93**, 1547–1552.
ter Keurs, M., Festen, J. M., and Plomp, R. (**1993b**). "Limited resolution of spectral contrast and hearing loss for speech in noise," J. Acoust. Soc. Am. **94**, 1307–1314.
Turner, C. W., and Brus, S. L. (**2001**). "Providing low- and mid-frequency speech information to listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **109**, 2999–3006.
Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). "Method for the selection of sentence materials for efficient measurement of speech reception threshold," J. Acoust. Soc. Am. **107**, 1671–1684.
Versfeld, N. J., and Dreschler, W. A. (**2002**). "The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners," J. Acoust. Soc. Am. **111**, 401–408.
Wagener, K. C., and Brand, T. (**2005**). "Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters," Int. J. Audiol. **44**, 144–156.
Warren, R. M. (**1970**). "Perceptual restoration of missing speech sounds," Science **167**, 392–393.
Watson, C. S., Qiu, W. W., Chamberlain, M. M., and Li, X. (**1996**). "Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition," J. Acoust. Soc. Am. **100**, 1153–1162.
Xu, L., Thompson, C. S., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**, 3255–3267.
Zurek, P. M., and Delhorne, L. A. (**1987**). "Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment," J. Acoust. Soc. Am. **82**, 1548–1559.

Jet offset, harmonic content, and warble in the flute

John W. Coltman[a)]
3319 Scathelocke Road, Pittsburgh, Pennsylvania 15235

(Received 19 May 2006; revised 11 July 2006; accepted 12 July 2006)

The effects of jet offset in the flute, directing the jet above or below the edge, were explored by two distinct means—experiments with a Boehm flute sounded by an artificial blower, and time domain simulation. Very large changes in harmonic content and dynamics were observed, changing greatly with blowing pressure. Warble, a modulation of the tone at frequencies of the order of 20 Hz, was observed both in the experiment and in the simulation. The phenomenon is explained as a beat between the frequency of a second harmonic generated by nonlinearity in the jet current and a neighboring partial sustained by jet feedback near the second mode resonance. A second type of warble, in which amplitude modulation occurs in all partials but with different phases, is yet to be explained. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2266562]

PACS number(s): 43.75.Qr, 43.75.Bc [NFH] Pages: 2312–2319

I. INTRODUCTION

The effects of jet offset, the passage of the average position of the jet above or below the edge on which it plays, have been described by Nolle[1,2] for several varieties of organ pipes. Fletcher and Douglas[3] provided a simplified theoretical treatment for a single case, and verified in part its applicability by experiment. The flute, however, differs from the organ pipes studied by Nolle in several ways. The organ pipe geometry is nearly symmetrical with respect to offset above or below the edge. In the flute, when the jet is above the edge it goes into a space bounded only by the cylindrical lip plate, when it is below the edge, it goes into a quite restricted hole. The edge in an organ pipe is straight; in the flute the edge is curved so that jet waves arrive at different times on different portions of the edge. The diameter to length ratio, which has a substantial effect on mode spacing, also differs; Nolle's pipe that came closest to flute dimensions had a D/L ratio 2.5 times that of the flute. The taper in the headjoint, the cork cavity, and the cavities below closed keys on the flute also affect the mode spacing. Finally, the blowing pressure used in the organ pipes was about three times that typical of the flute, requiring a much larger lip-to-edge distance. All of these differences make it desirable to examine the effects of offset in the flute, especially since flutists and flute pedagogues are well aware that tone quality and response are affected by what they term "angle," the directing of the jet more or less into the embouchure hole. The investigation was carried out in two distinct ways; by use of the time domain simulation described by Coltman[4] and by experiment using a conventional Boehm flute with an artificial blower. In the course of this work, the phenomenon of "warble" was encountered. Warble is the self-modulation of the tone at frequencies in the neighborhood of 20–30 Hz. An explanation of the origin of this phenomenon is given.

[a)]Electronic mail: coltmanjw@verizon.net

II. TIME DOMAIN SIMULATION

A time-domain program (Coltman[4]) that simulates the behavior of the flute was used to explore the effects of jet offset on tonal and dynamic response of the flute. The program is oversimplified in several ways. Nolle[5] and Coltman[6] describe many anomalies in the jet shape which call into question the often-used constant jet width and shape that is independent of amplitude. The model does not take into account the losses associated with vortex shedding at the edge (Verge *et al.*[7]). Complete symmetry of the region surrounding the edge is also assumed, so that in this respect it corresponded more nearly to the case of the organ pipe. However, the mode stretch, lip-to-edge distance, and blowing pressure were characteristic of the flute. The program had earlier predicted quite well many experimentally known aspects of the response of the flute, and it is believed that the results reported here may be taken as representative of the qualitative behavior of the flute with respect to changes in offset.

The program has the advantage that various parameters such a jet delay, jet current, jet offset, etc., can be independently varied while leaving other parameters strictly fixed. The major variable here was jet offset, the amount by which the jet centerline (when not oscillating) is displaced from the edge. The units of these displacements are half-jet-widths, that is, a displacement of 1 means that the centerline of the nonoscillating jet would strike above the edge by an amount equal to one half the width of the undisturbed jet, so that it would just be blowing entirely outside the flute. A displacement somewhat less than 1 would have a small part of the undisturbed jet blowing below the edge and into the flute. Negative displacements are similar except that the jet blows mostly below the edge and into the flute. Flutists ordinarily do not use offsets above the edge, and speak of increasing the angle when blowing more deeply into the embouchure hole. With these definitions, such an increase in angle corresponds to a negative offset of increasing magnitude.

Figure 1 shows the harmonic composition of the radiated sound intensity level as a function of offset for the case where the effective blowing pressure was close to that which would sound the pipe at its passive resonance frequency. The

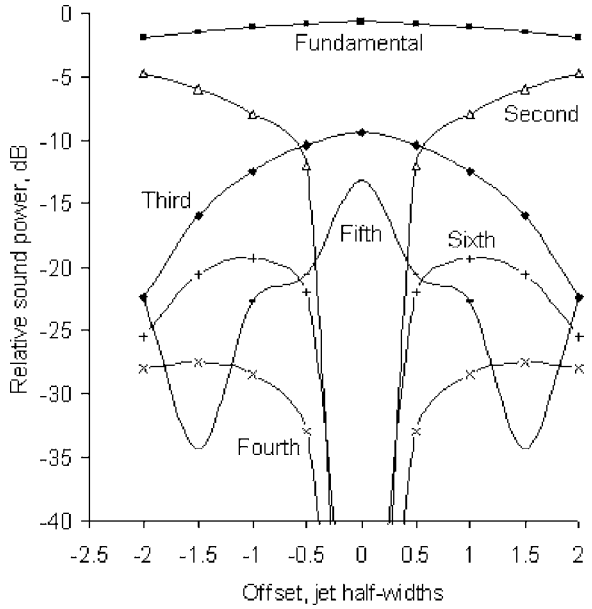


FIG. 1. Harmonic composition vs offset. Time-domain run with jet wave delay=5/16 cycle.

note imitated was first-register F4, nominal frequency 349.2 Hz. The numbers here represent the contribution of each harmonic to the total radiation intensity, e. g., a value of −3 dB would mean the harmonic provided half of the total radiated intensity. That total may of course vary substantially with offset, but a composition chart is more meaningful in presenting a measure of the tone quality. All such charts in this paper are similarly constructed. While the model generates the pipe current wave form, the radiation resistance goes as the square of the frequency, so the radiated power for a given harmonic of pipe current is proportional to the square of its frequency. The model assumes a reflection coefficient whose dissipation is independent of frequency. Accordingly the values of radiated power were also corrected for differences in resonance quality factors of the various modes, known from an earlier calculation for this fingering. The latter corrections were modest, less than 1 dB.

The most pronounced effect displayed is the increase of the even harmonics at the expense of the odd ones. With no offset, the wave form of the current (volume velocity) in the flute column is symmetrical, nearly triangular, and has essentially no even harmonics. Offset introduces even harmonics, especially the second. The effect here follows that predicted by Fletcher and Douglas[3] with respect to the increase from zero to substantial amounts as the offset increases. Positive and negative offsets give identical values, except for a phase shift of 180° for the even harmonics. However the results do not show the sharp minima at certain offsets in other harmonics that Fletcher and Douglas[3] would expect from the zeros in their theoretical driving function. The single experimental result shown by them exhibited no such minima either. The reason for this is that their theoretical driving pressure levels are plotted as a function of offset for a fixed, and rather small, jet amplitude. However, when the offset is changed, the jet amplitude will take up whatever new value balances the input energy with the dissipation losses, so that it is by no means fixed at the modest level that was assumed in deriving the driving function. Using the time domain model and attempting to keep the jet amplitude constant at the value specified for the driving function simply resulted in failure of the system to oscillate for all but the smaller values of offset.

TABLE I. Jet wave delay=5/16 cycle.

Jet offset	Frequency (Hz)	Radiated intensity	Second harmonic/ fundamental	Loudness (sones)
0	351.7	13.6	0	21
0.5	351.6	13.8	.08	28
1.0	351.2	13.8	.21	28
1.5	350.7	13.4	.35	27
2.0	350.0	12.4	.52	26

Table I examines in more detail the behavior of the important second harmonic as a function of jet offset. All other parameters, jet gain, jet wave delay, jet current, and tube length were fixed. Negative offsets give identical results, so only positive offsets are shown here.

The second column gives the frequency change as a function of offset. The change in frequency is at most a few cents, quite negligible. While some flutists state that the pitch goes flat as the angle increases, this is due to their covering more of the mouth hole as they roll the flute in to change the angle. The total radiated intensity in arbitrary units (column 3) is calculated from the sum of the radiated intensities of each of the harmonics. It will be seen that under these conditions the total radiated intensity does not change appreciably with offset. Column 4 relates the intensity of the second harmonic to that of the fundamental. The change in harmonic content with offset changes substantially the tone color or quality of the perceived sound. It also changes the perceived dynamic level—the ear ascribes a greater loudness to complex sounds than to simple ones. Quantitatively, the radiated intensities of the several harmonics were converted to phons (a measure of how the ear responds to various frequencies) and then to sones, a measure of perceived loudness. In making this calculation, the relative intensities were multiplied by a constant before converting to sound pressure levels (SPLs). The constant was chosen to bring the SPLs to the neighborhood of 70 dB re 20 μPa. The loudnesses in sones are given in column 5.

Table II is similar to Table I, except that the jet wave delay has been shortened to a value of 3/16 cycle rather than 5/16. This corresponds to increasing the blowing pressure by a factor of 2.8, while narrowing the lip aperture to keep the

TABLE II. Jet wave delay=3/16 cycle.

Jet offset	Frequency (Hz)	Radiated intensity	Second harmonic/ fundamental	Loudness (sones)
0	355.5	16.6	0	24
0.5	355.2	23.6	1.08	33
1	353.6	27.2	2.07	32
1.5	354.6	29.7	2.63	36
2.0	353.7	42.4	4.63	36

jet current unchanged. The increase in blowing pressure sharpens the frequency by about 18 cents, though again the frequency is little affected by changes in offset. The radiated intensity (column 3) is substantially increased over that in Table I, and in contrast is greatly affected by the amount of offset. This change is due primarily to the great strengthening of the harmonics, particularly the second. For offset=2, the internal pressure amplitude of the second harmonic is 75% of the amplitude of the fundamental. The factor of 4 due to the radiation resistance and a modest increase of 22% for the increased Q are responsible for the large values in column 4.

These large values result in large changes in tone quality and perceived loudness. To give some impression of a loudness change from 24 to 36 sones, that is approximately equivalent to three flutists playing the same note instead of one flutist. While the same decrease in jet delay could be obtained by leaving the pressure alone and decreasing the lip-to-edge distance from say 7 to 4.2 mm, such a change would also change the amount of embouchure hole coverage, flattening the pitch considerably. However one expects the increase in harmonic content due to offset to take place whatever the means of decreasing the jet delay.

At some short jet delay times and large offset, another phenomenon comes into play, as reported in Coltman.[4] Here the jet motion is a complex wave excited in part by the large second harmonic. Rather than simply switching up and down at the fundamental frequency, it actually crosses the edge four times in one cycle of the fundamental frequency, generating a second tone at twice the frequency of the fundamental. This is really a multiphonic, two tones generated simultaneously by feedback. The second tone may be locked in at just one octave above the first, and may be stronger in intensity. Complex waves on the jet have been treated in detail by Kaykayoglu and Rockwell[8] and mode-locking of such multiphonics has been treated by Fletcher.[9] The process is facilitated by mode stretch designed into the flute itself (Coltman[10]). This mode stretch, or progressive sharpening of the resonance frequencies of the upper modes, is provided largely by the taper in the headjoint. Many players use this multiphonic technique for most of the notes in the first register, giving what some call a rich tone. The ability of the jet to carry simultaneously waves of more than one frequency has often been neglected in theoretical treatments of jet blown instruments, but it can be an important feature in their operation.

For some very particular input conditions the phenomenon of warble was exhibited—a relatively slow but continuing modulation of the wave form. This will be treated in more detail in Sec. IV.

III. EXPERIMENTAL INVESTIGATION

Figure 2 shows a cross section of the artificial blower in position with respect to the flute. The blower was pivoted on two small wires that were concentric with the line through the center of the blowing slit at the exit face, so that rotation around this pivot changed the offset but did not change the distance from the lip to the edge. In order to prevent acoustic

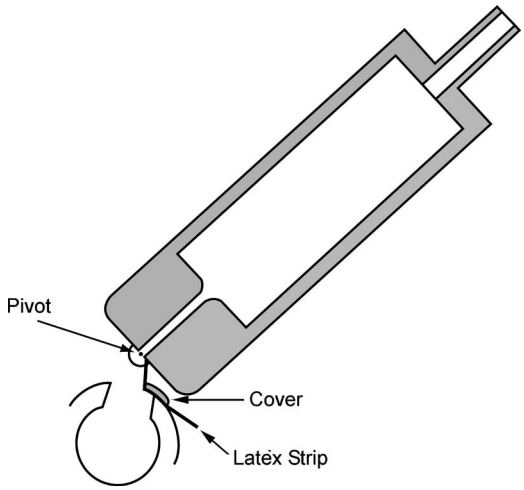


FIG. 2. Cross section of the artificial blower and flute embouchure.

current from escaping before exciting the jet, a thin latex strip 2 cm wide, shown edgewise in Fig. 2, was cemented to the face of the blower just below the slit. It was stretched to extend down below an adjustable cover that determined the coverage of the embouchure hole. This cover was set 8.8 mm from the edge for all the measurements reported here. The blower used was that employed earlier in measurements of jet profiles by Coltman.[6] The slit was rectangular with dimensions 9.9 by 1.2 mm. A scale was provided to indicate the angle of the blower. A small tube, not shown, permitted connection of the plenum to a Dwyer Magnehelic pressure meter. Blowing air was provided by a compressor from a household refrigerator, buffered by a holding tank. At any setting, the pressure was stable to better than 2 Pa.

Microphone placement poses some problems. The flute has two closely equal sound sources; the embouchure hole and the holes at the far end. They produce complex interference patterns that make pickup at any point where both can be heard highly dependent on position. Accordingly, a small tie-pin microphone (Radio Shack cat. #33 1062), 7.5 mm in diameter was rigidly mounted on the member supporting the flute opposite the center of the F hole, the first open hole for the fingering of the note F_4, diameter 14 mm. The face of the microphone was perpendicular to the tube axis, with its center 7 mm from the tube wall. In this position, the sound picked up by the microphone was found to be unaffected by variable reflections. A magazine waved as close as 4 in. from the microphone made no appreciable change in its output. There was a question whether the signal might be affected by streaming at the rather narrow opening between the hole wall and the key cover. A separate experiment was conducted in which the flute was driven by a dynamic driver while a second capillary microphone picked up the pressure inside the tube. The two microphone signal amplitudes remained closely proportional over a range of pressures much higher than those of the blown flute, so streaming was not a problem.

The amplified microphone signal was fed into the computer sound card for recording as a .wav file, using 8 bit amplitude and 22 050 samples/s. With the blowing pressure set, short recordings at each of some 8 or 10 offset positions

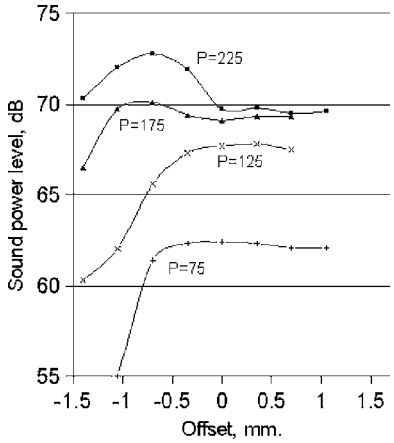


FIG. 3. Relative sound power level as a function of offset and plenum pressure (Pascals). Lip-to-edge distance 8.5 mm.

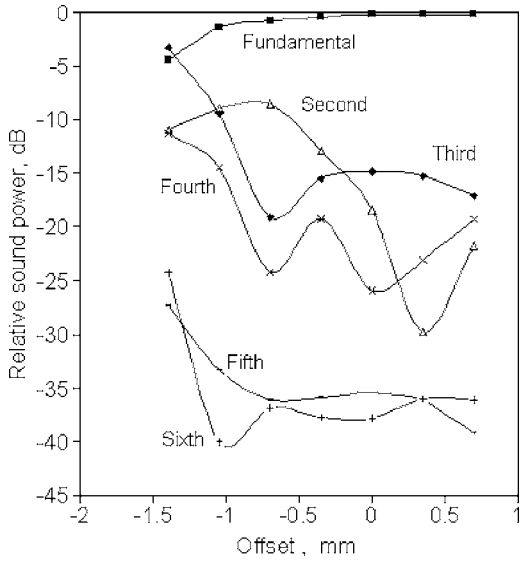


FIG. 4. Harmonic composition as a function of offset. Plenum pressure 125 Pa. Lip-to-edge distance 8.5 mm.

were made in quick succession. The recorded wave forms were examined for uniformity and lack of noise, and clean portions were selected for Fourier analysis.

Fourier analysis was carried out using the commercial mathematical program MATHCAD PLUS 5.0. A length of 4096 samples, corresponding to about 64 cycles of the fundamental frequency, was used for most analyses, with a Hanning function for apodization. While the position and height of the highest peak of a partial are often used as measures of the spectral composition of a tone, a more accurate measure was used here. The group of Fourier coefficients corresponding to a particular partial was isolated, and the sum of the squares of their amplitudes taken as the measure of intensity. Multiplying each squared coefficient by its frequency, and dividing their sum by the sum of the squares, returns an average frequency that is very precise, even for a relatively small number of samples. For example, the frequencies of a fundamental and its first harmonic, measured in this way from the recordings, typically departed by less than one part in 10 000 from a ratio of exactly 2.

A. Lip-to edge distance 8.5 mm

Two sets of measurements at two different lip-to-edge distances were made. In all of the measurements, the Reynolds numbers based on the slit height were within the range of 900–1600, so the jet is expected to be laminar rather than turbulent (Verge *et al.*[7]). The first set of measurements was taken with a lip-to-edge distance of 8.5 mm, corresponding to a ratio of distance to jet thickness of 7.1. The total radiated sound intensity level varied with plenum pressure and offset as shown in Fig. 3. In converting the total intensities to sound power levels a constant number of decibels was added to bring the SPLs to a value typical of what a listener might hear. Offsets here are given in millimeters rather than fractions of the jet width, which is not well known. The jet slot was 1.2 mm wide so a jet half -width may be assumed to be 0.6 mm or more. The asymmetry in this plot is marked; positive offsets made little change, while negative offsets (blowing more deeply into the embouchure hole) made substantial changes in the radiated SPL. Four sets of measurements, using plenum pressures of 75, 125, 175, and 225 Pa, were taken; only two are reported in detail here.

Figure 4 shows the harmonic composition of the radiated power measured at a plenum pressure of 125 Pa. This pressure is low, corresponding to what I use for the dynamic *piano* for this note. Flutists vary, of course, but measurements of several flutists indicate that my own technique is not far from average. The zero offset in this chart, and others taken in the experimental series, is not well defined—it is difficult to know exactly where the jet slit is pointing, so this may be off 0.2 mm or so. The real offset of the jet may differ from the point where the slit is pointing, as found by Nolle.[2] No attempt was made here to find the "real" zero of the jet offset. However, the positions of the minima of the even harmonics shown in Fig. 7 imply that the real offset zero occurs very close to the chosen zero. Changes in offset are accurate to better than 0.1 mm since the angle scale is well defined.

The behavior exhibited by Fig. 4 is far from symmetric, most of the large changes taking place in the negative half of the diagram. Particularly noteworthy is the behavior of the third harmonic, which rises steeply as the negative offset increases, so that at an offset of −1.5 mm it is the strongest partial. Substantial changes are evident in the fifth and sixth harmonics, though these are always so weak as to have a minor effect on tone quality. Note from Fig. 3 that the total power for this blowing pressure and −1.5 mm offset is some 8 dB down from that at zero offset.

Figure 5 gives results for a plenum pressure of 225 Pa, about the highest I would use in playing this note. For negative offset the second harmonic is dominant, slightly exceeding the fundamental, while the third harmonic is about 10 dB weaker than it was for the lower pressure. At the positive offset of 1.05 mm the flute sounded strongly with a large second harmonic, whereas with a blowing pressure of 125 Pa (Fig. 4) at this offset the note did not sound at all.

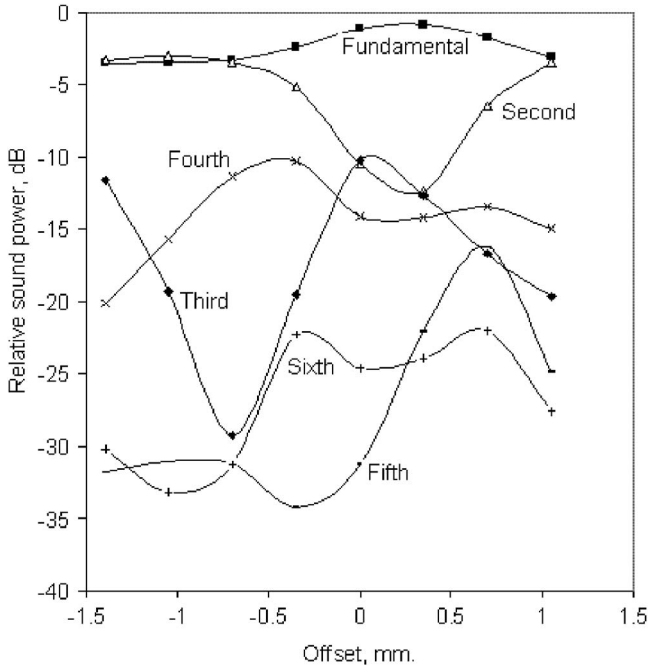


FIG. 5. Harmonic composition as a function of offset. Plenum pressure 225 Pa. Lip-to-edge distance 8.5 mm.

B. Lip-to-edge distance 4.5 mm

Here the ratio of the lip-to-edge distance to the jet thickness is 3.8. Sets of recordings were made for each of four blowing pressures. The sound power levels are shown in Fig. 6. While the levels are on an arbitrary base, the same base was used for Fig. 3, so these are directly comparable. Much the same power levels were obtained for the two distances, and again most of the change with offset occurs in the region of negative offsets. The breaks in the curves for plenum pressures 225 and 175 Pa indicate that between these the flute is overblowing, sounding F_5, as explained in more detail in the discussion of Fig. 8 to follow.

Figure 7 shows the harmonic composition for a plenum pressure of 75 Pa. This is appropriate to the small lip to-edge distance, the sounding frequency being very close to that of

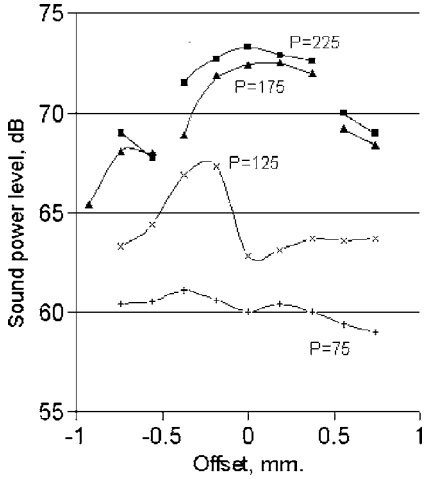


FIG. 6. Relative sound pressure as a function of offset and plenum pressure (Pascals) for lip-to edge distance of 4.5 mm. In the central portions of the upper two curves the flute is overblown, sounding an octave higher.

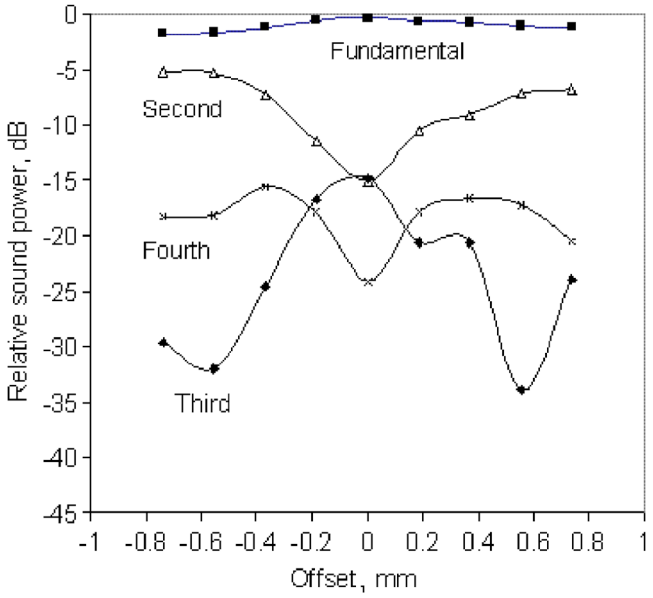


FIG. 7. Harmonic composition as a function of offset. Plenum pressure 75 Pa. Lip-to-edge distance 4.5 mm.

the passive resonance for this note. The chart shows a close resemblance to that obtained in the time domain study for a jet wave delay of 5/16 cycle. The chart is nearly symmetrical with respect to positive or negative offsets and the fundamental is dominant for all offsets. The even harmonics show large dips at zero offset. Though their amplitudes do not approach zero at zero offset, they do undergo a phase shift of 180° either side of zero, as expected from theory (Nolle[2], Fletcher and Douglas[3]) and exhibited in the above-noted time domain simulation.

At higher blowing pressures and some offsets, the flute overblows, that is, goes into the second mode and sounds the octave. This occurs at offsets zero and either side, the extent increasing with blowing pressure. Figure 8 displays the harmonic composition for a high plenum pressure, 225 Pa. For

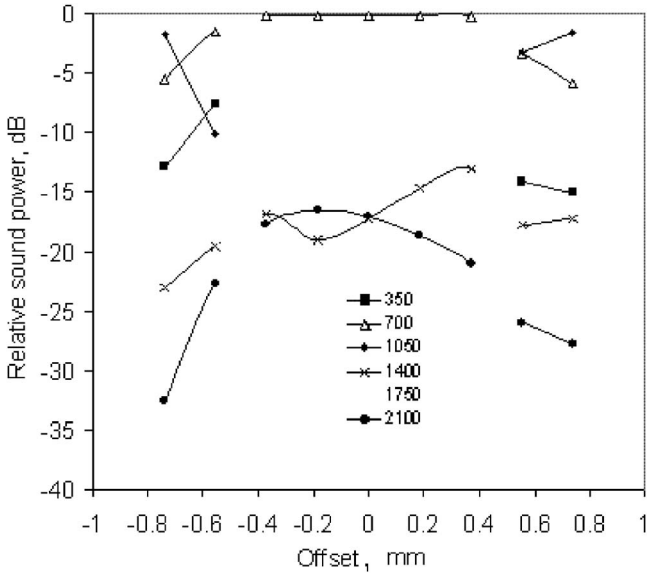


FIG. 8. Harmonic composition as a function of offset. Plenum pressure 225 Pa. Lip-to-edge distance 4.5 mm.

offsets between −0.4 and 0.4 mm the flute is sounding the octave. What was previously the second partial is now the fundamental. For this reason, the points are not labeled "fundamental," etc., but rather with their nominal frequencies of 350 Hz and its harmonics. Note that for large offsets, the flute still sounds F_4 at 350 Hz. However, the second and third harmonics of that note are here far stronger than the fundamental. It is evident that in this region we have a locked-in multiphonic, implying feedback at these frequencies as well as at the fundamental. The second harmonic when sounding F_4 approaches the strength of the fundamental of the F_5 around offset 0.5 or −0.5 mm, so that a smooth transition to the octave is easily made. Benade (Ref. 11, p. 71) describes the qualitative features of such a transition, and how the transition can be so subtle that listeners may disagree on whether or not a transition has occurred.

IV. WARBLE

A steady modulation of the sound from an organ pipe, at frequencies in the neighborhood of 20 Hz, occasionally takes place for certain adjustments during the voicing of organ pipes. Nolle[2] calls this mode-shift or buzz, but warble appears to be a more appropriate term. It does not appear to be used in playing the transverse flute, where flutter-tonguing is used to obtain a similar effect, but it does occur in some Native American Indian flutes. Some players on this instrument considerate it a desirable effect, an addition to their artistic repertoire. A description of the history of warble and its usage among Native Americans and what factors enter in the design of the flutes, together with recordings of the sound, may be accessed at Gatliff.[12]

During the offset experiments, warble was observed for some particular settings. The example we treat here was obtained with a lip-to-edge distance of 8.5 mm, plenum pressure 140 Pa, and an offset of 0.56 mm. A longer-than-usual recording containing 17 000 sample points was made to obtain higher resolution in frequency, and to include several cycles of the warble. Passive resonance frequencies and Q's of the first three modes of the flute were measured for help in analysis. These were obtained by exciting the flute at the mouth hole with a small loud speaker and observing the response with the tie-tack microphone inside the flute tube about 5 cm below the open F-hole. The resonance frequency and Q were obtained by fitting the microphone response versus frequency to a theoretical resonance curve.

The Fourier spectrum of the warble tone showed a clean unmodulated fundamental at 340.64 Hz. The second partial consisted of two well-resolved peaks, the first at 681.3 Hz, exactly twice the fundamental, while the second peak, of slightly greater amplitude, had a frequency of 713.8 Hz. The difference in these two frequencies, 32.5 Hz, closely approached the observed warble frequency of 33 Hz. By setting to zero all the Fourier components except those in the neighborhood of the two peaks, and performing an inverse Fourier transform, a wave form nearly 100% modulated at the difference frequency was obtained. It is evident that this warble was caused by the beating of the two frequencies of the second partial: one an exact harmonic produced by nonlinear current in the jet sweeping at the fundamental frequency, and the other a multiphonic which is not locked in, but is maintained by feedback from waves at the second-mode frequency carried also by the jet.

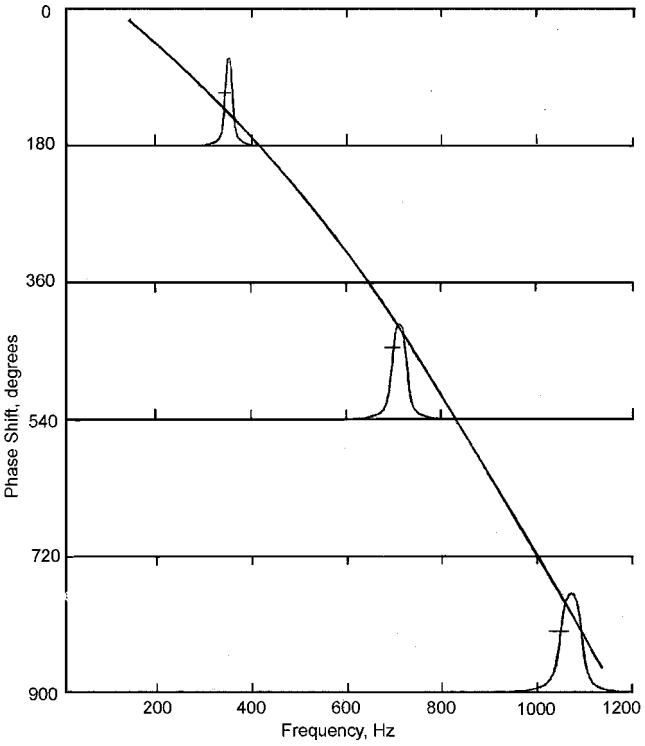


FIG. 9. Feedback phase conditions for three modes of the flute.

The third mode peak was very strong, slightly greater than the fundamental. Its frequency, 1054.4 Hz, was not harmonically related to that of the fundamental, and apparently it was sustained also by feedback from waves carried by the jet. A very small satellite peak, some 27 dB down in intensity, was seen at 1022.1 Hz, the third harmonic of the fundamental. Presumably it was also beating with its neighbor, but with an amplitude too small to be evident.

An analysis of this situation was pursued using the phase shift diagram from Coltman (Ref. 13, Fig. 4, p. 727). In any feedback process, the total phase shift around the loop must equal 360°, or integral multiples thereof. In the case of jet blown instruments, 180° of the phase shift is contributed by the reversal of currents at the mouth. For example, in a vertical edge-tone setup, current from the jet inserted into the space to the right of the edge causes an immediate current flowing from right to left across the lip, thereby initiating jet motion to the left. The delay on the jet need only be another 180° for first-stage oscillation. The diagram of Fig. 4 of Coltman (Ref. 13) was extended by adding second and third stage conditions as shown here in Fig. 9. Frequency is plotted on the x axis, phase shift (downward) on the y axis. The humps are plots of the phase delay between jet current and the mouth current it generates in the pipe. That phase delay, θ, is calculated using Eq. (1), reproduced from Eq. (2) of Coltman (Ref. 13) after correcting a typographical error in sign,

$$\tan\theta = -R/(\cot\beta x(R^2 + \tan^2\beta L) - \tan\beta L). \quad (1)$$

Here L is the apparent length of the resonating pipe, inferred from its passive resonance frequency, mode number n, and velocity of sound c in the pipe. $R = n\pi/2Q$, and $\beta = \omega/c$. For the three modes, the measured passive resonance frequencies and Q's were, respectively, 343.8, 38, 696.8, 45, and 1048, 49. The resultant L's were 502, 495, and 494 mm, respectively. The jet is assumed to insert its current between two points on either side of the edge spaced by an effective distance x, which we estimated here as 2 cm. This is the value used in Coltman[13] that gave good correspondence to measurements. A hump is plotted for each of the three modes, the second and third being on axes displaced from the first by 360° and 720°, respectively. The passive resonance frequency, which is marked with short horizontal line on the hump, lies at a lower value than that of the maximum phase shift.

The long curved line crossing all three humps represents the delay on the jet. The pipe can only sound at frequencies where the sum of the jet wave delay and the pipe current delay equals 180° or 180° plus 360° or 720°. This condition is met at the points where the jet wave delay curve intersects a hump curve, and the sounding frequency of that mode is the frequency where the intersection occurs. The jet wave delay curve was drawn using the experimentally determined curves of Fig. 4 of the reference as a model. It was drawn so that the straight portion passes through the third and second mode humps at the measured frequency of their sounding, and the line was modestly curved to go through the sounding frequency of the first mode. Because of the cavalier fashion in which this line was drawn to fit the data, and the uncertainties in calculating the mode humps, the diagram does not qualify as a theoretical analysis of the observed behavior. It does however, illustrate how it is possible that a first mode can sound below its passive resonance frequency, while the jet is also generating multiphonics sounding at frequencies slightly higher than the passive resonance frequencies of higher modes.

After drawing this diagram, I realized that a slight increase in blowing velocity would displace the line slightly upward and to the right, so as to miss the hump of the second mode altogether. Repairing to the laboratory, I found an increase in blowing velocity of 4% was enough to stop the warble altogether. Fourier analysis of the recorded new sound showed the second partial was now a single peak at exactly twice the frequency of the fundamental. The third partial, though weak, consisted of two peaks, one at the third harmonic, and the other higher by 27.4 Hz, showing that the jet delay line missed the second mode peak but still intersected the third mode. This predictive ability strengthens belief in the correctness of the above-presented picture.

As mentioned in Sec. II, warble was found also in the time domain simulation, at the rather low frequency of 3.6 beats/s. Making a record of the wave from that simulation and subjecting it to Fourier analysis revealed a fundamental of 349.7 Hz, and a second partial with two peaks. One at 699.3 Hz had twice the fundamental frequency, the other was lower at 695.8 Hz. Their difference, 3.5 Hz, corresponded well with the observed warble frequency. It is evident that the same process was at work—a beat between an exact second harmonic produced by jet nonlinearity and a second mode independently sustained by feedback involving second mode waves carried by the jet.

Analyses were performed on three recordings of warble produced by Native American flutes. One of these exhibited all of the above-discussed characteristics. The other two were distinctly different. Warble modulation was strong in the fundamental as well as in the second and third partials. The modulation frequencies were identical, but interestingly, not in phase. Thus at the time the fundamental was at a maximum, the second partial was near a minimum, a half cycle later the second partial was dominant and the fundamental recessive. The warble then consisted not only of amplitude modulation, but modulation of the tone quality.

This behavior strongly resembles that of the "wolf tone" occurring in string instruments, as discussed at some length in Benade (Ref. 11, pp. 567–572). The theory there calls for a second system, coupled to the primary vibrator, having a resonance frequency close to that of the tone being sounded. I was able to produce warble of this type by placing the open end of a quarter-wave resonant tube about 2 cm from the open end of a flute being blown by the artificial blower. It therefore seemed worthwhile to look for a second resonator in the Native American flute.

A Native American flute typically has the jet formed by a channel under a block of wood, often carved in the shape of an animal or a bird, mounted on the outside of the body. Air is brought into this channel from a rather long chamber of the same diameter as the main tube, provided with a small hole at the end into which the player blows. Such a chamber might be a candidate for the coupled resonating system. However, calculation of the resonance frequency of the chamber on one of the warbling flutes showed it to lie well below the sounding frequency. Filling the chamber with rice had no effect on the warble.

It appears that two distinct kinds of warble can occur in flutes of this type. I offer no explanation for the second kind.

V. SUMMARY

Jet offset, together with blowing pressure, plays a major role in determining the harmonic content and dynamics of the tone produced by the Boehm flute. Time domain simulation showed the expected rise of even harmonics with offset, to the extent that the radiated power of higher harmonics could exceed that of the fundamental by more than four times. Alterations in harmonic content can bring additional changes in the loudness perceived by a listener. With an artificial blower and a real flute, the effects were quite asymmetric, being most substantial for offsets in which the jet is directed into the embouchure hole. At the higher blowing pressures, the radiated power of higher harmonics often exceeds that of the fundamental. In particular, with a large offset and strong blowing, the strengths of the even harmonics can come to match the harmonics of the overblown second register note. Then with a slight change in offset a transition to the octave note is so smooth as to make the listener unsure

just when it occurred. Altogether, jet offset combined with blowing pressure seems to be the dominant variable under the control of the player in shaping tone quality and dynamics.

Under certain circumstances warble occurred both in the experiment and in the simulation. The phenomenon is explained as a beat between the frequency of a second harmonic generated by nonlinearity in the jet current and a neighboring partial sustained by jet feedback near the second mode resonance. A second type of warble, in which amplitude modulation occurs in all partials but with different phases, is yet to be explained.

ACKNOWLEDGMENT

The cooperation of Robert Gatliff in supplying information and samples of warble in Native American flutes is gratefully acknowledged.

[1]A. W. Nolle, "Some voicing adjustments of flue organ pipes," J. Acoust. Soc. Am. **66**, 1612–1626 (1979).
[2]A. W. Nolle, "Flue organ pipes: Adjustments affecting steady waveform," J. Acoust. Soc. Am. **73**, 1821–1832 (1983).
[3]N. H. Fletcher and L. M. Douglas, "Harmonic generation in pipes, recorders, and flutes," J. Acoust. Soc. Am. **68**, 767–771 (1980).
[4]J. W. Coltman, "Time domain simulation of the flute," J. Acoust. Soc. Am. **92**, 69–73 (1992).
[5]A. W. Nolle, "Sinuous instability of a planar air jet: Propagation parameters and acoustic excitation," J. Acoust. Soc. Am. **103**, 3690–3705 (1998).
[6]J. W. Coltman, "Jet behavior in the flute," J. Acoust. Soc. Am. **92**, 74–83 (1992).
[7]M. Verge, B. Fabre, A. Hirschberg, and P. J. Winands, "Sound production in recorder-like instruments. I. Dimensionless amplitude of the internal acoustic field," J. Acoust. Soc. Am. **101**, 2914–2925 (1997).
[8]R. Kaykayoglu and D. Rockwell, "Unstable jet-edge interaction. 2. Multiple frequency pressure fields," J. Fluid Mech. Digital Archive **169**, 151–172 (1986).
[9]N. H. Fletcher, "Mode locking in nonlinearly excited inharmonic musical oscillators." J. Acoust. Soc. Am. **64**, 1566–1569 (1978).
[10]J. W. Coltman, "Mode stretching and harmonic generation in the flute," J. Acoust. Soc. Am. **88**, 2070–2073 (1990).
[11]A. H. Benade, *Fundamentals of Musical Acoustics* (Oxford University Press, New York, 1976).
[12]R. Gatliff, "The warble of Native American flutes," Flutetree.com www.flutetree.com/nature/Warble.html. Last accessed July 10, 2006.
[13]J. W. Coltman, "Jet drive mechanisms in edge tones and organ pipes," J. Acoust. Soc. Am. **60**, 725–733 (1976).

Transmission loss in manatee habitats

Jennifer L. Miksis-Olds[a)]
Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882

James H. Miller[b)]
Department of Ocean Engineering and Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882

(Received 22 February 2006; revised 29 June 2006; accepted 6 July 2006)

The Florida manatee is regularly exposed to high volumes of vessel traffic and other human-related noise because of its coastal distribution. Quantifying specific aspects of the manatee's acoustic environment will allow for a better understanding of how these animals respond to both natural and human-induced changes in their environment. Transmission loss measurements were made in 24 sampling sites that were chosen based on the frequency of manatee presence. The Monterey-Miami Parabolic Equation model was used to relate environmental parameters to transmission loss in two extremely shallow water environments: seagrass beds and dredged habitats. Model accuracy was verified by field tests at all modeled sites. Results indicated that high-use grassbeds have higher levels of transmission loss for frequencies above 2 kHz compared to low-use sites of equal food species composition and density. This also happens to be the range of most efficient sound propagation inside the grassbed habitat and includes the dominant frequencies of manatee vocalizations. The acoustic environment may play a more important role in manatee grassbed selection than seagrass coverage or species composition, as linear regression analysis showed no significant correlation between usage and either total grass coverage, individual species coverage, or aerial pattern. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2258832]

PACS number(s): 43.80.Ev, 43.80.Nd [WWA] Pages: 2320–2327

I. INTRODUCTION

In order to better understand how sound may affect manatees in critical habitats, it is necessary to quantify the acoustic propagation loss characteristics of these extremely shallow-water regions. Transmission loss is particularly important to characterize because the sonar equation incorporates this term when relating received levels to both source levels and issues of detection (Urick, 1983). Geographically speaking, shallow water refers to the inland waters of bays and harbors and to coastal waters less than 200 m deep (Etter, 1996). The depth of manatee habitats covers only the shallowest 5% of that range.

Compared to sound propagation in deep water, the propagation of sound in shallow water is complicated. The difficulty in characterizing transmission loss in shallow water regions is due to the complex variability of environmental conditions in space and time, as well as the interactions between the upper and lower boundary layers. The range of detection in shallow waters is severely limited by high attenuation resulting from repeated interactions with the bottom and by limited water depths, which do not affect the long-range propagation paths in deep water (Etter, 1996). The challenges associated with characterizing sound propagation and signal detection in shallow water has resulted in numerous theories and mathematical models aimed at integrating acoustic and boundary conditions with transmission loss (Officer, 1958; Brekhovskikh, 1960; Urick, 1983; Frisk, 1994; Etter, 1996).

Unfortunately simple transmission loss models are not accurate for the complex intra-coastal environments that manatees inhabit. Nowacek *et al.* (2001) found that frequencies of sound produced by boats are attenuated in manatee habitats to a greater degree than would be predicted by simple transmission loss models. More detailed mathematical models are needed. Recently, modified parabolic equation (PE) models have been used successfully in shallow-water environments (Jensen, 1984; Etter, 1996; Smith, 2001). These models are based on a solution of the parabolic versus elliptic-reduced wave equation, which is used with ray theory and normal mode models. The PE models are most appropriate for use in range-dependent environments and can be used over a broad frequency band (Jensen, 1982; Collins and Chin-Bing, 1990; Orchard *et al.*, 1992; Etter, 1996; Smith, 2001).

The Monterey-Miami Parabolic Equation (MMPE) model was the specific PE model used in this study. The MMPE model produces solutions which are just as accurate as a benchmark quality model given a real ocean environment with inherent uncertainties. The efficiency of this model in producing both continuous wave and broadband pulse predictions makes it an attractive and powerful tool for ocean acoustic propagation modeling (Smith, 2001). The MMPE model is a numerical, far-field approximation of the horizontal acoustic propagation from a source in which the pressure field is represented by an outgoing Hankel function slowly modulated by an envelope function (Smith and Tap-

[a)]Current affiliation: School for Marine and Science Technology, University of Massachusetts Dartmouth, 706 S. Rodney French Blvd., New Bedford, MA 02744; electronic mail: jmiksis@umassd.edu
[b)]Electronic mail: miller@oce.uri.edu

pert, 1993). The current version of MMPE is a two-dimensional PE model that employs a split-step Fourier algorithm and assumes the surface is a perfect reflector due to a pressure release boundary. Input parameters needed to run the model are: sound speed profile, range-dependent bathymetry contour, sediment properties (sound speed, sound speed gradient, density, compressional attenuation, shear speed, and shear attenuation), source depth, and source type. The MMPE model also allows for an additional bottom layer to be present on top of the deep basement layer to allow for the effect of sediment or grass layers (Smith, 2001).

Understanding how sound is propagated in different manatee habitats is critical in order to more clearly understand the impact of human activities on manatees and the manatee communication system. For example, watercraft collisions have become the leading cause of adult mortality (Ackerman *et al.*, 1995; Reynolds, 1995). The question that naturally arises from this is whether or not manatees are hearing the noise produced from approaching boats in enough time to swim out of harm's way. The root of this question is the detection of sound signals. Quantifying sound propagation in different habitats provides one element of information necessary for determining received levels and ultimately the probability of manatees detecting approaching sound sources. Establishing whether manatees are discriminating between habitats that provide either more or less efficient sound propagation is a second logical element that needs to be addressed. The purpose of the study was to investigate manatee habitat use in relation to transmission loss.

II. METHODS

A. Site selection and habitat use

Sound propagation loss was investigated in two manatee habitats: seagrass beds and dredged habitats. These habitat types were chosen because of their biological importance to manatees. Animals typically feed in grassbeds and rest or socialize in dredged habitats (Koelsch, 1997). Habitats used by manatees in the Sarasota Bay, FL area were identified from aerial survey data available from Mote Marine Laboratory for the years 2000–2003 (Gannon *et al.* (in press); Lefebvre *et al.*, 1995). A total of 24 sites were selected for acoustic and environmental sampling: 13 grassbeds and 11 dredged habitats. Grassbed sites were defined by the presence of seagrass within the site, and dredged habitats were areas that had been dredged for human use and were characterized by the presence of a fine sediment layer. There were two selection criteria for site selection. First, manatees had to be observed in a site more than once over the 4-year survey period from 2000–2003. Second, the site had to be accessible by the 5.2 m (17 ft) research vessel.

The percentage of surveys in which animals were sighted in the selected grassbeds ranged from 5.3% to 78.9%. The percentage of surveys in which animals were observed in the dredged habitats ranged from 5.3% to 39.5% (Table I). The 13 grassbed sites included the five most heavily used grassbeds identified from the aerial surveys, one of which was in a manatee sanctuary (Pansy Bayou Grassbed or Pansy GB). The grassbed sites also included the five least used grassbeds in Sarasota Bay meeting the selection criteria. The 11 dredged habitat sites included the three most heavily used dredged basins/canals, one of which was in a manatee sanctuary (Pansy Dredged Basin or Pansy DB). The dredged sites also included the four least used dredged habitats in Sarasota Bay meeting the selection criteria.

TABLE I. Selected grassbed and dredged habitat sites with associated usage patterns.

Site	Grassbed	Usage (%)	Site	Dredged habitat	Usage (%)
A	City Island Grassflats (CIGF)	78.9	**C**	Pansy DC	39.5
B	Pansy Bayou GB	73.7	**W**	Buttonwood Canal	23.7
V	Buttonwood Harbor S	44.7	**U**	Bowlees Creek	15.8
H	S. Sarasota Bay	44.7	**K**	Cluster	13.2
I	W. Roberts Bay	44.7	**M**	E. Roberts Bay	13.2
T	Bowlees GB	18.4	**L**	Phillipi Creek	13.2
N	SE Sarasota Bay	18.4	**J**	Cocoanut Bayou	10.5
S	Airport GB	15.8	**R**	Whitacker Bayou	5.3
D	CIGF East	10.5	**Q**	Hyatt Basin	5.3
F	SW Bird Key	10.5	**E**	S. Lido Canal	5.3
X	New Pass GB	7.9	**P**	Harbor Acres	5.3
G	Down South Lido	7.9			
O	E. Sarasota Bay	7.9			

B. Transmission loss

The MMPE model was used to model the sound propagation loss at sites within the Sarasota Bay area (Smith, 2001). Transmission loss in a seven-octave frequency band was modeled from 250 Hz to 20 kHz over a range of 100 m. Transmission loss was quantified for eight frequencies: 250, 500 Hz, 1, 2, 4, 8, 16, and 20 kHz. The 20 kHz frequency was chosen as the maximum because this was the maximum frequency output of the broadcasting system used during the field test validations. Initial environmental parameters were collected during the summer of 2003 for application in the MMPE model. A SBE 25 SEALOGGER CTD was used to monitor salinity, temperature, and sound speed profiles in each site over the course of the season. Each environmental input parameter was averaged for a 6-month time period, and the average sound speed value was used in the MMPE model for each site. The largest difference between the average 6-month sound speed profile and any individual sample within each site was less than 1.5%; therefore the seasonal variation in the sound speed profiles had a negligible effect on model predictions. Bathymetry data were obtained by doing transects across each site. A bathymetry reading was recorded approximately every 10 m. Sediment properties were obtained from the Sarasota Bay National Estuary Program (Culter and Leverone, 1993). Distribution of sediment grain sizes in each site was identified from Culter and Leverone (1993). The proportion of grain sizes in each site was then used to estimate sediment sound speed, sound speed gradient, density, and compressional attenuation loss from Hamilton (1980).

The modeled transmission loss range in all sites was approximated from the distance between the closest boat

channel and the farthest possible manatee position within a given site. A point source at a depth of 0.75 m was used in all models in order to most closely simulate the depth of an outboard motor. A 50 m sediment layer was used in all dredged habitat model runs. The 50 m sediment layer width was chosen because it was the minimum layer width that produced no interaction with the rock layer deep below the sediment layer. In seagrass beds, the transmission loss was modeled with a 0.3 m grass layer on top of a 50 m sediment layer. All grassbed sites were modeled with the same grass layer acoustic properties. These properties were approximated for turtle grass (*Thalassia testundinum*), the dominant seagrass species in the 13 selected grassbed sites. Grass layer velocity (1450 m/s), density (0.90 g/cm^3), and attenuation loss (0.17 dB/km/Hz) were derived from grass blade density and physiological and biomechanical properties of turtle grass in Sarasota Bay, FL (Tomasko *et al.*, 1996; Sabol (private communication)). Density values were taken directly from measured values, whereas velocity and attenuation loss values were estimated based on the cross-sectional ratio of gas-filled lacunae and plant tissue. It was assumed from previous work on the acoustic reflectivity of aquatic vegetation that plant tissue had the acoustic properties of seawater, while the lacunae had acoustic properties of air (Sabol *et al.*, 1997; Kopp, 1998; Sabol *et al.*, 2002). The 0.90 g/cm^3 density value reflects the density of the grass blades themselves. *In situ* biomass density is variable and was not measured directly; consequently *in situ* biomass was not used as a model input. However, *in situ* biomass can be approximated if the area grass blade density is known. For example, an approximate 3 kg/m^2 grass density roughly corresponds to 1.1% of the volume of a 1 m^2 by 0.3 m deep volume of water. Using a seawater density of 1.0247 g/cm^3, the layer density would be approximately 1.0233 g/cm^3 (0.011 $\times$0.9 g/cm^3+0.989$\times$1.0247 g/cm^3).

The MMPE model outputs transmission loss in three forms: TL at a single frequency versus range and depth, TL at a single range versus frequency and depth, and TL at a single depth versus frequency and range. All results in this study were based on the output of TL at a single frequency versus range and depth. In order to quantitatively compare TL at a specified range and frequency between sites, TL was averaged over the depth of the water column at ranges of interest. The concept of averaging over the depth of the water column also has a biological basis, as the direct channels of manatee sound reception are not completely clear. Manatees have been shown to sense sound pressure levels in the region of the head, but it has also been suggested that the short hairs uniformly spaced over their body may detect particle displacement (Gerstein *et al.*, 1999). All dB units were converted to intensity before averaging and re-converted back to dB units for final calculations.

A difference technique was used to validate the MMPE model outputs. Difference techniques measure the distance between the model prediction and field measurements in terms of dB difference at a given range (Etter, 1996) . Model accuracy was verified at all sites by recording a broadband signal transmitted from an anchored boat a known distance away. The broadcast signal was a 1 s frequency sweep spanning 20 Hz–22 kHz. In 2003, the frequency sweep was introduced by a J-9 underwater transducer, which is capable of producing sounds in the range of 40 Hz–20 kHz with a source level near 160 dB re 1 μPa at 1 m.[1] In 2004, a Lubell LL916 underwater loudspeaker system was used as a source. This system had a 200 Hz–20 kHz frequency range with an output source level of 180 dB re 1 μPa at 1 m.[1] Neither transducer had a flat frequency response in the 20 Hz–22 kHz range, so a source level measurement for the frequency sweeps was recorded at a 1 m distance in each site to obtain a frequency-dependent source level for transmission loss calculations (Fig. 1). Figure 1 was processed to provide the source level in dB re 1 μPa rms at 1 m.

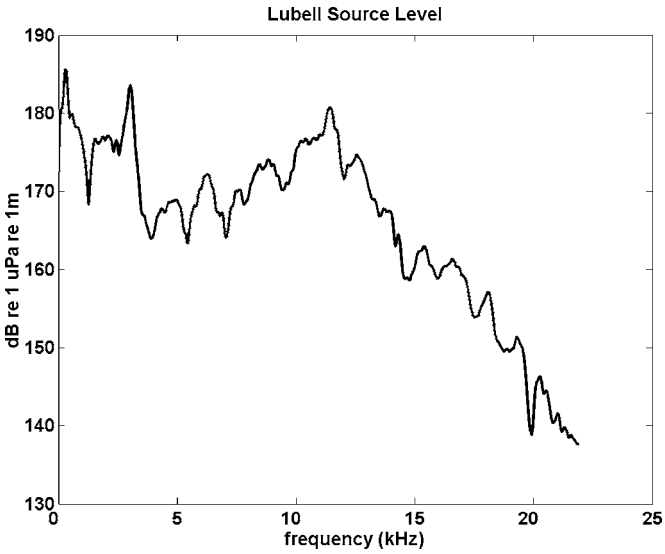


FIG. 1. Source level of 1 s frequency sweep transmitted from the Lubell LL916 transducer. Raw data were processed to provide the source level in dB re 1 μPa rms at 1 m.

Transmission loss was calculated by subtracting the received levels of signals recorded at a distance of 5, 10, 25, and 50 m from the 1 m source level at each of the eight modeled frequencies in all sites. Additional measurements at 100 m were made in two dredged habitat sites. All settings of the broadcasting system remained constant throughout the study. The dynamic range of the recording system was varied to prevent overloading the system. The recording hydrophone was a HTI-99-HF hydrophone with built-in preamplifier and had a 2 Hz–125 kHz frequency range and −178 dB re 1V/μPa sensitivity. The recording system was a National Instruments PCMCIA DAQ Card-6062E used in conjunction with a Dell Inspiron 8100. This system had a frequency response of 5 Hz–250 kHz with a selectable input voltage range. All recordings were sampled at a rate of 200 kHz.

Model accuracy was evaluated by examining field measurements with respect to model output as a function of range and frequency (Fig. 2). Figure 2 illustrates how the MMPE model outputs for each model run were viewed in relation to the field measurements. Each MMPE model output was examined on two levels: TL at the depth of the hydrophone and TL depth averaged over the water column as a function of distance. In grassbeds the hydrophone depth was 1 m, and the depth of the hydrophone in dredged habi-

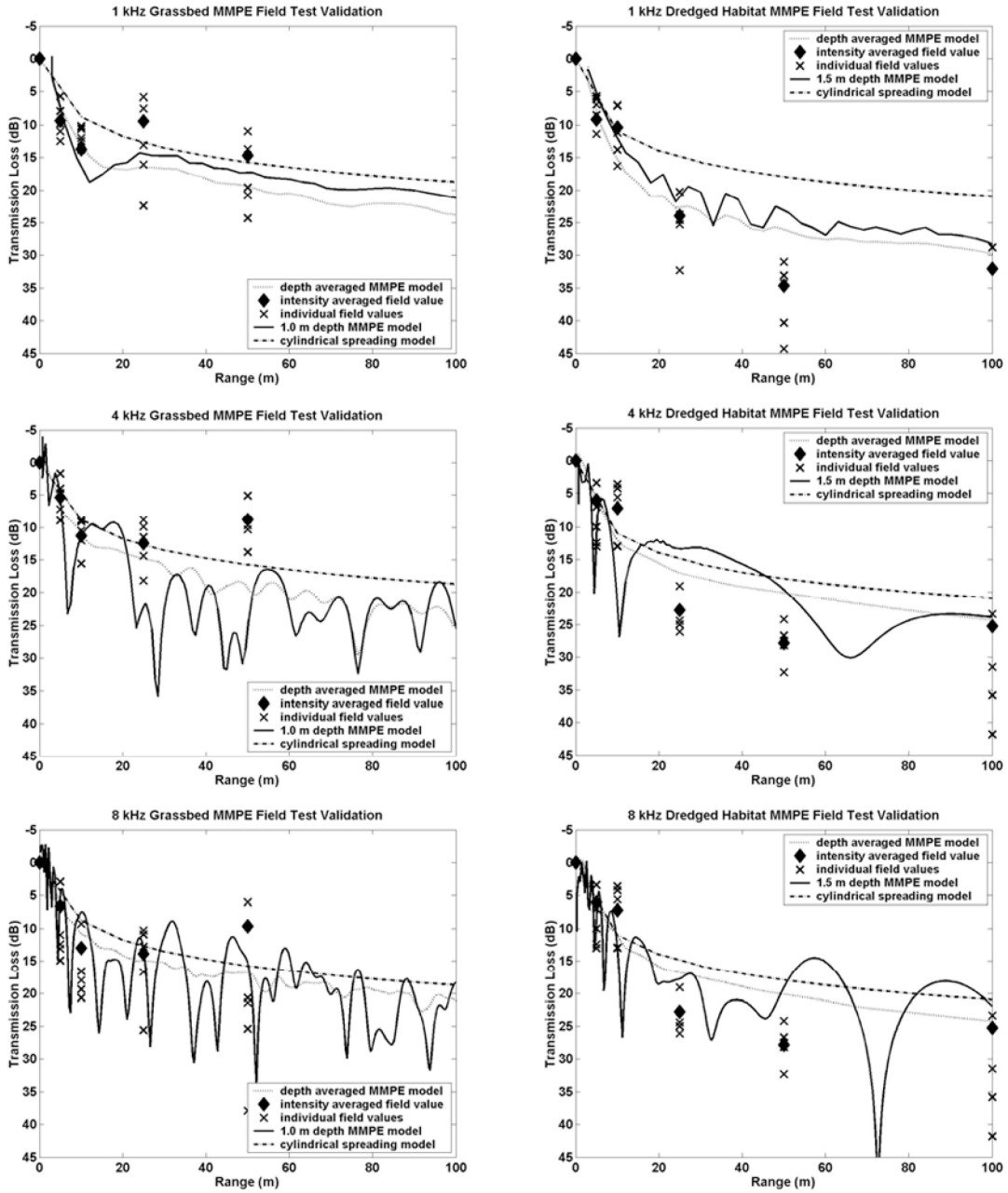


FIG. 2. Model predictions and field measurements for transmission loss at threefrequencies in a single grassbed and dredged habitat. The intensity averaged field value at a specified distance is represented by a diamond. "×" symbols represent the individual measurements from which the average was calculated. The solid line represents the MMPE model output at the depth of the hydrophone making the field measurements (1.0 m in grassbeds and 1.5 m in dredged habitats). The dotted line is the depth averaged TL estimated in the water column by the MMPE model over the 100 m range. The cylindrical spreading model is presented by a dash-dot line for comparison to the MMPE model results. Input parameters for the model run in each site are detailed in Table II. Note: transmission loss was so great in grassbeds at 100 m that no field calculation was possible.

tats was 1.5 m. A cylindrical spreading model (TL $=10\log(r)+10\log(h)$) was included for comparison. Deviations between the model and field measurements were calculated at the hydrophone depth [hydrophone depth=1 m in grassbeds and 1.5 m in dredged habitats] (Figs. 2 and 3).

C. Seagrass habitat density estimates

Seagrass density was estimated for the area's three most prominent seagrass species: turtle grass (*Thalassia testudinum*), manatee grass (*Syringodium filiforme*), and shoal grass (*Halodule wrightii*). In-water estimates were obtained using standard procedures for shoot density and biomass (Tomasko and Dawes, 1989). A 1 m^2 quadrant, which was divided into 25 equal 20×20 cm squares, was cast six times in each seagrass habitat. The total grass coverage, as well as individual grass species and macroalgae coverage, was evaluated for each quadrant toss. The values were then averaged to determine a final species and total grass coverage percentage for each site.

Seagrass patterns within each site were also estimated from an aerial survey flown on 16 June 2004. Grass patterns were evaluated on a 5-point scale ranging from sparse to

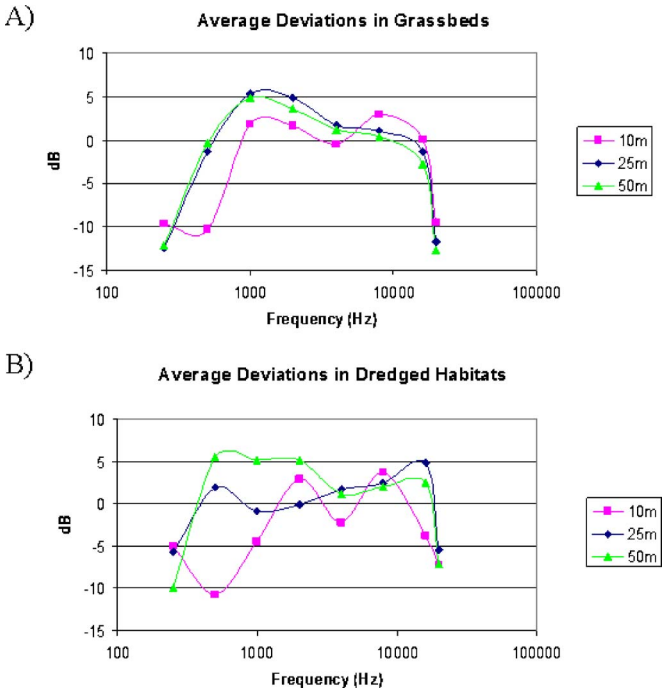


FIG. 3. Average deviations of field measurements from MMPE model predictions at the hydrophone depth for transmission loss in grassbeds (A) and dredged habitats (B). Hydrophone depth in the grassbeds was 1 m. Hydrophone depth in the dredged habitats was 1.5 m. Deviations are shown for measurements at three specified ranges. Negative values indicate the model overestimated the TL, and positive values indicate the model underestimated the TL.

dense grass coverage. The categories were: (1) sparse, (2) sparsely patchy, (3) densely patchy, (4) continuous (moderate cover), and (5) dense. All references to seagrass habitat quality in this work pertain only to the parameters of density and species composition. It was assumed that quality is dependent on the availability of the three most highly consumed seagrass species in Florida and does not take into account other previously used parameters to assess seagrass quality such as shoot age, weight of plant material, time spent chewing, etc. (Bengtson, 1983).

III. RESULTS

Initial MMPE model results comparing TL across habitat types supported documented evidence from previous manatee habitat TL experiments which reported higher levels of transmission loss in grassbeds compared to dredged habitats (Nowacek, 2001). Modeled and measured transmission loss was often greater in grassbeds than in adjacent dredged basins or canals at close ranges and always greater at the range of 100 m. The transmission loss in grassbeds at 100 m was so great that signals could not be detected for TL calculations (Fig. 2). This pattern was consistent for all frequencies modeled. Model results also indicated that the highest TL occurred at frequencies below 2 kHz, whereas the most efficient frequencies of sound propagation were 2–20 kHz in both grassbeds and dredged habitats.

When deviations between the field measurements and model calculations at the hydrophone depth were averaged over all the sites as a function of habitat type, range, and frequency, results indicated that the MMPE model was most accurate for frequencies from 1 to 16 kHz (Fig. 3). In this frequency range, average deviations were predominantly within ±5 dB. Deviations were averaged without taking the absolute values of the differences to identify biases. Negative deviation values indicated the model overestimated the TL, and positive values indicated the model underestimated the TL. An example of model parameters with their estimated errors for both a grassbed and dredged habitat site are presented in Table II. The origin of deviations between the model and field measurements is most likely due to errors in the environmental input parameters to the computational model.

TABLE II. Example of MMPE model input parameters with their estimated errors for both a grassbed and dredged habitat site. Depth errors are based on a 0.1 m oscillation of the research vessel due to waves, which adds a greater error component to shallow water depths. Note: depth is the only range-dependent variable in these very shallow, well-mixed environments.

	GB site			Dredged site		
	Value	Max error	% error	Value	Max error	% Error
Depth at range						
1 m	2.0	0.1	5	1.9	0.1	5
5 m	2.3	0.1	4	1.8	0.1	6
10 m	1.7	0.1	6	1.8	0.1	6
25 m	1.7	0.1	6	2.0	0.1	5
50 m	1.7	0.1	6	2.6	0.1	4
100 m	1.1	0.1	9	2.6	0.1	4
Sound speed (m/s)	1549.0	0.01	<1	1544.0	0.01	<1
Grass layer velocity (m/s)	1450.0	50	3			
Grass layer density (g/cm^3)	0.90	0.05	6			
Grass layer attenuation (dB/km/Hz)	0.04	0.005	12.5			
Sediment velocity (m/s)	1700.0	50	3	1590	50	3
Sediment density (g/cm^3)	1.8	0.05	6	1.59	0.05	6
Sediment attenuation (dB/km/Hz)	0.06	0.005	8	0.1	0.005	5

TABLE III. Regression analysis p values for transmission loss and usage comparisons at specified distances and frequencies. **Shaded** values show significant relationships at the 95% significance level.

Frequency	10 m	20 m	25 m	50 m	100 m	200 m
Grassbeds						
250 Hz	**0.019**	0.32	0.70	0.42	0.76	0.63
500 Hz	0.57	0.41	0.79	0.33	0.57	0.78
1 kHz	0.14	0.13	0.23	**0.04**	0.08	**0.04**
2 kHz	0.096	**0.035**	**0.04**	0.09	**0.004**	**0.001**
4 kHz	**0.030**	**0.032**	**0.04**	**0.01**	**0.045**	**<0.001**
8 kHz	**0.037**	**0.037**	**0.03**	**0.02**	**0.037**	**0.006**
16 kHz	**0.029**	**0.025**	**0.03**	**0.02**	**0.016**	**0.001**
20 kHz	**0.034**	**0.033**	**0.03**	**0.02**	**0.015**	**0.002**
Dredged habitats						
250 Hz	0.35	0.37	0.10	0.87	0.17	0.30
500 Hz	0.58	0.58	0.74	0.90	0.66	0.54
1 kHz	0.77	0.95	0.80	0.83	0.97	0.67
2 kHz	0.32	0.12	0.14	0.13	0.21	0.27
4 kHz	0.48	0.36	0.39	0.41	0.48	0.61
8 kHz	0.56	0.58	0.51	0.65	0.72	0.85
16 kHz	0.41	0.33	0.35	0.41	0.56	0.68
20 kHz	0.49	0.40	0.42	0.50	0.68	0.76

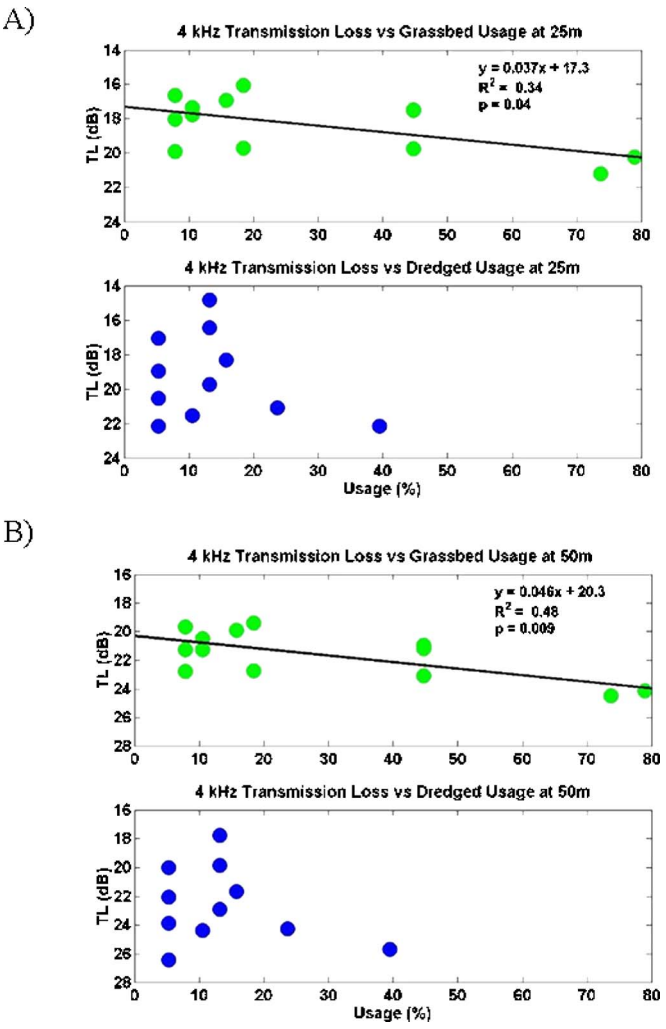


FIG. 4. 4 kHz transmission loss at 25 m (A) and 50 m (B) as a function of manatee site usage in grassbeds and dredged habitats. Solid regression lines indicate significant relationships in grassbeds.

Depth-averaged TL was calculated from the MMPE model outputs at distances of 10, 20, 25, 50, 100, and 200 m for each of the eight specified frequencies within each grassbed and dredged habitat site. Transmission loss in shallow water can be a complicated function of frequency due to the interference of the acoustic normal modes with varying phase velocities. For frequencies, separated by an octave, the transmission loss can be well modeled as statistically independent for waveguide depths greater than a wavelength (Bongiovanni *et al.*, 1996). Regression analyses were performed within each habitat type and at each frequency and distance in order to determine if TL was significantly correlated with manatee usage. Usage was defined as the percentage of time manatees were present at a site during aerial surveys from 2000 to 2003. Results showed a significant correlation between usage and TL in grassbeds at all investigated distances for frequencies from 4 to 20 kHz (Fig. 4 and Table III). Sites that were used more heavily by manatees tended to have higher levels of transmission loss. Significance was observed at some distances, but not all, for frequencies of 1–2 kHz. For all significant regressions, R squared values ranged from 0.32 to 0.71. There was no significant correlation between usage and TL in grassbeds at frequencies below 1 kHz or in dredged habitats at any frequency or distance.

Total seagrass coverage and individual seagrass species coverage varied widely among the 13 seagrass habitats sampled (Table IV). One hundred percent coverage was seen in four sites ranging in usage from 7.9% to 44.7%. The two most heavily used grassbed sites (A and B) had a total coverage of 91.3% and 75%, respectively. Linear regression analysis showed no significant correlation between usage and either total grass coverage, individual species coverage, or aerial pattern. This indicates that usage is not a function of seagrass habitat quality in relation to density parameters, but does not necessarily reflect patterns of usage in relation to other unmeasured parameters of seagrass quality (shoot age, plant weight, etc).

IV. DISCUSSION

Both model results and transmission loss field experiments showed that TL was greater in grassbeds than in adjacent dredged basins. In addition, the most TL occurred for frequencies below 2 kHz, whereas the least TL was seen for frequencies from 2 to 20 kHz. From a manatee's point of view this would mean that the sounds traveling through the environment least efficiently in both habitats are the lower frequency sounds, which overlap with dominant boat noise frequencies (Richardson *et al.*, 1995; Gerstein, 2002). Conversely, those frequencies that travel through the environment best are those that overlap the peak and fundamental frequencies (peak: 1–12 kHz; fundamental: 1–9 kHz) of manatee vocalizations (Nowacek *et al.*, 2003). A relatively quiet frequency band has been documented between 1 and 4 kHz in many terrestrial and ocean environments, and this may be one reason why bird and mammal vocalizations fall in these frequencies (Bradbury and Vehrencamp, 1998). It appears that the manatee communication system has adapted

TABLE IV. Grassbed quality estimates in relation to % usage. Total coverage of *Thalassia testudinum*, *Syringodium filiforme*, *Halodule wrightii,* and macroalgae values are in % of total quadrant covered by species. Aerial pattern values are based on a 1–5 point scale. The categories were: (1) sparse, (2) sparsely patchy, (3) densely patchy, (4) continuous (moderate cover), and (5) dense.

Site	% usage	Total coverage	Thalassia	Halodule	Syringodium	Macroalgae	Aerial pattern
A	78.9	91.3	75.3	29.3	8	45.3	5
B	73.7	75	46.6	92.2	0	4.7	4.5
D	10.5	100	52.3	34	0	29.3	4.5
F	10.5	100	96.7	38	33.3	2	3.5
G	7.9	100	98.7	16.7	0	76	5
H	44.7	100	83.3	16.7	33.3	9.3	4
I	44.7	84	50	17.3	0	41.3	3
N	18.4	69.3	47.3	46	16.7	12.7	1
O	5.3	98	53.3	28.7	32.7	27.3	3.5
S	15.8	98	96	28	2.7	0.7	4
T	18.4	83.3	78	33.3	0	47.3	4
V	44.7	96	86	34	0	50.7	5
X	7.9	85.7	68.6	30.3	0	24.6	3

to capitalize on the acoustics of the shallow water habitats they inhabit over evolutionary time. The presence of lower frequency boat noise in manatee habitats is a relatively new pressure on an evolutionary time scale, and its effects are yet to be fully understood.

Significant correlations of transmission loss and usage were found in grassbed habitats but not in dredged habitats. The exact cause of this observation is not known but could be related to how animals are using each habitat. Manatees typically feed in grassbeds and engage in play or rest while in dredged habitats (Koelsch, 2001). It is possible that manatees are selecting grassbeds that attenuate high levels of noise, allowing them to tolerate higher noise levels while meeting nutritional requirements. Grassbeds in Sarasota Bay, FL tended to be louder than dredged habitats due to the loud broadband noise produced by snapping shrimp (*Alpheus* and *Synalpheus sp*), which becomes stronger with decreasing depth (Richardson *et al.*, 1995; Camp *et al.*, 1998; Miksis-Olds, 2006). Manatees may be exposed to lower noise levels when resting or playing in dredged habitats and therefore become less selective in the acoustic properties of the habitat.

One question that naturally arises from the observed relationship is what factor is more dominant in driving the manatee grassbed usage, sound propagation or habitat quality? Dense grassbeds attenuate sound more than sparse grassbeds which may create a quiet area near a high noise traffic zone, yet a sparse grassbed may be located near a less busy boating area but propagate more noise. Dense grassbeds may also complicate the noise and sound propagation by the diversity of fauna living within the habitat (i.e., snapping shrimp, toadfish). Analysis of the seagrass coverage and species composition indicated no correlation between quality, as defined previously, and grassbed usage. This suggests that propagation characteristics associated with transmission loss play a more dominant role in habitat selection than the parameters of seagrass quality investigated in this study.

Making field measurements and using models to determine the transmission loss of a signal in manatee habitats are only two of many elements that must be quantified in order to ultimately answer questions pertaining to habitat selection and signal detection by an animal. Another major factor is noise. The actual range of effective signal transmission in the noisy, shallow-water areas manatees inhabit is dependent on the area noise levels, acoustic propagation loss characteristics, and frequency and amplitude of the signals being produced. Environmental parameters such as water depth, salinity, temperature, bottom type, and wind speed will also affect sound absorption and attenuation. Consequently, sound transmission is different for varying wavelengths in different manatee habitats, and different habitat types may make it easier or more difficult for manatees to detect either conspecific vocalizations or approaching vessels. Compared to the laborious technique of measuring transmission loss in the field, the MMPE model provides a relatively simple and accurate method for quantitatively assessing the transmission loss component involved in issues of signal detection and the impact of noise sources on manatees and other marine mammals living in shallow water environments, given accurate environmental parameters. A firm grasp on environmental noise levels, in addition to transmission loss characteristics, in specific habitats will build upon the work done here and is needed in order to more fully understand questions pertaining to manatee habitat selection and signal detection in these shallow habitats.

ACKNOWLEDGMENTS

The authors would like to acknowledge the following: Kevin Smith (Naval Postgraduate School) and Gopu Potty (URI Ocean Engineering) provided invaluable information regarding the MMPE modeling. The Manatee Research Program at Mote Marine Laboratory contributed aerial survey information and field interns. The authors would also like to acknowledge Bruce Sabol (Department of the Army, Engineer and Research Development Center) and Brad Robins (Mote Marine Laboratory) for their contribution to the estimates of seagrass acoustic properties and density measure-

ments. Special thanks are also extended to Peter Tyack (WHOI), John Reynolds (Mote Marine Laboratory), David Farmer (URI), Cheryl Wilga (URI), Bruce Sabol and two anonymous reviewers for comments on previous versions of this manuscript. This research was supported by a P.E.O. Scholar Award and National Defense Science and Engineering Fellowship awarded to Jennifer Miksis.

[1]Source level estimates at the lower end of the frequency band are not expected to be applicable, as the shallow water depths do not allow for free-field source level measurements.

Ackerman, B. B., Wright, S. D., Bonde, R. K., Odell, D. K., and Banowetz, D. J. (**1995**). "Trends and patterns in mortality of manatees in Florida". in Population Biology of the Florida Manatee (*Trichechus manatus latirostris*), National Biological Service, Information and Technical Report No. 1, pp. 1974–1992.

Bengtson (**1983**). "Estimating food consumption of free-ranging manatees in Florida," Journal of Wildlife Management **47**: 1186—1192.

Bongiovanni, K., Badiey, M., and Seigmann, W. L. (**1996**). "Interference patterns of frequency dependent transmission loss for shallow-water propagation," J. Acoust. Soc. Am. **99**, 2523–2529.

Bradbury, J. W., and Vehrencamp, S. L. (**1998**). *Principles of Animal Communication* (Sinauer Associates, Sunderland, MA).

Brekhovskikh, L. M. (**1960**). *Waves in a Layered Media* (Academic, New York).

Camp, D. K., Lyons, W. G., and Perkins, T. H. (**1998**). Checklists of Selected Marine Invertebrates of Florida, FMRI Technical Report No. TR-3, Florida Department of Environmental Protections.

Collins, M. D., and Chin-Bing, S. A. (**1990**). "A three-dimensional parabolic equation model that includes the effects of rough boundaries," J. Acoust. Soc. Am. **87**, 1104–1109.

Culter, J. K., and Leverone, J. R. (**1993**). "Bay bottom habitat assessment," Mote Technical Report No. 303, Mote Marine Laboratory, Sarasota, FL.

Etter, P. C. (**1996**). *Underwater Acoustic Modeling* (E & FN SPON, London).

Frisk, G. V. (**1994**). *Ocean and Seabed Acoustics* (PTR Prentice-Hall, Englewood Cliffs, NJ).

Gannon, J. G., Scolardi, K. M., Reynolds, J. E., III, (in press). "Habitat selection by manatees in Sarasota Bay, Florida," Marine Mammal Science.

Gerstein, E. R. (**2002**). "Manatees, bioacoustics and boats," Am. Sci. **90**, 154–163.

Hamilton, E. L. (**1980**). "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**, 1313–1340.

Jensen, F. B. (**1982**). "Numerical models of sound propagation in real oceans," Proceedings of the MTS/IEEE Oceans 82 Conference, pp. 147–154.

Jensen, F. B. (**1984**). "Numerical models in underwater acoustics", in *Hybrid Formulation of Wave Propagation and Scattering* (Martinus Nijhoff, Dordrecht), pp. 295–335.

Koelsch, J. (**1997**). "The seasonal occurrence and ecology of Florida manatees (*Trichechus manatus latirostris*) in coastal waters near Sarasota, FL," M.S. thesis, University of Southern Florida, Tampa, 121 pp.

Koelsch, J. (**2001**). "Reproduction in female manatees observed in Sarasota Bay, Florida," Marine Mammal Sci. **17**(2), 331–342.

Kopp, B. S. (**1998**). "The effect of nitrate fertilization and shading on physiological and mechanical properties of eelgrass (*Zostera marina*)," Ph.D. dissertation, University of Rhode Island, Narragansett.

Lefebvre, L. W., Ackerman, B. A., Portier, K. M., and Pollock, K. H. (**1995**). "Aerial survey as a technique for estimating manatee population size and trend - problems and prospects," in Population Biology of the Florida Manatee (*Trichechus manatus latirostris*). National Biological Service, Information and Technical Report No. 1, pp. 63–74.

Miksis-Olds, J. L. (**2006**). "Manatee response to environmental noise," Ph.D. dissertation, University of Rhode Island, Narragansett., 248 pp.

Nowacek, D. P., Buckstaff, K. C., Johnson, M. P., and Wells, R. S. (**2001**). Transmission loss of vessel noise in manatee environments, Report No. 721, Mote Marine Laboratory, Sarasota, FL.

Nowacek, D. P., Casper, B. M., Wells, R. S., Nowacek, S. M., and Mann, D. A. (**2003**). "Intraspecific and geographic variation of West Indian manatee (*Trichechus manatus spp.*) vocalizations (L)," J. Acoust. Soc. Am. **114**(1), 66–69.

Officer, C. B. (**1958**). *Introduction to the Theory of Sound Transmission* (McGraw-Hill, New York).

Orchard, B. J., Siegmann, W. L., and Jacobson, M. J. (**1992**). "Three-dimensional time-domain paraxial approximations for ocean acoustic wave propagation," J. Acoust. Soc. Am. **91**, 788–801.

Reynolds, J. E., III (**1995**). "Florida manatee population biology: Research progress, infrastructure, and applications for conservation and management," in Population Biology of the Florida Manatee (*Trichechus manatus latirostris*), National Biological Service, Information and Technical Report No. 1. pp. 6–12.

Richardson, W., Greene, C., Malme, C., and Thomson, D. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).

Sabol, B., McCarthy, E., and Rocha, K. (**1997**). "Hydroacoustic basis for detection and characterization of eelgrass (*Zostera marina*)," Proceedings of the Fourth International Conference Remote Sensing for Marine and Coastal Environments, 17–19 March, Vol. 1, pp. 679–693.

Sabol, B., Melton, R. E., Chamberlain, R., Doering, P., and Haunert, K. (**2002**). "Evaluation of a digital echo sounder system for detection of submersed aquatic vegetation," Estuaries **25**(1), 133–141.

Sabol, B. (private communication).

Smith, K. B. (**2001**). "Convergence, stability, and variability of shallow water acoustic predictions using a split-step Fourier parabolic equation model," J. Comput. Acoust. **9**, 243–285.

Smith, K. B., and Tappert, F. D. (**1993**). "UMPE: The University of Miami Parabolic Equation Model, Version 1.3," Marine Physical Laboratory, SIO, Technical Memorandum No. 432, February 1993.

Tomasko, D. A., and Dawes, C. J. (**1989**). "Evidence for physiological integration between shaded and unshaded short shoots of *Thalassia testudinum*," Mar. Ecol.: Prog. Ser. **54**, 299–305.

Tomasko, D. A., Dawes, C. J., and Hall, M. O. (**1996**). "The effects of anthropogenic nutrient enrichment on turtle grass (Thalassia testudinum) in Sarasota Bay," Estuaries **19**, 448–456.

Urick, R. J. (**1983**). *Principles of Underwater Sound* (Peninsula, Los Altos, CA).

Simulating the effect of high-intensity sound on cetaceans: Modeling approach and a case study for Cuvier's beaked whale (*Ziphius cavirostris*)

P. Krysl[a)]
University of California, San Diego, 9500 Gilman Drive No. 0085, La Jolla, California 92093-0085

T. W. Cranford
San Diego State University, 5500 Campanile Drive, San Diego, California 92182

S. M. Wiggins and J. A. Hildebrand
Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, No. 0205, La Jolla, California 92093-0205

(Received 23 December 2005; revised 16 June 2006; accepted 29 June 2006)

A finite element model is formulated to study the steady-state vibration response of the anatomy of a whale (Cetacea) submerged in seawater. The anatomy was reconstructed from a combination of two-dimensional (2D) computed tomography (CT) scan images, identification of Hounsfield units with tissue types, and mapping of mechanical properties. A partial differential equation model describes the motion of the tissues within a Lagrangean framework. The computational model was applied to the study of the response of the tissues within the head of a neonate Cuvier's beaked whale *Ziphius cavirostris*. The characteristics of the sound stimulus was a continuous wave excitation at 3500 Hz and 180 dB re: 1 μPa received level, incident as a plane wave. We model the beaked whale tissues embedded within a volume of seawater. To account for the finite dimensions of the computational volume, we increased the damping for viscous shear stresses within the water volume, in an attempt to reduce the contribution of waves reflected from the boundaries of the computational box. The mechanical response of the tissues was simulated including: strain amplitude; dissipated power; and pressure. The tissues are not likely to suffer direct mechanical or thermal damage, within the range of parameters tested. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2257988]

PACS number(s): 43.80.Gx, 43.80.Nd, 43.80.Lb, 43.80.Jz [WWA] Pages: 2328–2339

I. INTRODUCTION

Recently, attention has been directed toward understanding the relationship between cetacean stranding and high-intensity sound, particularly the potential for midfrequency naval sonar to cause stranding of beaked whales (Frantzis, 1998; NOAA, 2001; Balcomb and Claridge, 2003; Hildebrand, 2005). The overall pattern of these strandings has raised concerns that sounds from sonar could directly or indirectly result in the death or injury of beaked whales, particularly Cuvier's beaked whale (*Ziphius cavirostris*). The connection between sound exposure and cetacean stranding is not yet understood in terms of physiological or behavioral responses to sound (Cox *et al.*, in press; Rommel *et al.*, in press). To address some of these issues we present an investigation of the direct impacts of sound on beaked whale head tissues, presumed to be the most susceptible to anthropogenic sound impact.

We aim to develop techniques to model how high-intensity underwater sound interacts with cetacean anatomy and physiology. In this study, we first evaluate how well modern finite element modeling tools are capable of simulating the interaction between anatomic geometry, tissue properties, and sound sources with various characteristics. Then we model the interactions between sound waves and anatomy within the head of a specimen of Cuvier's beaked whale (*Ziphius cavirostris*). We chose this species because it has stranded in the greatest numbers during incidents associated with exposure to midfrequency sonar (Cox *et al.*, in press). The specimen of *Ziphius cavirostris* modeled in this study was the subject of computed tomography (CT) scans and tissue property measurements reported elsewhere (Soldevilla *et al.*, 2005). Consequently, more is known about the anatomy and physical properties of this specimen than any other single beaked whale.

In conjunction with our primary objective of developing modeling tools for the study of cetaceans, we investigated whether sounds with characteristics similar to those used in midfrequency naval sonar are capable of producing direct effects on cetacean tissues. We seek to understand whether vibrations of high-displacement amplitude or high-strain amplitude could lead to large amounts of energy being dissipated in the soft tissues causing thermal damage. Our model is formulated as follows. Consider an animal submerged in seawater; the animal's head is exposed to sound excitation which arrives in the form of continuous planar sound waves of a given frequency, direction, and sound pressure level.

[a)]Electroni mail: pkrysl@ucsd.edu

 0001-4966/2006/120(4)/2328/12/$22.50

We simulated the steady state vibration response of the tissues and assessed the quantitative mechanical characteristics of such motion.

Modeling sound propagation using numerical methods has been used previously to investigate sound generation and propagation pathways in a dolphin's head. Aroyan *et al.* (1992) used a two-dimensional numerical model to test ideas about dolphin sonar signal generation and emission (Cranford, 1988). Aroyan (2001) evaluated the competing and prevailing hypotheses for sonar signal transmission and reception (Norris, 1964; Brill and Harder, 1991; Cranford *et al.*, 1996) by creating a simulated three-dimensional acoustic propagation model. Primarily, these studies demonstrated the promise of using numerical methods to query anatomic systems about acoustic questions. One drawback in the studies of Aroyan and colleagues is that they used a simplified estimate of tissue properties. They assumed that tissues behave like fluids and did not consider the complex interactions between the solid properties of tissues and organs. The combination of techniques developed for the current study provides a more realistic approximation of tissue properties than that used by previous authors.

Finite element modeling (FEM) has become an established tool for an ever-widening sphere of applications. For instance, it has long been used by engineers to simulate the effects of vibrations on structures like buildings, bridges, and machines. Nowadays it is also being increasingly used to study the effects of stress/strain regimes on biological structures. The formulation of a biomechanical model involves a number of factors, including: the geometrical complexity of the anatomy; the heterogeneity, anisotropy, and rate-dependent response of the tissues; and computational costs. Also, the infinite extent modeling of the surrounding media (water) must be considered, since the sound waves that impinge upon the animal are reflected and scattered and propagate away in all directions, as well as being transmitted into the tissues.

This paper opens with a brief recapitulation of the partial differential equations that describe a model for deformations of tissues due to acoustic vibrations. It proceeds with a discussion of a constitutive equation, and formulates a strategy for the treatment of boundary conditions and a description of finite element discretization. The paper concludes with a case study, in which the modeling approach is used to assess the effects of selected high-intensity sound on the head anatomy of a neonate Cuvier's beaked whale *Ziphius cavirostris*.

II. MODEL

We present a brief description of the adopted partial differential equation (PDE) model that is used to describe the motion of the soft and hard tissues (and the surrounding water), both during the steady state forced vibration produced by progressive sound waves, and during any transient vibrations.

A. Governing equations

We consider the model of small-strain and small-displacement deformation of continuous media. Even for sound pressure levels ~200 dB, we would not expect the amplitude of the displacements to be larger than a few dozen micrometers and our results for the case study confirm this. A strictly Lagrangean description is adopted, for both solids and fluids (i.e., the particles of the continuum are tracked in time, as opposed to the Eulerian approach in which the continuum is watched as it flows through a control volume fixed in space). The continuum occupies volume V, whose bounding surface is S, and the time-dependent deformation is followed in the interval $0 \le t \le t_{\text{final}}$. The vector of engineering strains $\varepsilon = \{\varepsilon_{xx}, \varepsilon_{yy}, \varepsilon_{zz}, \gamma_{xy}, \gamma_{xz}, \gamma_{yz}\}^T$ (the superscript T indicates transpose) is defined as

$$\varepsilon = Bu,$$

where $u = \{u_x, u_y, u_z\}^T$ is the displacement vector, and

$$B = \begin{bmatrix} \partial/\partial x & & \\ & \partial/\partial y & \\ & & \partial/\partial z \\ \partial/\partial y & \partial/\partial x & \\ \partial/\partial z & & \partial/\partial x \\ & \partial/\partial z & \partial/\partial y \end{bmatrix},$$

is the symmetric gradient matrix operator. The balance equation is

$$\rho \ddot{u} = B^T \sigma,$$

where ρ is the mass density, $\sigma = \{\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{xy}, \sigma_{xz}, \sigma_{yz}\}^T$ is the stress vector, and $\ddot{u}$ is the second order time derivative of the displacement vector. Note that body forces are omitted in the balance equation. The constitutive equation is in general a relationship between stress and the independent variables strain and strain rate

$$\sigma = \sigma(\varepsilon, \dot{\varepsilon}).$$

The boundary conditions on $y \in S$ are written for the traction vector

$$(P_n \sigma)_i = \bar{t}_i,$$

where

$$P_n = \begin{bmatrix} n_x & & & n_y & n_z & \\ & n_y & & n_x & & n_z \\ & & n_z & & n_x & n_y \end{bmatrix}$$

is the projector on to the direction of the outer normal n, and $\bar{t}_i$ is the prescribed value of the traction component. For the velocity vector boundary conditions, where $\bar{v}_i$ is the prescribed velocity component

$$\dot{u}_i = \bar{v}_i.$$

Likewise, the initial conditions are

$$\dot{u}(x, t=0) = \bar{v}(x),$$

and

$$\sigma(x, t=0) = \bar{\sigma}(x).$$

B. Constitutive equation for soft tissue

To complete our model for time-dependent deformation of soft tissue, we need a constitutive equation with estimates for the associated material constants. Several good models of dissipative behavior of soft tissues exist, for instance, Rubin and Bodner (2002). However, for a given class of tissue, many material parameters are required, few of which have been thoroughly measured. Soldevilla *et al.* (2005) recently characterized the soft tissues of a neonate beaked whale (the same specimen used in our study), providing a map from CT scan data of x-ray attenuation Hounsfield units to mass density, dilatational sound speed, and the Young's modulus. Based on Soldevilla *et al.* (2005) results, the tissue properties are assumed to be isotropic, making, it is possible to extract from these experiments a full set of material parameters for the classical PDE of inhomogeneous isotropic elastodynamics. Therefore, as a first step we propose to use the constitutive model of isotropic viscoelastic response

$$\sigma = Kmm^T\varepsilon + 2GI_0I_d\varepsilon + 2\eta I_0I_d\dot{\varepsilon},$$

where K is the bulk modulus, G is the shear modulus, η is the dynamic viscosity, $m=[1,1,1,0,0,0]^T$, $I_0 = \frac{1}{2}\,\mathrm{diag}([2,2,2,1,1,1])$, $I_d=I-\frac{1}{3}mm^T$ is the deviatoric projector, I is the identity, and $\dot{\varepsilon}$ is the strain rate vector. Note that bulk viscosity has been omitted because it is generally considered less important at lower frequencies than shear viscosity (it is two or more orders of magnitude smaller; in addition, the dilatational strain rate is typically much lower in magnitude than shear strain rate). The use of this constitutive equation (under the label "Voigt's model") recently has been justified in transient elastography of soft tissues by Catheline *et al.* (2004).

C. Solution strategy for a bounded volume

The intent is to obtain a solution to the above equations that represents a steady-state forced vibration produced as a response to the continuous sound excitation. The solutions to the above PDE model are both progressive and standing waves. It is reasonable to assume that the surrounding seawater is infinite in extent (i.e., the free surface is ignored). The computational domain could also be infinite, and there are several possible approaches to address this problem [refer for instance to the review by Astley (2000)]: boundary element methods; methods that combine finite and infinite elements; or methods that combine finite elements with boundary elements (Wagner, 2004). Alternatively, the domain may be truncated, with the infinite part replaced by appropriate boundary conditions applied to the finite part; for instance the nonreflecting boundary conditions of Bayliss *et al.* (1982), or the perfectly matched layer formulation of Festa and Vilotte (2005). As a matter of expediency, this study made use of the finite-domain discretization, while we are evaluating all of the mentioned approaches as a means of controlling the approximation error.

During steady-state vibration, the plane sound waves that impinge upon the tissues are converted into a mixture of dilatational pressure and shear waves. These waves are reflected from, and transmitted across any impedance-mismatch interface, in particular the interface between soft tissue and bone, and between soft tissue or bone and an airspace. Any waves that reflect off the anatomy and arrive at the boundary of the computational volume will be reflected back into the volume, unless we control for them, for instance, by using the perfectly matched layer (PML) approximation (Festa and Vilotte, 2005) which introduces an anisotropic lossy material along the boundary to absorb oncoming waves. Since PML approximations tend to be computationally expensive and complex, and since we are not interested in the propagation of the reflected waves through the surrounding water, we make the entire water volume into an absorbing layer by suitably increasing its (Newtonian) viscosity to damp shear deformations. It should be noted that the sound waves of the acoustic stimulus will also be damped since the rate of shearing is nonzero in the forcing planar wave. The deviatoric components of the reflected waves that have left the tissue sample and are propagating toward the boundary, or of the waves that have been reflected from the boundary, will be strongly damped. Clearly, this is an approximation in that the reflected dilatational waves are not being damped until they get converted fully or partially into motions with nonzero shear (rate) components.

With the described approximate treatment of the effect of the finite extent of the computational box, it is possible to formulate a solution strategy: given any consistent set of boundary and initial conditions, integrate the governing equations in time until a steady state is reached. The quantities of interest are then directly available in the solution data. This approach resembles the dynamic relaxation solution techniques applied to static problems (e.g. Řeřicha, 1986).

D. Finite element discretization

Standard Galerkin discretization of the weak form of the PDE model leads to a system of ordinary differential equations

$$M\ddot{U} = F,$$

where M is the time-independent mass matrix; $\ddot{U}$ is the vector of all free components of the displacement; and F is the vector of corresponding effective forces (external and internal). The discretization in time follows the classical Newmark (centered differences) template, from which we obtain the time stepping scheme

$$A^{(n)} = M^{-1}F^{(n)}$$

$$U^{(n+1)} = U^{(n)} + \Delta t V^{(n)} + \frac{\Delta t^2}{2}A^{(n)},$$

$$V^{(n+1)} = V^{(n)} + \frac{\Delta t}{2}(A^{(n-1)} + A^{(n)})$$

where $A^{(n)}, V^{(n)}$ are approximations to the second and first order derivative (acceleration and velocity) of the vector of unknown displacements $U^{(n)}$ and Δt is the time step. For the prescribed velocity degrees of freedom, $V_j^{(n)}=\bar{V}_j^{(n)}$, the first equation is replaced with $A_j^{(n)}=\frac{2}{\Delta t}(\bar{V}_j^{(n)}-V_j^{(n-1)})-A_j^{(n-1)}$. The

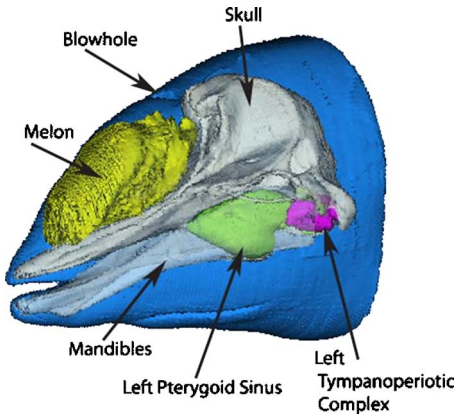


FIG. 1. (Color online) Left anterolateral view of the head of a female neonate Cuvier's beaked whale (*Ziphius cavirostris*). The image was reconstructed from CT scans by segmentation of the anatomical structures. The structural color scheme is as follows: skin=blue; skull=white; mandibles =light blue; melon=yellow; left pterygoid sinus=green; left tympanoperiotic complex=magenta.

initial values $U^{(0)}$ and $V^{(0)}$ (and also the initial stress) are determined by the initial conditions.

Hexahedral isoparametric eight-node brick elements are used throughout the mesh. The use of the simple isotropic constitutive equation allows us to formulate an effective computational procedure to deal with almost incompressible materials. In the materials that possess nonzero shear modulus (connective tissue, muscle, and fats) volumetric locking is mitigated by using the selective reduced integration procedure of Hughes (2000) (one-point integration of the volumetric response, and full $2\times2\times2$ quadrature of the shear terms). The volumetric term in the elements that represent water (zero elastic shear stiffness) are stabilized by mixing a very small fraction of fully integrated forces with the one-point-rule integrated response.

III. CASE STUDY—*Ziphius cavirostris*

The head of a neonate Cuvier's beaked whale (specimen NMFS field number KXD0019) *Ziphius cavirostris*, from which tissue properties were measured by Soldevilla *et al.* (2005), is examined in the present computational study (Fig. 1). We compute the mechanical response of the soft tissues of the head, including those associated with peripheral sound reception. The acoustic stimulus was a series of planar sound waves from the right side of the computational box with a given frequency (3500 Hz), and a given intensity (sound pressure level (SPL) =180 dB re 1 μPa). The characteristics of the acoustic stimulus roughly match estimates of potential maximum received sound pressure levels during exposure to mid-frequency sonar associated with the Bahamas beaked whale strandings (NOAA, 2001).

Several structures within the head of the neonate Cuvier's beaked whale (Fig. 1) are of interest in the production or reception of sound (Cranford and Amundin, 2003). Sound production is associated with air movement within a set of sacs located beneath the blowhole, between the melon and the skull. The melon is composed of fatty tissues and acts as an acoustic channel for sound propagating out of the head. The lower half of the head contains the mandibles, acoustic fats associated with sound reception, air-filled pterygoid sinus, and the bony ear (tympanoperiotic) complex. The juxtaposition of the air-filled sinus and the bony ear complex represents the greatest impedance mismatch within this animal.

To construct an anatomical model, the head of the neonate Cuvier's beaked whale was scanned with x-ray computed tomography (CT), as reported by Soldevilla *et al.* (2005). The data were collected continuously with 152 transverse scans along the longitudinal axis. Each of these transverse scans was 5 mm thick, collected every 5 mm. The GE Lightspeed scanner used a 500 mm diameter field of the view scan region and a scanning protocol as reported in Soldevilla *et al.* (2005). In addition to the anatomic data, tissue samples from the neonate Cuvier's beaked whale were studied for physical properties including density, sound velocity, and Young's modulus.

A. Data processing

The computational domain is a rectangular three-dimensional box representing a volume of water with the specimen inside. The dimensions of the computational box can be found in Fig. 2 (all dimensions in millimeters). The anatomy is defined by a three-dimensional array of voxels as generated by a CT scan (referred to hereafter as raw data). The voxel values are in the Hounsfield units that may be mapped to material density and other material parameters as described by Soldevilla *et al.* (2005). The matrix size is 512×512 pixels in each transverse or cross-sectional scanning plane. The resolution within each transverse plane is 1.5 mm square pixels.

The original CT data have been minimally processed to produce the input three-dimensional (3D) voxel array. All of the voxels external to the boundary of the head were converted to a Hounsfield value corresponding to seawater. Thus, the specimen is, in this simulation, immersed in an environment of sea water near sea-surface pressure. The addition of water outside the specimen ensures adequate space between the specimen and the bounding box for the simulations. In this study we provided a space of approximately 30 to 80 mm between the specimen and the bounding box.

Another adjustment to the CT data was applied to the air sinuses that occur normally within the head and air-filled voids created by the postmortem nature of the specimen. The pterygoid sinuses and any air-filled voids were assigned an artificial Hounsfield value of −2000 (a value not found in the original CT image data) so that these voxels could be recognized and dealt with appropriately during the mesh generation process. Consequently, we could fill in voids that are not normally present, for example, those left after blood drained from them. This step also provided a means to access these internal regions, separately or collectively, in order to test ideas of their effect on the system. No further editing of the CT data was undertaken for this study.

B. Generation of the finite element meshes

The input 3D voxel array was resampled using linear interpolation to produce voxels of desired resolution. The

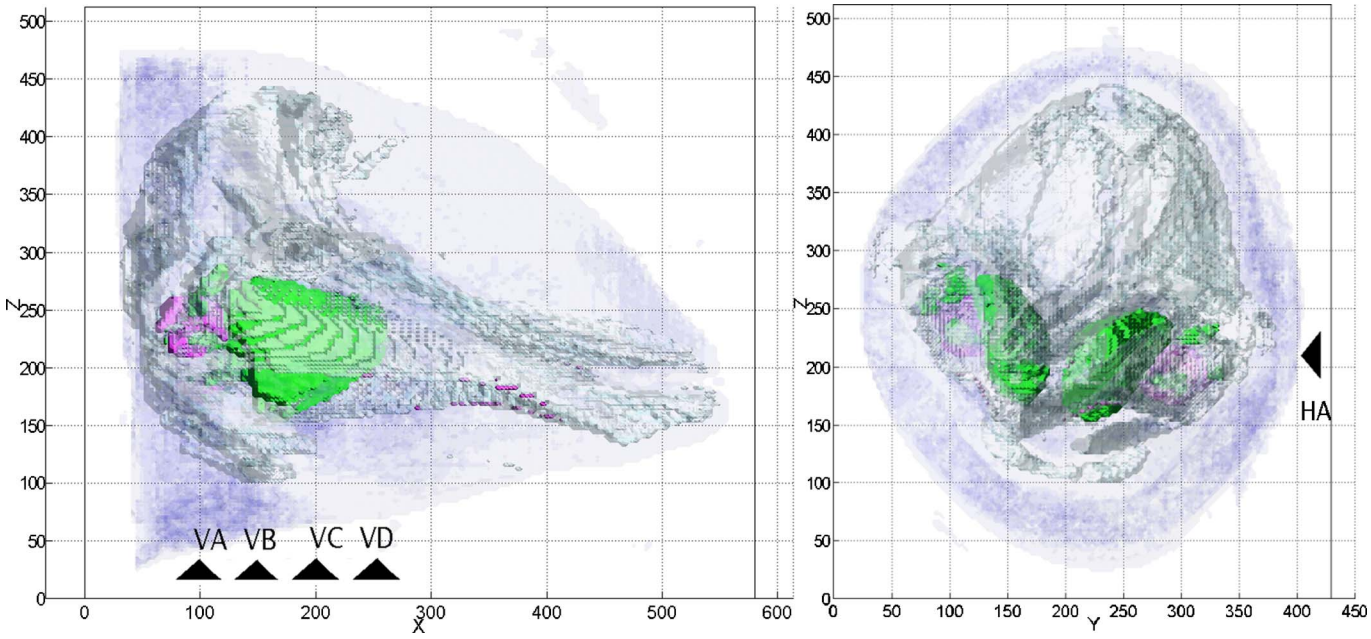


FIG. 2. (Color online) The head within the computational volume or "box." The frontal view is in the right panel, and right lateral view is in the left panel. The soft tissues are displayed with some transparency to reveal the internal structure, especially the pterygoid sinuses (air cavities) in green, and the high-density tympanoperiotic complexes (ear bones) in magenta. The horizontal slice location is indicated with "HA" in the right panel. The vertical slice locations (perpendicular to the longitudinal axis of the animal) are indicated with "VA," "VB," "VC," and "VD" in the left panel. The geometry is shown at the sample resolution corresponding to the finest mesh used in this paper $N=152$. Some visible artifacts appear, for instance the topographic curves along the surface of the mandible in the left panel.

voxels of the resampled array were then converted directly into finite elements, each of which was assigned a single material property based on the value of the x-ray attenuation Hounsfield unit in the voxel.

Four different meshes have been used, with N number of cells in the longitudinal direction $N \in \{80, 99, 123, 152\}$. Correspondingly, the total number of finite element nodes varied between 340,000 and 2.5 million, and the spatial resolution varied between 6.91 and 3.61 mm. The finite elements were nearly cubic. In the transverse plane the finite elements are lower resolution than is available from the original scan; in the longitudinal direction the resolution is comparable (5 mm in the scan compared to 6.91 to 3.61 mm for the finite elements).

C. Initial and boundary conditions

For simplicity, the computational box is assumed initially at rest, and unstressed. Such an assumption makes it easy to reconcile the boundary and the initial conditions. It also implies that the specimen and its bounding box are near the sea surface and that the system is not exposed to significant hydrostatic pressure. In this case, we used the geometry of the air sinuses as we found them. The geometry of air spaces, such as the pterygoids or peribullary sinuses (Fraser and Purves, 1960) will be different at depth, depending upon the pressure and other factors determined by the animal during a dive cycle. We did not test a variety of geometric configurations for partially collapsed air sinuses during this initial study, although this is planned for future studies.

Along all the bounding surfaces of the computational box, all three velocity components are prescribed as $v_x(t)=0$ and $v_z(t)=0$, and

$$v_y(y,t) = v_A[1 - \exp(-\Sigma t^2)]\sin(\gamma y - \varpi t),$$

where $\gamma = 2\pi f/c$ is the wave number; $c=1507\ \mathrm{ms}^{-1}$ is the sound speed in seawater at 15 °C; y is the left-to-right coordinate; ϖ is the angular frequency; $f=3500$ Hz is the frequency of the sound signal; $\Sigma = 6.1 \times 10^7\ \mathrm{s}^{-2}$ is a constant related to the signal ramp-up; $v_A = p_A/\rho c$ is the velocity amplitude of the plane wave generated by pressure with amplitude $p_A = 10^3$ Pa (corresponding to SPL of 180 dB re1 μPa). The initial conditions are consequently $\bar{v}(x)=0$, and $\bar{\sigma}(x)=0$. These initial and boundary conditions correspond to plane sound waves propagating in the left-to-right direction (transverse) with respect to the animal, with an exponential ramp up from a rest/unstressed state, up to full power within a fraction of a millisecond.

D. Material parameters

The material parameters needed for the isotropic constitutive equation described earlier are the density ρ, bulk modulus K, shear modulus G, and dynamic viscosity η. The sample density can be mapped from the CT image using a conversion from the Hounsfield units (Soldevilla *et al.*, 2005). Since the dynamic viscosity is not available as a map of the Hounsfield units, and needs to be estimated from the literature, we use the map proposed by Soldevilla *et al.* (2005) to assign representative "average" mechanical properties to tissues in the following groups: hard bone, soft bone, connective tissue, muscle, and acoustic fats/blubber.

To derive the stiffness moduli from the Young's modulus E as measured by Soldevilla *et al.* (2005), we compute the

TABLE I. Tissue parameters used for model.

Tissue type	Mass density ρ (kg m^{-3})	Bulk modulus K (MPa)	Shear modulus G (MPa)	Dynamic shear viscosity η (Pa s)
Hard bone	3000	2772	2083	900
Soft bone	2000	83	38	100
Connective	1075	2821	0.041	20
Muscle	1000	2102	0.033	20
Fats	950	1861	0.022	20

bulk modulus K from the speed of sound (connective tissue 1620 ms^{-1}, muscle 1456 ms^{-1}, and acoustic fats/blubber 1400 ms^{-1}) using

$$c_1 = \sqrt{\frac{K}{\rho}},$$

and further the Poisson's ratio ν from the formula

$$K = \frac{E}{3(1-2\nu)}.$$

So far, the viscosity values of these tissues have not been measured, and we employ values that are representative of the viscosities obtained experimentally for human [breast: $0.55 \leq \eta \leq 4.0$ Pa s; see Sinkus *et al.* (2005)], and bovine [biceps femoris: $3 \leq \eta \leq 17$ Pa s; Catheline *et al.* (2004)] tissues, and which are similar to the high content of high viscosity fats in the tissues of the odontocetes. It is noteworthy that this parameter is likely important and needs to be experimentally determined.

We assume Young's modulus and Poisson's ratio of $E=5$ GPa and $\nu=0.2$ for the hard bone, and $E=0.1$ GPa, and $\nu=0.3$ for the soft bone (because the specimen is a neonate, and the soft bone is mostly cartilaginous in a neonate mammal: These properties are estimates taken from Vincent, 1990). The damping in the bone has to be estimated from published data, in particular for human and bovine bones. The loss tangent for the bones in the kilohertz range is taken from Garner *et al.* (2000) as $\tan\delta \approx 0.01$, which gives the dynamic viscosity η using the approximate formula, $\eta \approx G\delta/2\pi f = E\delta/4(1+\nu)\pi f$, where f is taken as the forcing frequency.

For the ***tissues*** we obtain the set of parameters in Table I.

For the surrounding ***sea water*** we take: $\rho = 1000$ kg m^{-3}, $K=2102$ MPa, $G=0$ MPa, and $\eta = 2 \times 10^{-3}$ MPa s^{-1}, which corresponds to a strongly viscous Newtonian fluid. The high viscosity is used to damp out reflected and scattered waves. Note that the pressure waves generated by the motion of the boundary (i.e., the external forcing) are also being damped, but since these are the forced part of the motion of the water, the effect of damping is minor. The reflected and scattered waves have a strong shear component which is effectively damped by the high viscosity of the surrounding water.

The pterygoid sinuses are assumed to be filled with air at atmospheric pressure (sea level air pressure), hence of negligible stiffness (air is much more compressible than water), and consequently the interior of these cavities is not modeled. This is presumed not to introduce significant error, since the lowest resonant frequencies of cavities of this size are on the order of 100 Hz. Using a free-spherical air bubble model with a 10 cm radius and atmospheric pressure one obtains an underestimate of 33 Hz of the natural frequency, and various models correcting for additional stiffness are available, including a heuristic relationship for a whole-lung resonance of Cudahy *et al.* (1999) [refer to Finneran (2003) and references therein]. These resonant frequencies are significantly lower than our chosen forcing frequency of 3500 Hz.

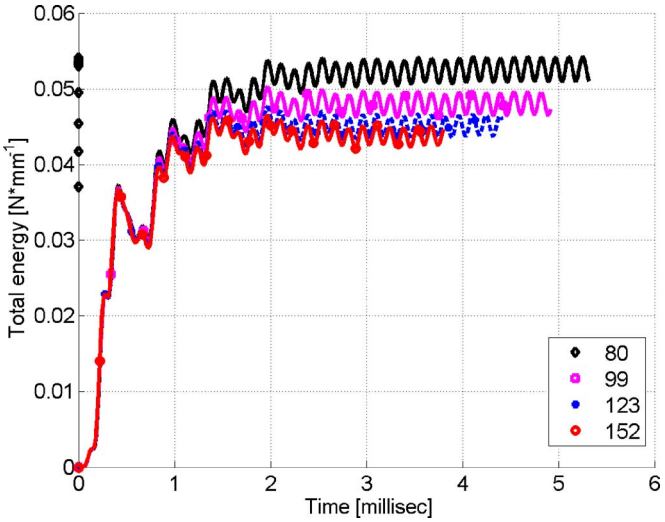


FIG. 3. Total energy in the computational box, including the surrounding water, as a function of time. The legend refers to the number of finite elements in the longitudinal direction N. The curves are ordered top to bottom as in the legend: the coarsest mesh $N=80$ is stiffer than the finest $N=152$.

For each resolution, the run was terminated when for a set number of cycles the total energy in the computational box oscillated around a constant value, indicating that a steady-state had been reached (hence at a different time instant for each resolution). A single processor personal computer (PC) with 2 GB of memory and a 3.4 GHz Pentium 4 CPU was used for these runs, with the longest run on the order of 40 CPU hours.

E. Results

The results of the simulations include the rate of steady-state energy dissipation at various spatial resolutions, maps of pressure distribution, heat distribution and dissipation, stress/strain regimes, and strain rates.

1. Total energy at steady state

The time variation of the total energy (potential plus kinetic) for the four mesh sizes used is shown in Fig. 3. Richardson's extrapolation (Roache, 1998) is used to estimate the convergent value of the total energy to assess the quality of the computed results. The normalized error in total energy may be defined as

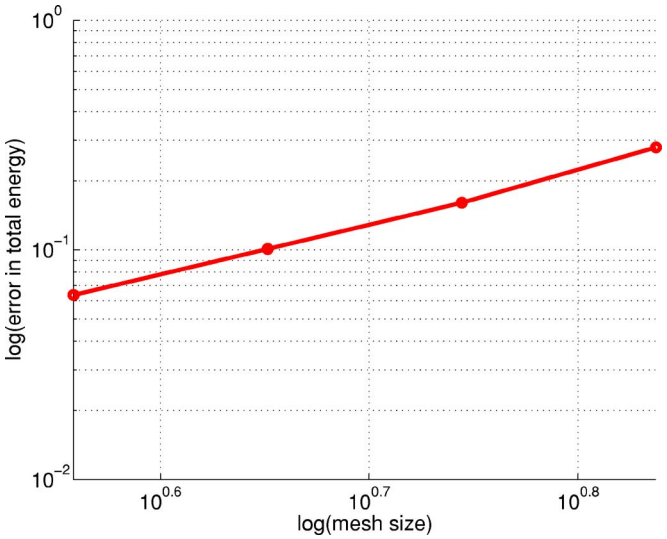


FIG. 4. The convergence plot for the total energy at steady state. Richardson extrapolation was used to estimate the converged total energy. The convergence rate is approximately 2.19.

$$e_h = \left| \frac{\Pi_R - \Pi_h}{\Pi_h} \right|$$

where Π_R is the estimated converged value; Π_h is the solution for mesh size h. The convergence graph is shown in Fig. 4, which indicates that the approximation error in the total energy is about 7% for the finest mesh. Clearly, accuracy in local quantities (as opposed to the global quantity—energy) is going to be lower. Therefore, more resolution (finer spatial mesh) is needed, and a transition to high-performance computer architectures (work in progress) would give us the ability to refine the computational meshes further, increasing accuracy in the results, including the bounds on the local quantities, for instance strain or dissipated energy density.

2. Pressure

Figure 5 shows the distribution of sound pressure for the finest mesh in a horizontal slice at the level of the sinuses and the ear bones (slice HA as shown in Fig. 2). The anatomic structures are displayed in color as surfaces for visualization purposes. Note that the highest pressure within the head, represented by a region of light gray pixels, is approximately 183 dB re: 1 μPa and is concentrated around the posterior aspect of the left ear (upper left region of Fig. 5). There is also a ridge of high pressure along the bottom of Fig. 5, which is the result of the acoustic stimulus planar wave entering the computational box from that side. In addition, there is a slight low-pressure depression around the posterior region of the right ear (Fig. 5). Note that the illustrations of pressure distribution are snapshots of a time-dependent process, where the maxima and minima at different locations are not synchronized.

The complex distribution of sound pressure is further illuminated in Fig. 6. It shows the pressure map at the same time instant as Fig. 5, but in transverse slices (perpendicular to the longitudinal axis of the animal), at various positions along the pterygoid sinuses (locations VA, VB, VC, and VD

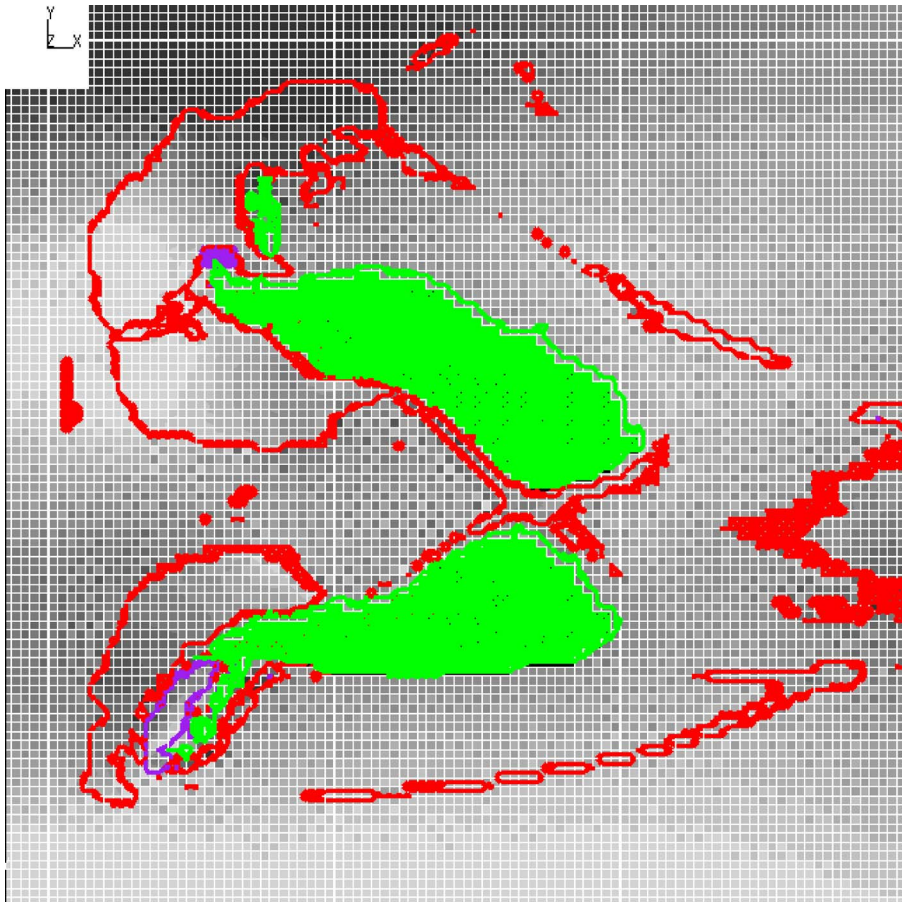


FIG. 5. (Color online) Pressure distribution for horizontal slice (HA) through the sinuses and the ear bones after the steady state has been reached. The left side of the animal is toward the top of the frame and the front of the animal is to the right of the frame. Light shade corresponds to positive pressure, dark shade corresponds to negative pressure, the pterygoid sinuses are shown in green, and the bony ear complexes in purple. The outlines of the skull bones are in red.

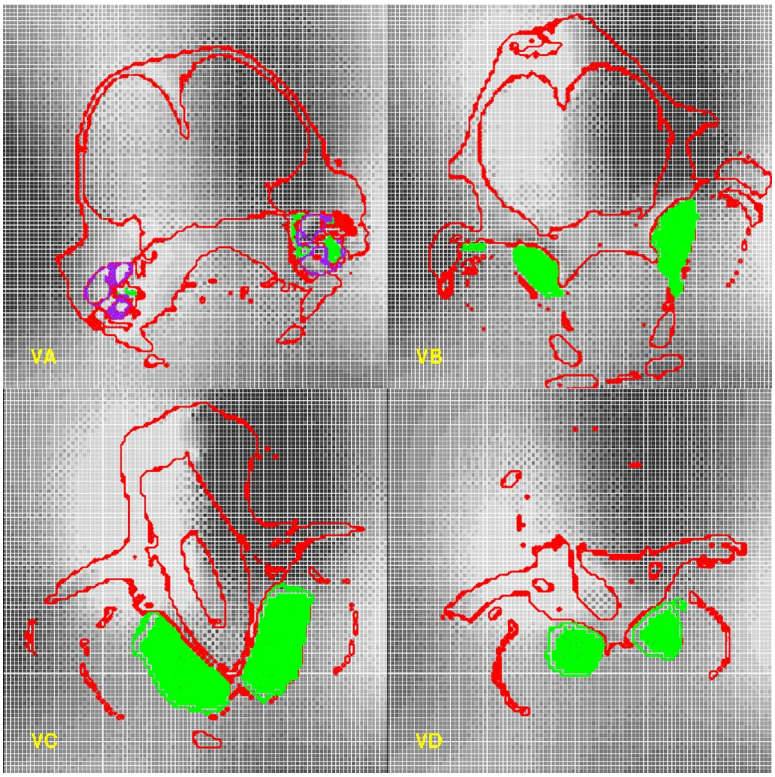


FIG. 6. (Color online) Pressure distribution at the same time as in Fig. 5. Vertical slices at different stations in the longitudinal direction along the sinus. (The locations of slices VA, VB, VC, and VD, are noted in Fig. 2. Light shade corresponds to positive pressure, dark shade corresponds to negative pressure, the pterygoid sinuses are shown in green, the bony ear complexes in purple, and the skull bones are outlined in red.

as shown in Fig. 2). The interaction of the sound pressure wave with the skull and other anatomical structures is clearly visible. For example, observe that the pressure generally varies more around the skull than in the soft tissues at the ventral side of the animal.

Note that at the frequency of 3500 Hz, the maximum pressure difference occurs within the space of the head. This can be seen as the darkest region on the right side of the head and the lightest region on the left side of the head (Fig. 6). To verify that this distribution of pressures is not just the "effect of the box," a simulation has been performed where the box was filled with seawater only. In this case, the distribution of pressures agreed with the prescribed motion (propagating planar wave). Consequently, we may conclude that the focusing of pressure is due to the presence of the tissues and not the dimensions of the box.

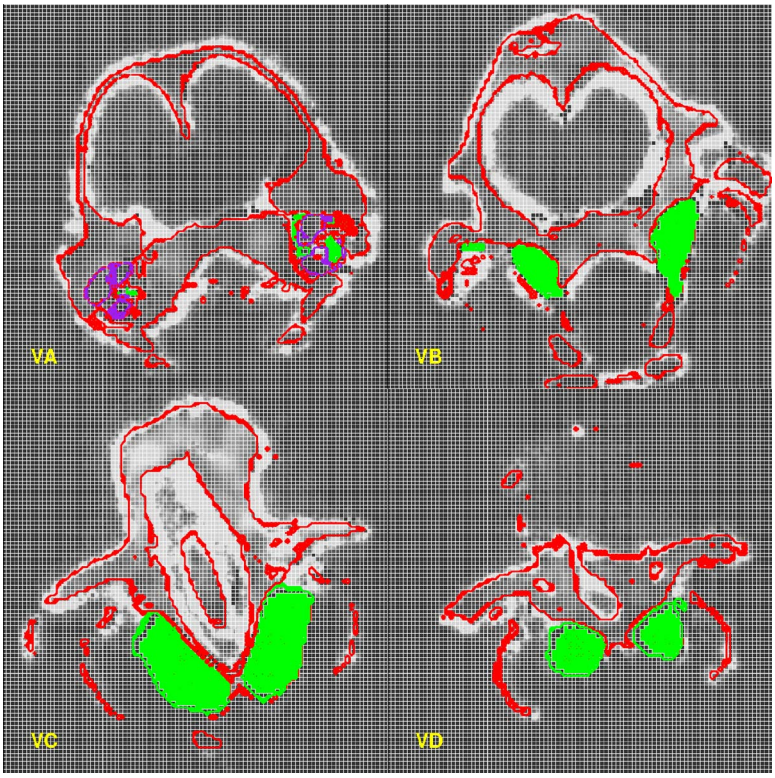


FIG. 7. (Color online) Energy dissipation density. Vertical slices at different stations in the longitudinal direction. (The location of each slice VA, VB, VC, and VD is shown in Fig. 2.) Light gray shades correspond to high density of energy dissipation, darker shade corresponds to lower density. The pterygoid sinuses are shown in green, the bony ear complexes in purple, and the skull bones are outlined in red.

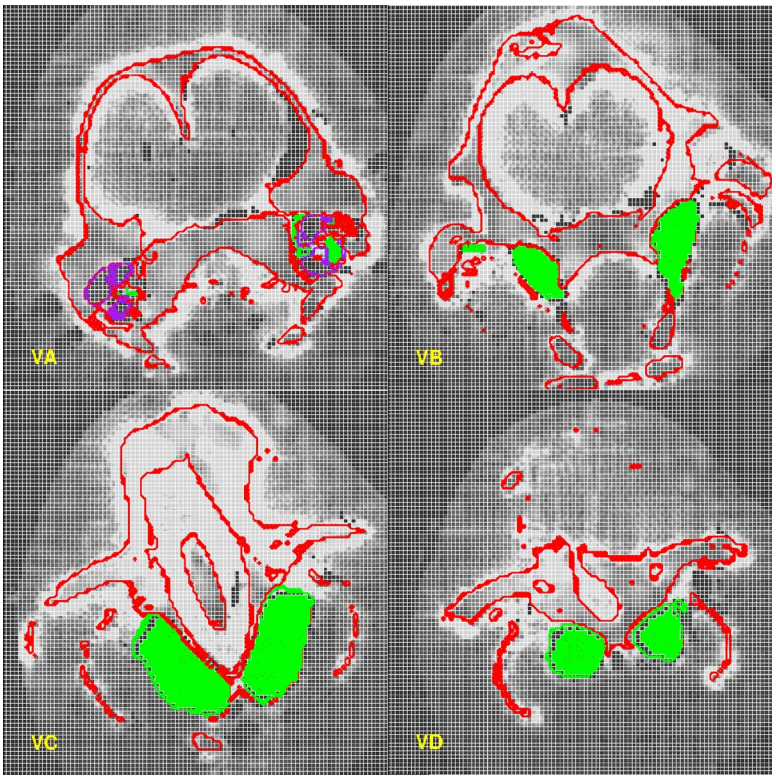


FIG. 8. (Color online) Maximum principal stretch displayed on vertical slices through the mesh. (The locations of the slices are displayed in Fig. 2.) High values (light shades) occur in the vicinity of the bone. Note that the distribution of the maximum principal stretch correlates with the distribution of the highest density energy dissipation. The pterygoid sinuses are shown in green, the bony ear complexes in purple, and the skull bones are outlined in red.

It also appears there is some "shielding" of acoustic pressure across the lower half of the head when compared to the upper half of the head, and appears as if it may be related to the presence of the pterygoid sinuses.

3. Local heating effects

It is known that the absorption of ultrasound at the bone/soft-tissue interface may lead to significant transient temperature rise (Myers, 2004). In the present case, the sound frequencies are much lower, but the mechanisms are similar: damping processes, primarily viscous, dissipate energy which is converted into heat. Therefore, as a first step one may look at the rate of energy dissipation to assess possible temperature increase as a result of local heating. Figure 7 illustrates the distribution of the dissipation energy density. Evidently, most of the energy is being dissipated in the soft tissue adjacent to the bone. In particular, it is interesting that the regions with the highest density of energy dissipation are within the brain at the edge of the braincase in VB of Fig. 7. In addition, panel VA of Fig. 7 shows a high-density region at the point where part of the brain (Vestibulo-cochlear nerve) passes through the internal auditory canal of the skull. This is the same region where a hemorrhage was found in a Blainville's beaked whale *Mesoplodon densirostris,* which stranded in association with naval sonar usage (NOAA, 2001).

These observations may also be correlated with the distribution of the *maximum principal stretch* (MPS). At any given point and time instant, the MPS is the largest of the stretches that the material experiences in any direction at that point. If we then consider the MPS for *all* points in the material, the greatest value is the quantity we refer to as the *largest* maximum principal stretch. Figure 8 demonstrates that the largest maximum principal stretch generally occurs next to the bones of the skull, including the mandibles.

The slope of the least-squares fit to the curves that express the time dependence of the largest local density of energy dissipation gives the maximum point wise *rate* of the dissipated energy density. An estimate of the maximum temperature rise rate can be estimated through the following expression (obtained from the linear heat conduction equation by neglecting heat diffusion and heat convection effects):

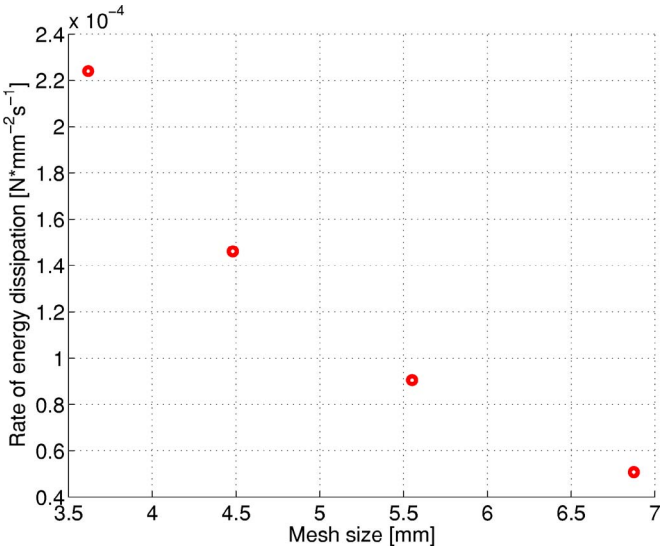


FIG. 9. Convergence of the local maximum of the point wise dissipated energy density. Finer meshes result in higher energy dissipation.

$$\dot{T} \approx \frac{\Phi_d}{\rho c_p},$$

where $\dot{T}$ is the rate of temperature increase; Φ_d is the density of the rate of dissipation; ρ is the mass density; and c_p is the specific heat at constant pressure. By reference to Fig. 9, we can estimate the maximum local rate of the density of energy dissipation as $\Phi_d \approx 4.0 \times 10^{-7}$ W mm^{-3}, which together with an estimate of the specific heat of soft tissues that is commonly used in the therapeutic ultrasound literature $c_p \approx 3500$ J kg^{-1} K^{-1}, and the mass density $\rho \approx 1000$ kg m^{-3}, yields the maximum local temperature increase rate $\dot{T} \approx 0.6 \times 10^{-4}$ K s^{-1}. Evidently, because heat diffusion and heat convection are being neglected this is an overestimate. Therefore, the temperature increase within the tissues due to sound excitation of expected exposure times on the order of 100 s appears to be a few millidegree Kelvin.

4. Injurious strains

The strain data is of lower accuracy than the energy data because strain results will depend strongly upon mesh size. We conjecture that there are no true stress singularities in the solution, and hence that the strain is finite everywhere. Consequently, the strain amplitudes should converge to finite values. However, for the strains the Richardson extrapolation is not available because our strain results are not in the asymptotic range yet, and may not be used to estimate converged values. To put the representation of the strains into context, consider that to begin to resolve the shear waves with speed of sound $c_s = \sqrt{G/\rho} \approx 5.7$ ms^{-1} (muscle), mesh resolution of approximately 0.4 mm would be required, whereas, the finest mesh size we used was 3.61 mm.

Nevertheless, since the response of the soft tissues at steady state is a mixture of dilatational modes and shear modes, some observations are possible based on the present range of resolutions (3.61–6.91 mm). Figure 8 illustrates the distribution of the *maximum principal stretch* at a particular time instant after the steady state has been reached. To quantify (see Fig. 10), the maximum principal stretch is approximately 1.5×10^{-4}.

IV. DISCUSSION

Our most important result is the success in combining CT scanning, tissue measurements, and finite element analysis tools to simulate the interactions between sound waves and anatomy. The innovation of combining these techniques has allowed us to ask questions that were not possible in previous studies (Aroyan *et al.*, 1992; Aroyan, 2001). For example, within the range of the physical parameters we tested, heating of the tissue and stretch (extensional strain) appear not to be factors that would lead to tissue damage in a neonate Cuvier's beaked whale. Although, wider ranges of the parameter values will need to be simulated to fully explore the possibility of injurious effects of these two factors. Another advancement from this work is the ability to pull CT data directly into the simulation software, with little or no need for data manipulation.

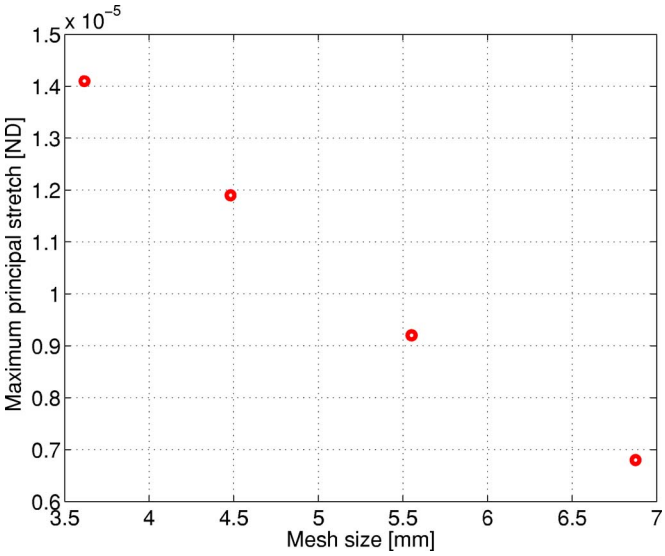


FIG. 10. Maximum point wise largest principal stretch during the steady-state vibration. Four different meshes.

At the intensity we tested (SPL=180 dB at 1 μPa), there is no indication of a significant deleterious effect on tissues from heating or strain. In ultrasound applications, a one minute exposure to ultrasonic frequencies can result in an increase of 6 °C, which is thought to have no deleterious effects on tissues (Hedrick *et al.*, 2005). In our case, the temperature increase due to a one minute exposure would probably amount to no more than a fraction of a degree Celsius. As such, temperature increase in tissue due to acoustic exposure is unlikely to be significant, with respect to possible injurious effects of strains alone, and within the acoustic parameters we tested and the restrictions noted above.

Biomechanics studies have established that injurious strains typically result when the stretch exceeds strains on the order of 10%. For instance, Sunderland (1978) published extensive work that show that nerve cells can typically sustain stretches of 20% while still functioning in an "elastic regime." The stretches in the present study (Fig. 8) are much lower, however, they are derived for a limited range of excitation parameters. If we extrapolate (under the assumption of linearity) from 180 to 220 dB re: 1 μPa, the pressure increases with a factor of 100. Correspondingly the strain amplitude may be expected to increase also with a factor of 100, placing the maximum stretch in the range of 1%. In addition, it is important to realize that at this point the strains produced in our model are not accurate owing to limitations in the mesh size, and we do not have a quantitative estimate of errors in strains. It is conceivable that our computed maximum stretches are in error, being underestimated.

One of the most interesting results of the simulations was the acoustic pressure distribution across the specimen. First, the maximum pressure differential is found within (across) the head in every transverse plane (perpendicular to the plane of the sound source) shown in Fig. 6. At this point in our investigation, we cannot determine whether these steep pressure gradients may have effects on sensory systems within the head. The dimensional extent of the maximum

differential pressure gradient is probably a function of the acoustic wavelength, which is related to the selected frequency (3500 Hz).

Another aspect of the simulated results of acoustic pressure distribution suggests that it might be possible to determine computationally whether a shielding or acoustic shadowing effect for the bottom half of the head exists. This might indicate a benefit if it confers protection from acoustic exposure to the hearing apparatus. With the present resolution it is possible to observe that the acoustic pressure varies more strongly at the top of the head, likely due to the presence of relatively stiff bones, but this issue deserves a more detailed examination.

V. CONCLUSIONS

The most promising outcome from this work is the broad spectrum of questions that can now be asked and answered. Prior to merging these techniques it was not possible to repeatedly query an anatomic system with questions of acoustic impact. At the given received level we tested (SPL=180 dB re: 1 μPa), there is no indication of a significant deleterious effect on the tissues. At the same time, neither can we determine what the simulations might reveal at higher intensities and/or different frequencies or mixtures of frequencies. However, the effect of a nonlinear dynamics upon the resonant vibrations has not been investigated. These are clearly interesting avenues for future investigations.

The effects of hydrostatic pressure on the geometry of air spaces, such as the pterygoids or peribullary sinuses, as described by Fraser and Purves (1960) will be different according to pressure and other factors determined by the animal during a dive cycle. With increasing hydrostatic pressure, the air within the sinuses may be systematically squeezed from the pterygoid sinuses to their respective peribullary sinuses. This is likely accomplished as blood fills the venous plexus and occludes the pterygoid sinuses as the air volume is compressed by the increased hydrostatic pressure with depth. We did not test a variety of geometric configurations for partially collapsed air sinuses during this initial study, but this is planned for future studies. As a consequence of the way we treat the internal sinuses as a special airspace (by assigning a unique Hounsfield value to the voxels corresponding to the sinuses), it may be possible to manipulate the size and shape of this internal cavity.

The source is lateral to the specimen in the example studied thus far. It would be interesting to investigate the effect of placing the source above or below the specimen, and how the acoustic shadowing and other parameters may change with frequency and intensity. It should be realized however that resonant vibrations of the tissues are excited by nonzero forcing irrespective of the direction of the oncoming sound wave. The resonant motion would not be generated only if the work-conjugate forcing identically vanished, which is statistically unlikely.

This specimen was CT scanned twice, once frozen and once thawed. Future studies should include a comparison of these data sets to understand the implications of the values that frozen tissue yields compared to thawed tissue. These two different conditions may help illuminate the implications of specimen geometry variations.

Increasing the resolution of the model significantly by uniform refinement (e.g., to finite element linear dimension of 0.5 mm) would improve the ability of the model to handle shear wave propagation. Likewise it would mean a substantial growth in the computational cost (computational grid of approximately 62 billion elements, compared to the 2.5 million elements used in this study). However, with adaptive, targeted refinement, the computational cost may be kept in check (Krysl *et al.*, 2003, 2004). A computational procedure of this nature is under development.

The treatment of the (semi-) infinite extent of the sea water environment is in this work only approximate. Use of infinite elements, coupling with boundary elements, and perfectly matched layers approximation are under consideration as alternatives.

It should also be noted that finite element simulation and analysis is an iterative process, beginning with a simplified model and moving to more complex objects and parameters with repeated iterations. This study represents the first step in a series. The anatomy and the range of parameters explored are necessarily simplified. So, these results can be understood as cracking the door open to reveal the most basic answers to the interaction between underwater sound and beaked whale anatomy. At the same time, the success of this first attempt encourages us to open the door more widely during successive iterations and sort through the cornucopia of remaining questions.

Numerical methods have been applied to simulate the effects of acoustic exposure in a complex biological system: the head of a neonate beaked whale *Ziphius cavirostris*. The presented approach offers promising prospects for future investigations for a variety of types of sounds and a variety of cetacean species.

ACKNOWLEDGMENTS

This work was supported by the U.S. Navy CNO-N45, with project management by Frank Stone and Ernie Young. The specimen of *Ziphius cavirostris* used in this study and in Soldevilla *et al.* (2005), was acquired with the help of Susan Chivers at the National Marine Fisheries Service (SWFC, San Diego) and is cataloged under the field number KXD0019.

Aroyan, J. L. (**2001**). "Three-dimensional modeling of hearing in *Delphinus delphis*," J. Acoust. Soc. Am. **110**(6), 3305–3318.

Aroyan, J. L., Cranford, T. W., Kent, J., and Norris, K. S. (**1992**). "Computer Modeling Of Acoustic Beam Formation In *Delphinus delphis*," J. Acoust. Soc. Am. **92**(5), 2539–2545.

Astley, R. J. (**2000**). "Infinite elements for wave problems: a review of current formulations and an assessment of accuracy," Int. J. Numer. Methods Eng. **49**, 951–976.

Balcomb, K. C. III and Claridge, D. E. (**2003**). "A mass stranding of cetaceans caused by naval sonar in the Bahamas," Bahamas J. Sci. **2**, 2–12.

Bayliss, A., Gunzberger, M., and Turkel, E. (**1982**). "Boundary conditions for the numerical solution of elliptic equations in exterior regions," SIAM J. Appl. Math. **42**, 430–450.

Brill, R. L., and Harder, P. J. (**1991**). "The effects of attenuating returning echolocation signals at the lower jaw of a dolphin *Tursiops truncatus*," J. Acoust. Soc. Am. **89**, 2851–2857.

Catheline, S., Gennisson, J. L., Delon, G., Fink, M., Sinkus, R., Abouelkaram, S., and Culioli, J. (**2004**). "Measurement of viscoelastic properties of homogeneous soft solid using transient elastography: An inverse problem approach," J. Acoust. Soc. Am. **116**(6), 3734–3741.

Cox, T. M., Ragen, T. J., Read, A. J., Vos, E., Baird, R. W., Balcomb, K., Barlow, J., Caldwell, J., Cranford, T., Crum, L., D'Amico, A., D'Spain, G., Fernández, A., Finneran, J., Gentry, R., Gerth, W., Gulland, F., Hildebrand, J., Houser, D., Hullar, T., Jepson, P. D., Ketten, D., MacLeod, C. D., Miller, P., Moore, S., Mountain, D., Palka, D., Ponganis, P., Rommel, S., Rowles, T., Taylor, B., Tyack, P., Wartzok, D., Gisiner, R., Mead, J., and Benner, L. (**2004**). "Understanding the Impacts of Anthropogenic Sound on Beaked Whales," J. Cetacean Res. Manage. **7**(3), 177–187

Cranford, T. W. (**1988**). "The anatomy of acoustic structures in the spinner dolphin forehead as shown by X-ray computed tomography and computer graphics," in: *Animal Sonar: Processes and Performance*, P. E. Nachtigall and P. W. B. Moore, eds., (Plenum, New York) pp. 67–77.

Cranford, T. W. and Amundin, M. E. (**2003**). "Biosonar Pulse Production in Odontocetes: The State of Our Knowledge," in *Echolocation in Bats and Dolphins*, J. A. Thomas, C. F. Moss, and M. Vater, eds. (The University of Chicago, Chicago) pp. 27–35.

Cranford, T. W., Amundin, M., and Norris, K. S. (**1996**). "Functional morphology and homology in the odontocete nasal complex: Implications for sound generation," J. Morphol. **228**, 223–285.

Cudahy, E. A., Hanson, E., and Fothergill, D. (**1999**). "Summary on the bioeffects of low-frequency waterborne sound," in Technical Report 3, Environmental impact statement for surveillance towed array Sensor system low-frequency active (SURTASS LFA) sonar.

Festa, G., and Vilotte, J. P. (**2005**). "The Newmark scheme as velocity-stress time-staggering: An efficient PML implementation for spectral element simulations of elastodynamics," Geophys. J. Int., **161**(3), 789–812.

Finneran, J. J. (**2003**). "Whole-lung resonance in a bottlenose dolphin (Tursiops truncatus) and white whale (Delphinapterus leucas)," J. Acoust. Soc. Am. **114**(1), 529–535.

Frantzis, A. (**1998**). "Does acoustic testing strand whales?," Nature (London) **329**, 29.

Fraser, F. C. and Purves, P. E. (**1960**). "Hearing in cetaceans: Evolution of the accessory air sacs and the structure and function of the outer and middle ear in recent cetaceans," Bulletin of the British Museum (Natural History) Zoology **8**, 1–140.

Garner, E., Lakes, R., Lee, T., Swan, C., and Brand, R. (**2000**). "Viscoelastic dissipation in compact bone: Implications for stress-induced fluid flow in bone," ASME J. Biomech. Eng., **122**(2), 166–172.

Hedrick, W. R., Hykes, D. L., and Starchman, D. E. (**2005**). *Ultrasound Physics and Instrumentation*, 4th ed. (Elsevier, New York).

Hildebrand, J. A. (**2005**). "Impacts of Anthropogenic Sound" in *Marine Mammal Research: Conservation beyond Crisis*, J. E. Reynolds *et al.*, eds. (The Johns Hopkins University Press, Baltimore, Maryland).

Hughes, T. J. R. (**2000**). *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis* (Dover, New York).

Krysl, P., Grinspun, E., and Schroder, P. (**2003**). "Natural hierarchical refinement for finite element methods," Int. J. Numer. Methods Eng., **56**(8), 1109–1124.

Krysl, P., Trivedi, A., and Zhu, B. Z. (**2004**). "Object-oriented hierarchical mesh refinement with CHARMS," Int. J. Numer. Methods Eng., **60**(8), 1401–1424.

Myers, M. R. (**2004**). "Transient temperature rise due to ultrasound absorption at a bone/soft-tissue interface," J. Acoust. Soc. Am. **115**(6), 2887–2891.

NOAA (**2001**). "Joint Interim Report Bahamas Marine Mammal Stranding Event of 14-16" March 2000. Washington, D.C., US Department of Commerce and US Navy, available at: www.nmfs.noaa.gov/prof-res/overview/Interim-Bahamas-Report.pdf.

Norris, K. S. (**1964**). "Some problems of echolocation in cetaceans." in *Marine Bio-acoustics* W. N. Tavolga, ed (Pergamon Press, New York) pp. 317–336.

Řeřicha, P. (**1986**). "Optimum load time history for nonlinear-analysis using dynamic relaxation," Int. J. Numer. Methods Eng., **23**(12), 2313–2324.

Roache, P. J. (**1998**). *Verification and Validation in Computational Science and Engineering* (Hermosa Publishers, Albuquerque, New Mexico).

Rommel, S. A., Costidis, A. M., Fernandez, A. J. F., Jepson, P. D., Pabst, D. A., McLellan, W. A., Houser, D. S., Cranford, T. W., van Helden, A. L., Allen, D. M., and Barros, N. B. "Elements of beaked whale anatomy and diving physiology, and some hypothetical causes of sonar-related stranding," J. Cetacean Res. Manage., in press.

Rubin, M. B., and Bodner, S. R. (**2002**). ""A three-dimensional nonlinear model for dissipative response of soft tissue," Int. J. Solids Struct. **39**(19), 5081–5099.

Sinkus, R., Tanter, M., Xydeas, T., Catheline, S., Bercoff, J., and Fink, M. (**2005**). "Viscoelastic shear properties of in vivo breast lesions measured by MR elastography," Magn. Reson. Imaging **23**(2), 159–165 Sp. Iss. SI.

Soldevilla, M. S., McKenna, M. E., Wiggins, S. M., Shadwick, R. E., Cranford, T. W., and Hildebrand, J. A. (**2005**). "Cuvier's beaked whale (*Ziphius cavirostris*) head tissues: Physical properties and CT imaging," J. Exp. Biol., **208**(12), 2319–2332.

Sunderland, S. (**1978**). *Nerves and nerve injuries*, 2nd ed. (Churchill Livingstone, Edinburgh).

Vincent, J. F. V. (**1990**). *Structural Biomaterials* (Princeton University Press, Princeton, NJ).

Wagner, M., Gaul, L., and Dumont, N. A. (**2004**). "The hybrid boundary element method in structural acoustics," Z. Angew. Math. Mech. **84**, No. 12, 780–796.

St. Lawrence blue whale vocalizations revisited: Characterization of calls detected from 1998 to 2001

Catherine L. Berchok[a)]
Graduate Program in Acoustics, The Pennsylvania State University, P. O. Box 30, State College, Pennsylvania, 16804-0030

David L. Bradley and Thomas B. Gabrielson
Applied Research Laboratory, The Pennsylvania State University, P. O. Box 30, State College, Pennsylvania, 16804-0030

(Received 11 December 2005; revised 7 July 2006; accepted 18 July 2006)

From 1998 to 2001, 115 h of acoustic recordings were made in the presence of the well-studied St. Lawrence population of blue whales, using a calibrated omnidirectional hydrophone [flat (±3 dB) response from 5 to 800 Hz] suspended at 50 m depth from a surface isolation buoy. The primary field site for this study was the estuary region of the St. Lawrence River (Québec, Canada), with most recordings made between mid-August and late October. During the recordings, detailed field notes were taken on all cetaceans within sight. Characterization of the more than 1000 blue whale calls detected during this study revealed that the St. Lawrence repertoire is much more extensive than previously reported. Three infrasonic (<20 Hz) and three audible range (30–200 Hz) call types were detected, with much time/frequency variation seen within each type. Further variation is seen in the form of call segmentation, which appears (through examination of Lloyd's Mirror interference effects) to be controlled at least partially by the whales. Although St. Lawrence blue whale call characteristics are similar to those of the North Atlantic, comparisons of phrase composition and spacing among studies suggest the possibility of population dialects within the North Atlantic. [DOI: 10.1121/1.2335676]

PACS number(s): 43.80.Ka [WWA] Pages: 2340–2354

I. INTRODUCTION

There are two main types of blue whale (*Balaenoptera musculus*) vocalizations reported in the literature: long duration, low-frequency calls that occur in highly patterned continuous series and short duration, higher-frequency calls that occur in more sporadically spaced groupings.

The low-frequency calls have been referred to by many different names: AB calls, 20-Hz pulses, commas, moans, songs, broadcast calls, and snapping shrimp (Kibblewhite *et al.*, 1967; Cummings and Thompson, 1971; Alling *et al.*, 1991; Cummings and Thompson, 1994; Nishimura and Conlon, 1994; Clark, 1995; McDonald *et al.*, 2001). They have been recorded in many areas of the world, and it is generally accepted that geographic variations exist (Thompson and Friedl, 1982; Clark, 1995; Rivers, 1997; Stafford *et al.*, 1999a, 1999b, 2001; Mellinger and Clark, 2003; McDonald *et al.*, in press).

The higher-frequency blue whale calls have also been described by a multitude of terms: D calls, downsweeps, FM downsweeps, short pulses, and arch sounds (Thompson *et al.*, 1996; Ljungblad *et al.*, 1997; Thode *et al.*, 2000; Mellinger and Clark, 2003). This call type also occurs worldwide but does not show the obvious geographic variation seen with the low-frequency vocalizations. Although these higher-frequency calls occur worldwide, they are mostly reported by studies conducted in coastal waters (Thompson *et al.*, 1996; Ljungblad *et al.* 1997; Teranishi *et al.*, 1997; Thode *et al.*, 2000; McDonald *et al.*, 2001; Bass and Clark, 2002). It should be noted, however, that with the exception of Stafford *et al.* (2001), many deep basin datasets have not yet been fully analyzed for this call type (Clark, 2003; Moore, 2005).

For simplicity, the low-frequency and higher-frequency blue whale call types will be referred to as infrasonic and audible calls, respectively. These terms are in reference to human hearing and are not meant to imply anything about the hearing ability of blue whales.

Prior to this study, the known vocal repertoire for St. Lawrence blue whales consisted of a single call type with little variation between calls (Edds, 1982). Although this dataset was small ($n=7$), the Edds recordings provided the earliest description of blue whale calls from the St. Lawrence and were the first to connect long-duration infrasounds to North Atlantic blue whales. The St. Lawrence has also been home to the longest biological field study of blue whales in the world. It was here the discovery was made that individual blue whales can be identified by their unique pigmentation patterns (Sears *et al.*, 1990). At the present time, 403 individuals have been identified, over 40% of these individuals have been genetically sampled through biopsy, and 26 years of behavioral data have been collected (Sears, 2005).

It was this wealth of biological information coupled with the limited amount of acoustical data for the St. Lawrence blue whale population that was the main impetus for this

[a)]Current address: Marine Physical Laboratory, Scripps Institution of Oceanography, 291 Rosecrans Street, San Diego, California, 92106. Electronic mail: cberchok@ucsd.edu

study. Another was an interest in learning where St. Lawrence blue whales go once they pass through the Cabot Straight into the North Atlantic. Although blue whales are seen regularly in the St. Lawrence from May until December, peaking from June through August, biological information is scarce for the North Atlantic (Sears and Calambokidis, 2002). However, acoustic recordings collected basinwide by the U.S. Navy's Sound Surveillance System (SOSUS) arrays have shown a concentration of blue whale vocalizations off the Grand Banks of Newfoundland from August through May, peaking from September until February (Clark, 1995). Analyses from similar arrays in the Pacific have revealed migratory patterns (Watkins *et al.*, 2000), which can be attributed to separate populations based on call characteristics (Stafford *et al.*, 1999a, 2001). If the St. Lawrence blue whale dialect is truly unique, it could be used to track this population as its members roam the North Atlantic.

Over 1000 vocalizations attributable to blue whales were detected over the four years of this study. Each call was characterized in terms of its frequency and time parameters, which were then used to organize the calls in categories. Statistics of the quantitative parameters for each category are listed along with the intercall interval lengths and patterning descriptions. The presence of call segmentation is also noted, and the influence of surface interference patterns on this segmentation investigated. In addition, this dataset is compared with recordings made in two recent North Atlantic studies (Mellinger and Clark, 2003; Nieukirk *et al.*, 2004), and the possibility of a regional dialect is discussed.

II. DATA COLLECTION, PROCESSING, AND ANALYSIS

A. Recording system

The calibrated recording system used for this study consisted of a single omnidirectional hydrophone (Geospace Corporation MP-18 piezoelectric transducer) suspended at a depth of 50 m from an 8 ft surface isolation buoy. The vertical motion of the system was further damped by attaching aluminum disks on the cable between the buoy and the hydrophone. Both boat-side and submerged preamplifiers were used to amplify and filter the received signal before it was recorded at a sampling rate of 44.1 kHz with a Sony PCM-M1 Digital Audio Tape (DAT) recorder. This provided a flat response with 3 dB down points at 5 and 800 Hz. Several attempts to increase the signal-to-noise ratio of the system were made over the course of this study by experimenting with different damping plate and circuitry designs. The majority of vocalizations were recorded on two main system types. The first (one 36 cm diameter plate, −130 dB *re* 1 V/μPa passband gain) performed well in calm water conditions, but changes in hydrostatic pressure caused by sensor motion in rougher seas generated large voltages that swamped the circuitry, riddling the recordings with signal cutout. The second system type (five 15 cm diameter plates, −140 dB *re* 1 V/μPa passband gain) provided a lower signal gain in calm seas than the first type, but effectively eliminated signal cutout in all sea states where recordings were attempted.

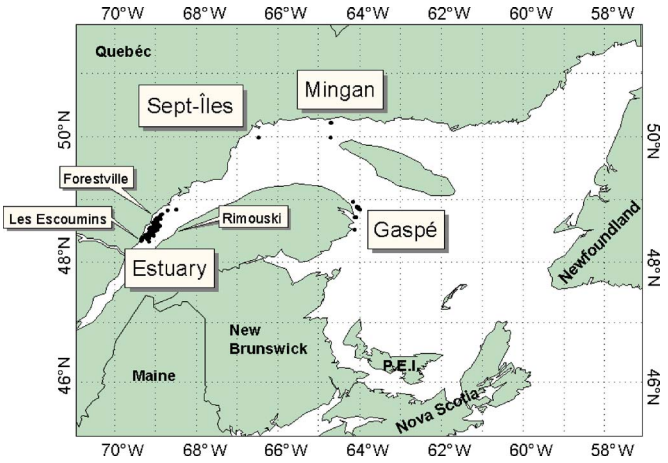


FIG. 1. (Color online) Locations of the Mingan, Gaspé, Sept-Îles, and Estuary study sites in the St. Lawrence of Quebéc. Black dots mark all hydrophone deployment locations. The majority of recording sessions occurred in the Estuary region bounded by the towns of Forestville, Rimouski, and Les Escoumins.

B. Field method

The field method for this study consisted of alternating periods of acoustic recordings with periods of photo-identification work. Close to 900 h were spent on the water throughout four field seasons (1998–2001), with over 100 h of acoustic recordings made in the presence of blue whales. Recordings were made at four different study sites in the St. Lawrence (Fig. 1), although the primary site was the estuary region from Les Escoumins to Forestville, Quebéc. Most fieldwork occurred between mid-August and late October. Effort was concentrated in the daylight hours, although some recordings were made a few hours after nightfall. A 16-ft rigid-hulled inflatable boat with a 70-HP two-stroke outboard motor was used as the observational and recording platform for most of the study.

Before each recording session, the research boat was positioned in an area where at least one blue whale was present, the engine turned off, and the recording system deployed. A Global Positioning System (GPS) receiver was used at ten-minute intervals throughout the session to determine the position of the research boat. During the session, detailed field notes were taken on the position (compass bearing and estimated distance), group size and composition, and behavior of all blue whales within sight. Notes were also taken on fin (*Balaenoptera physalus*), humpback (*Megaptera novaeangliae*), and minke (*B. acutorostrata*) whales, harbor porpoises (*Phocoena phocoena*), Atlantic white-sided dolphins (*Lagenorhynchus acutus*), and seals (*Halichoerus grypus, Phoca vitulina, and P. groenlandica*) sighted during the recording sessions. Both the time stamp of the DAT recorder and the visual field data were referenced to GPS time to synchronize acoustic and visual observations. Each recording continued until the blue whales moved out of visual range, weather conditions worsened, or an approaching cargo ship forced retrieval of the system. Between recording sessions, individual identities of the blue whales were confirmed when possible through the photo-identification methods described by Sears *et al.* (1990).

In addition to the acoustical and biological data collection, weather and sea conditions and all vessels within visual range were noted. A temperature profile to 100 m was also taken once a day, and a check of the propagation conditions was made through the use of a light bulb implosion point source [see Heard *et al.* (1997) for characteristics of imploding light bulbs].

C. Data processing and analysis

Prior to the call detection process, the recordings were transferred from DAT tape tracks to computer wave files. Each wave file was then electronically antialiasing filtered, downsampled by a factor of 30 or 300 (for either audible or infrasonic call analysis), and broken into two-minute segments to circumvent the memory limitations of MATLAB®. This process yielded two datasets with sampling rates of 1.6 kHz and 160 Hz, respectively. For both datasets, the time series and spectrogram of each two-minute segment were visually inspected for vocalizations as the segment was cycled through a series of digital FIR (Finite Impulse Response) bandpass filters and its audio track was played through headphones. This playback was sped up to reduce processing time and enable infrasonic calls to be heard.

The time series of all detected calls were extracted into a master data file along with their associated recording and signal processing information. From this master file, the filtering band was fine-tuned for each call, minimum/maximum frequency and beginning/end times of the call selected on the resulting spectrogram, and the time span and frequency band calculated. In addition, a trace line along the medial line of each spectrogram was generated by taking the highest-amplitude frequency of each time slice of the spectrogram (between the beginning and end time/frequency points of the call), smoothing with a five-point moving average, then fitting this curve with a set of straight line segments. These trace lines allowed for calculation of sweep rates as well as easier comparison of call shapes.

Received levels were calculated in terms of average call power, total call energy, and maximum rms power.[1] The average call power was calculated by taking the integral of the power spectral density of the windowed time series (boxcar with 5% of each end Hanning tapered, size=call time span) between the minimum and maximum frequency limits of the call. The ambient noise power was calculated before and after each call, and this average was subtracted from the average power of each call to give a corrected average power value. Multiplication of this corrected average power value by the time span of the call gave the corrected total call energy. For calculation of the maximum rms power, a spectrogram of the filtered call was computed (Infrasounds: 512 point FFT, 95% overlap, 499 points zero padding; Audibles: 1024 point FFT, 97.7% overlap, 4096 points zero padding; all use Hanning window), and the time slice containing the maximum value was found. The maximum rms power of the call was then calculated by taking the integral of the power spectral density of this time slice.

In addition to these quantitative measurements, the calls were sorted qualitatively into call type categories and assigned call rankings from 1–5. In assigning the rankings, two criteria were used: whether the quality of the signal was sufficient to identify it as a legitimate call, and whether the quality of the signal was sufficient to provide an accurate measurement of its parameters. Calls of Rank 1–3 met both criteria but varied in the amount of noise present (with Rank 1 calls having the least amount of noise). Rank 4 calls have very low signal to noise ratios and were sometimes recognized only through contextual clues. These Rank 4 calls were not used for the signal description statistics due to their poor quality, but they were still definite blue whale calls and were therefore used to calculate intercall intervals and patterning. Calls of Rank 5 did not meet either criterion and so were left out of any statistical analysis. After the calls were assigned call types, they were arranged sequentially to calculate intercall intervals and examine call patterning.

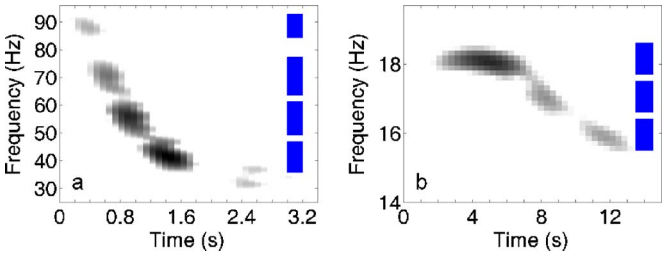


FIG. 2. (Color online) Examples of call segmentation in (a) infrasonic and (b) audible calls. A time-collapsed representation of this segmentation (used in Fig. 3) is located to the right of each call. Spectrogram parameters were (a) 1024-point FFT, 97.7% overlap, Hanning window with 4096 points zero padding, and (b) 512-point FFT, 95% overlap, Hanning window with 499 points zero padding.

D. Interference pattern (Lloyd's mirror) analysis

Another call characteristic observed in this study was a variation in amplitude that gave a segmented appearance to the calls (Fig. 2). This segmentation can be produced by the whale itself, an environmental effect, or some combination of the two. The purpose of this analysis was not to explain all call segmentation but to determine whether there are any cases in which interference effects [Urick, 1983 (Chap. 5)] cannot explain the observed segmentation. To this end, interference patterns for a variety of source/receiver geometries were modeled, with contributions from both surface and bottom reflection paths considered.

For the surface reflection case these patterns were calculated by summing the pressure equations for the direct and surface-reflected paths of a spherical wave for each source/receiver geometry. For each source depth, the frequencies of the pressure minima were then identified and superimposed at that depth onto a plot of depth versus frequency (producing the dotted lines seen in Fig. 3). These depth/frequency plots were generated for a variety of ranges from 25–300 m. The nodal pattern of every segmented call was then visually compared to each of these plots to determine the possible depths/ranges of the vocalizing whale (Fig. 3). Segmented calls in which no reasonable[2] source positions/movements could be found to explain all of their segmentation were flagged.

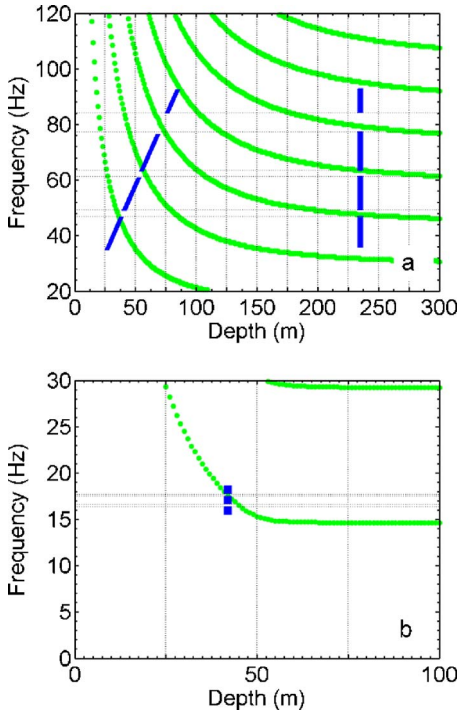


FIG. 3. (Color online) Comparison of pressure minima (dotted lines) versus frequency and depth to nodal patterns (time collapsed representation shown with solid boxes) of the calls from Fig. 2. (a) This example calculates the pressure minima at a range of 100 m. If the source depth is held constant, the nodal pattern of the audible downsweep call from Fig. 2(a) lines up with the pressure minima lines at ~230 m depth or greater, physically impossible since the seafloor depth was 100–140 m in this area. Assuming a variable source depth, the nodal pattern fits at a more reasonable source depth (25–90 m); however, this depth change over the 2 s length of call would require the whale to move at a speed of ~30 m/s—almost three times the fastest speed reported for a blue whale (Gambell, 1979). All other source ranges would have to be examined in this way (as well as cases of the whale moving in both range and depth) before surface interference effects can be ruled out. (b) This example calculates the pressure minima at a range of 5 m. Even at this extremely close range, the two pressure minima curves are still too far apart to cause the two nodes in the infrasonic downsweep call of Fig. 2(b), so at least one call node was produced by the whale itself.

A comparison of the impact of sea surface waves to call segmentation was not done for two reasons. First, although surface interference effects require a flat surface, this flatness is relative to acoustic wavelength. This flatness assumption is met for the infrasounds with their 80 m wavelengths, and during most recording sessions this assumption is also met for the audible calls with their 30 m wavelength. Second, the point of this analysis was to determine if any of the segmentation seen in the calls was created by the whales themselves. Falsely assuming a flatness condition overestimates the number of calls that can be explained by surface interference effects, therefore giving a conservative estimate of the number of calls with whale-generated segmentation.

The same general analysis techniques were used to examine the effect of bottom reflections on call segmentation. A perfectly reflecting seafloor was assumed, which produces the strongest potential interference effect. In reality, the reflection coefficient of the seafloor is most likely less than 1: results from the light bulb implosion measurements show bottom-reflected waveforms with amplitudes reduced to approximately 15% of the direct path. However, because reflection coefficients are frequency dependent, and the light bulb implosion spectra do not contain frequencies below 30 Hz, these results cannot be applied to the infrasonic calls. In any case, the perfectly reflecting boundary assumption gives a conservative estimate of the number of calls that cannot be explained by interference effects.

Because the audible calls analyzed for segmentation came from a specific field observation, the range used in the calculations could be safely limited to 50–400 m. The water column depth (necessary for the bottom-reflection calculations) was set at 105 m. This value was determined through travel time difference measurements of three light bulb implosions made after the encounter and is in agreement with the bathymetrical charts from that area. Specific source locations could not be determined for the infrasounds so the segmentation analysis for this call type was extended out to a maximum range of two kilometers (the detection range in the study area was limited by high ambient noise and poor propagation conditions). An examination of the position of all recordings containing segmented infrasounds led to the decision to use a water column depth of 300 m for this analysis.

For multinodal calls in both the surface and bottom reflection cases, it was too complicated to find solutions to the nonstationary whale cases through visual inspection, so an automated program was written to iteratively search for solutions. The process involved calculating pressure minima with the same method used for the visual comparison analysis, except range versus frequency plots were created every 0.1 m from the surface to the seafloor depth. All plots were then curve fit and range values were interpolated for the minimum and maximum frequency of each node. The maximum possible distance the whale could travel was then computed based on the time difference between nodes and a reasonable[2] swim speed for the whale. Any range/depth combinations that resulted in a distance greater than this maximum were rejected. In addition, any surface interference pattern solutions that required unreasonable[2] source depths or ranges were also rejected.

III. RESULTS

A. Infrasonic calls

Infrasonic calls are typically low in frequency (<20 Hz) and long in duration (>5 s). They can be found singly or arranged into regularly repeating patterns. Call segmentation was analyzed for most call types.

1. Infrasonic call types

St. Lawrence blue whale infrasounds are divided into four specific types: monotonic (n=433), downsweep (n=113), hybrid (n=22), and other (n=151), where n is the total number of calls of Rank 1–4 detected. The calls found in the first three categories have been attributed to blue whales in the literature. The last category contains calls that have not been previously reported and also do not have visual confirmation as to their source. This category is included to show the variety of sounds recorded that have some similarities to known blue whale infrasounds.

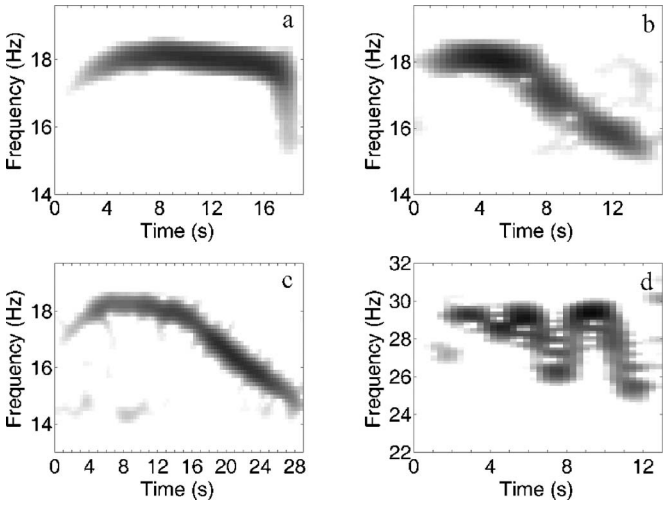


FIG. 4. Spectrograms of St. Lawrence infrasonic blue whale sounds. (a) Monotonic infrasound. (b) Downsweep. (c) Hybrid (note the longer duration: this call type is actually a monotonic infrasound joined to a downsweep). (d) Other (one typical example). All spectrograms were generated with a 512-point FFT, 95% overlap, and Hanning window with 499 points zero padding.

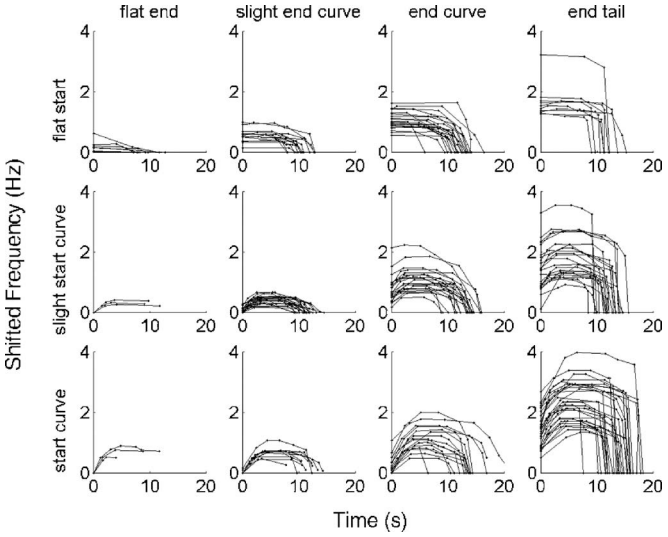


FIG. 5. Trace lines of infrasonic monotones, showing the variation between subtypes. Calls are divided into subtypes based on the shape of their leading and trailing edges. Four discrete curve shapes are used: flat, slight curve, curve, and tail (a sharp dropoff in frequency seen only at the end of the calls [see Fig. 4(a) for a spectrogram)]. To allow an easier comparison of call shapes, trace lines are shifted so that the minimum frequency of each call is zero.

a. Monotonic. The most common type of infrasonic sound was the monotonic call, comprising 76% of all infrasounds detected that can be attributable to blue whales [see Fig. 4(a) for one example of a monotonic infrasound and Table I for a summary of the quantitative parameters of all calls of this type]. These calls are similar to the North Atlantic part A phrase unit described by Mellinger and Clark (2003). They occur with a mean peak frequency of 18.1±0.4 Hz and a duration of 13.8±2.3 s.

Although these calls are labeled "monotonic," there is a small change in frequency with bandwidths ranging between 0.8 and 4.6 Hz (mean 2.0±0.8 Hz), with most of this frequency modulation occurring at the leading and trailing edges of the call. When the shapes of these edges are compared between calls (Fig. 5), it becomes apparent that the monotonic call type is anything but uniform. Also, it should be noted that although discrete edge shapes (flat, slight curve, curve, and tail) are used in Fig. 5, the similarity between adjacent categories shows that call variations extend across a continuum.

b. Downsweep. Infrasonic downsweep calls [Fig. 4(b)] occurred less frequently than the monotonic type, making up just 20% of all known blue whale infrasounds detected (see Table I for a summary of the quantitative parameters of this call type). These calls are similar to the part B phrase units of the North Atlantic (Mellinger and Clark, 2003). Compared to the monotonic calls, infrasonic downsweeps are slightly shorter in mean duration (12.5±2.3 s) with a lower mean peak frequency of 17.0±0.9 Hz but wider mean bandwidth of 3.3±0.5 Hz. As the name implies, all calls in this category sweep downward in frequency.

Differences between the calls are seen primarily in the initial 3–5 s of the calls (roughly the first third); all calls finish with a long downswept section of approximately 10 s. Some calls [Fig. 6(a)] appear to be missing the initial segment and are approximated by a single line of slope −0.3 Hz/s. Others [Fig. 6(b)] begin with an initial segment with zero slope, followed by a long linear segment of slope

TABLE I. Quantitative parameters for infrasonic calls. Mean ± s.d., range, and median values given for each parameter. The variable *m* indicates the number of calls of Rank 1–3 used in these statistics.

			Frequency (Hz)			Received level (dB)		
Call type	Duration (s)	Bandwidth (Hz)	Minimum	Maximum	Peak	Avg. power (*re* 1 μPa^2)	Total energy (*re* 1 μPa^2/Hz)	Max, rms Power (re 1 μPa^2)
	13.8±2.3	2.0±0.8	16.6±0.8	18.6±0.3	18.1±0.4	108	120	110
Monotonic	4.6–19.1	0.8–4.6	14.3–20.0	16.1–21.6	15.6–21.1	72–136	81–148	75–138
m=187	13.9	1.8	16.8	18.6	18.0	103	115	106
	12.5±2.3	3.3±0.5	15.0±0.5	18.3±0.7	17.0±0.9	102	113	106
Downsweep	5.8–16.8	2.2–4.3	12.7–16.0	16.3–19.0	14.0–19.6	80–109	91–120	84–114
m=62	12.9	3.4	15.0	18.4	17.3	101	111	105
	22.8±3.4	3.7±0.5	15.0±0.4	18.7±0.2	17.6±0.7	99	113	104
Hybrid	17.3–28.2	3.2–4.4	14.2–15.5	18.4–18.8	16.4–18.3	73–106	85–120	78–111
m=8	22.2	3.5	15.0	18.7	17.9	98	112	102

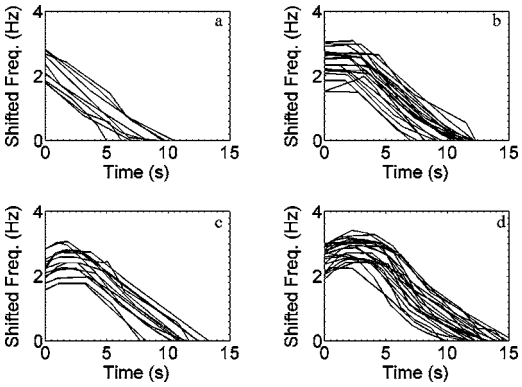


FIG. 6. Trace lines of infrasonic downsweep calls, showing the variation between subtypes. Calls are divided into subtypes based on the shape of their leading and trailing edges. Four shapes are used: (a) straight leading and trailing, (b) flat leading, straight trailing, (c) arched leading, straight trailing, and (d) arched leading, curved trailing. To allow an easier comparison of call shapes, the trace lines are shifted so that the minimum frequency of each call is zero.

−0.3 Hz/s. The rest of the infrasonic downsweep calls begin with a gently arching first segment. This is followed by either a linear downsweep of slope −0.3 Hz/s [Fig. 6(c)] or a section that begins with slope −0.4 Hz/s but decreases to −0.2 Hz/s by the end of the call [Fig. 6(d)].

c. Hybrid. Hybrid calls [Fig. 4(c)] were the least common of all infrasounds, occurring just 4% of the time. A hybrid call is actually a call phrase (see the "Infrasonic call spacing and sequencing" section) consisting of a monotonic call followed by a downsweep with no time interval between. Because it is difficult to tell where this crossover occurs, quantitative call statistics cannot be accurately calculated for each phrase component. For this reason, the phrase components are measured together as a single hybrid call. The mean duration of hybrid calls is 22.8±3.4 s with a mean peak frequency of 17.6±0.7 Hz and a bandwidth of 3.7±0.5 Hz (see Table I for hybrid call quantitative parameters).

d. Other. The "other" category [an example of which is shown in Fig. 4(d)] contains a variety of sound types. None have direct field evidence linking them to blue whales other than that they were detected in recordings made when blue whales were present. They are included here as questionable sounds to promote discussion of whether their source is a blue whale, another whale species, another biological source, or noise. No call parameter statistics are presented in Table I for these calls due to their great variability. Instead, spectrograms of representative calls from each type will be shown in the context of two-minute spectrograms. The first type is the wiggle (n=43, Ranks 1–3), which varies between being highly convoluted [Fig. 7(a)] to slightly kinked [Fig. 7(b)]. These calls had mean durations of 10±4 s (range 5–22 s), with a mean peak frequency of 23±7 Hz and a bandwidth of 4±2 Hz. The second questionable call type (n=19, Ranks 1–3) includes calls that are downswept with shorter durations (mean: 7±2 s, range: 4–11 s) and higher frequencies (mean peak: 22±4 Hz) than infrasonic downsweep calls. They are also longer in duration and lower in frequency than the audible downsweep calls described below. Examples of these questionable downsweep calls (indicated by upper left and center arrows) can be seen along with an infrasonic monotonic/downsweep pair (lower two arrows) in Fig. 7(c). The third questionable call type (n=17, Ranks 1–3), shown by the upper rightmost arrow in Fig. 7(c), is short in duration (mean: 4.5±1 s, range: 3–6 s), narrow in bandwidth (mean: 2.0±0.4 Hz), and has peak frequencies similar to the other questionable call types (mean: 19±5 Hz).

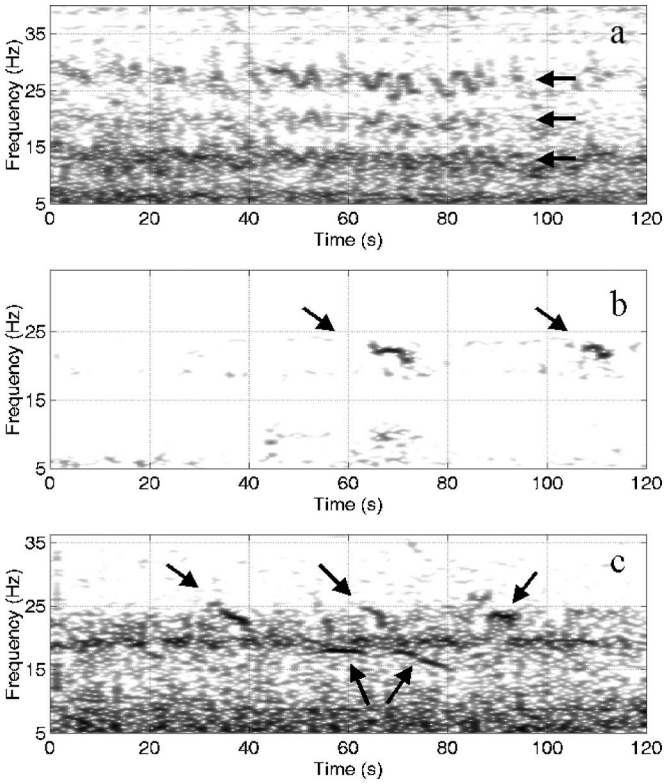
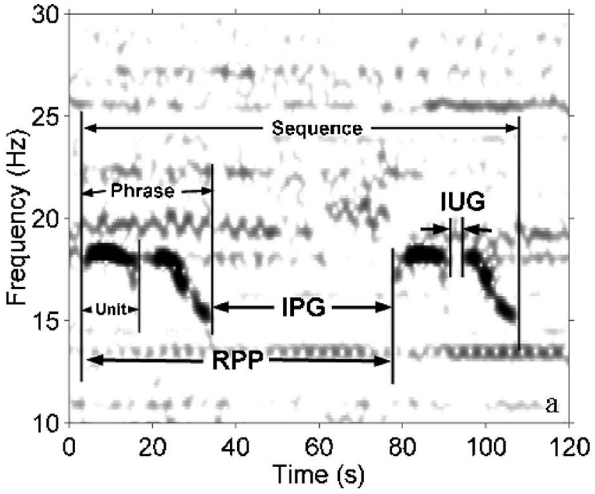


FIG. 7. Some examples of possible blue whale infrasounds. Although blue whales were present during these recordings, no field evidence can directly identify them as the source of these calls. (a) Highly convoluted wiggle call (along frequency lines marked by arrows). (b) Slightly kinked wiggle call (arrows). (c) Questionable downsweep (upper left and center arrows) and short (upper right arrow) calls, shown with an infrasonic monotonic/downsweep phrase (lower two arrows). All spectrograms generated with a 1024-pt FFT, 85% overlap, and a Hanning window with 499 points zero padding.

e. A note on 9 Hz sounds. All segments containing infrasonic downsweep calls were rescanned for the presence of the 9 Hz sound described by Mellinger and Clark (2003). Most of these segments had high levels of ambient noise around 9 Hz, so the prevalence of occurrence for this call type is unknown. One definite and 16 questionable sounds similar to the 9 Hz sound were detected in total.

2. Infrasonic call spacing and sequencing

The terminology used to describe infrasonic call spacing and sequencing follows roughly from that of Mellinger and Clark (2003) and is illustrated in Fig. 8. *Units* are individual calls,[3] separated by interunit gaps and grouped into phrases. *Phrases* are separated by interphrase gaps and grouped into sequences. These *sequences* are separated by intersequence gaps and grouped into series. *Interunit gaps* (IUG) are measured from the end of one unit to the beginning of the next unit. *Interphrase* (IPG) and *intersequence* gaps (ISG) are measured from the end of the last unit of one phrase or sequence to the beginning of the first unit of the next phrase or sequence. The difference in length between these three intercall interval types can be seen in the histograms of Fig.



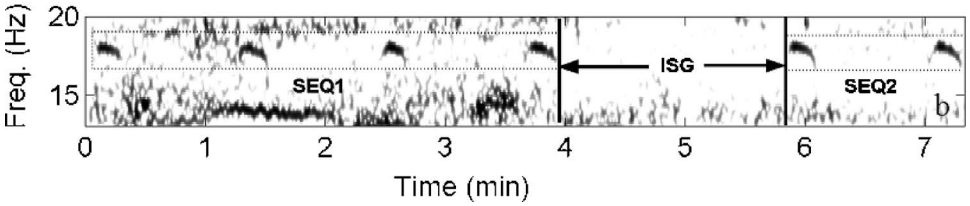


FIG. 8. An illustration of call and interval terms for infrasonic calls. (a) A sequence of two double unit phrases. (b) A series of two sequences (SEQ1 and SEQ2) of 4 and 2 single unit phrases, respectively. IUG: interunit gap, IPG: interphrase gap, RPP: regular phrase period, ISG: intersequence gap.

9. IUGs form the first peak at 4 s, while IPGs appear as the second peak at approximately 58 s [Fig. 9(a)]. ISGs are far more variable but do show a main peak at around 3 min and a slight second peak near 10–12 min [Fig. 9(b)]. These ISG peaks coincide well with the typical surfacing and dive times, respectively, for St. Lawrence blue whales (Sears and Calambokidis, 2002).

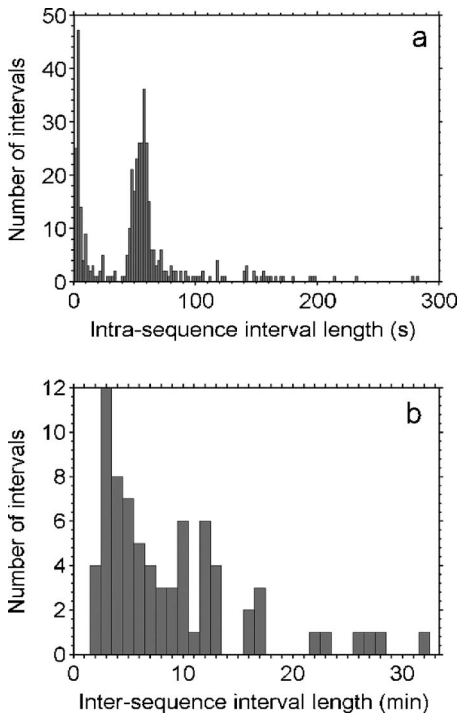


FIG. 9. Histograms of intercall intervals between all infrasonic calls. (a) The intrasequence interval shows peaks at both 4 and 58 s that represent interunit gaps (IUG) and interphrase gaps (IPG), respectively. (b) The intersequence interval shows a main peak at 3 min and a slight second peak near 10–12 min.

Because it is not known whether blue whales cue in to the actual interval space between phrases (IPG) or the timing from the start of one phrase to the next (the regular phrase period), the latter is also reported. The *regular phrase period* (RPP) is measured from the start of the first unit of a phrase to the start of the first unit of the next phrase (Mellinger and Clark, 2003). Results for the RPP are broken into phrase types, since phrase length contributes to the RPP length. The RPP was found to be 76.3±8.8 s (n=75) for all AB phrases, 72.9±8.5 s (n=8) for all AA phrases, and 72.3±9.1 s (n=159) for all single-A phrases. Additional measurements were made of the RPP for the subset of AB phrases followed by an AB phrase (79.4±5.4 s, n=32) and for the subset of single-A phrases followed by a single-A phrase (71.5±10.3 s, n=120) to facilitate comparison with other studies.

The basic structure of infrasonic call sequences is revealed when interval lengths between specific call types are measured. The A-A interval shows a slightly bimodal distribution [Fig. 10(a)] with the greatest peak at 60 s. This peak represents the IPG where each phrase is made up of a single-A call. The smaller peak below 20 s comes from IUGs where the phrase is composed of a pair of A calls. In contrast, the A-B interval [Fig. 10(b)] is almost exclusively an IUG, with a main peak at 3 s (the additional peak at 0 s comes from the 17% of all AB phrases that are of the hybrid call type). The single peak at 50 s for the B-A interval (histogram not shown) indicates that this is predominantly an IPG.

A total of 458 phrases were found in the recordings. These phrases were primarily composed of single-A units (67%) and AB (including the hybrid call) pairs (23%). The rest consisted of single-B and AA pairs (about 5% each) or

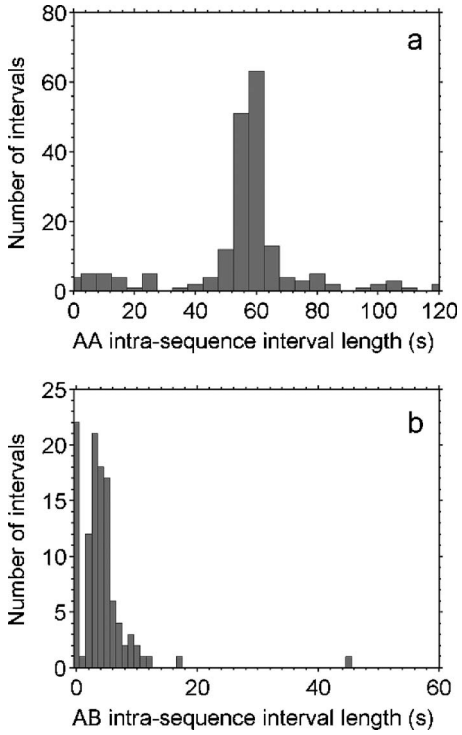


FIG. 10. Histograms of intrasequence intervals between specific types of infrasonic calls. (a) The histogram of AA intervals shows a main peak at 60 s that represents an inter-phrase gap where each phrase is composed of a single A unit. The interunit gap for AA phrases is shown by the smaller peak below 20 s. (b) The histogram of AB intervals shows a main peak at 3 s that represents an interunit gap (the peak at 0 s represents the interunit gap of those AB phrases composed of hybrid calls).

AAB and ABA groupings (less than 1% each). These phrases combined to form 157 sequences. The majority of sequences (77%) were considered to be complete (i.e., there was at least 100 s, the maximum RPP length, on either side of the sequence). Of those sequences that were complete, over 50% were made up of a single phrase, as shown in Fig. 11. The maximum number of phrases found in a sequence was 12. Complete multiphrase sequences were either composed entirely of single-A phrases (43%) or a combination of single-A, AA, AB, and hybrid call phrases (56%). Although patterning of phrases was seen in some of the sequences, most did not continue the same patterning for the entire length of the sequence. However, it is possible that the repeating pattern is longer than most of the call sequences detected. For example, one incomplete sequence showed strong evidence that the repeating pattern was eight phrases long.[4]

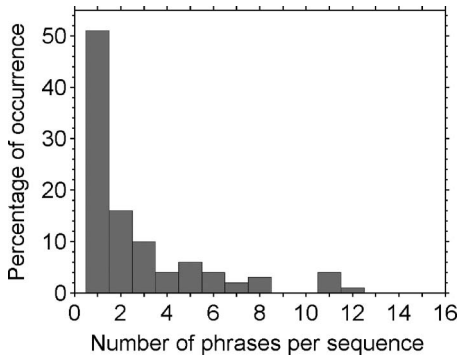


FIG. 11. Histogram showing the phrase composition of sequences. Sequences varied from 1 to 12 phrases in length. Over 50% of all sequences are composed of a single phrase.

The most common initial phrase in multi-phrase sequences with complete beginnings ($n=61$) is the single-A (69%), followed by the AB and hybrid phrases (11% each), and the single-B (7%). The single-A is also the most common end phrase (71% are single-A; 16% are AB) in multiphrase sequences with complete endings ($n=77$). Since the single-A and AB phrases are the most common of all phrase types, it is not surprising that they are also the most common starting and ending phrases. However, although hybrid calls are not common, 55% were found to occur at the start of a sequence, versus 30% in the middle, and 15% at the end.

Because very few recording sessions were of sufficient length to capture both the start and end of a series, all series ($n=72$) are included in the dataset. The series seem to be composed predominantly of single phrases (39%), with the longest series containing 46 phrases. The majority of series were made up of single sequences, with a few containing up to eight sequences. Sequences from the same series were more likely to be similar than those from different series, although very few consisted of the exact same phrase patterning. Of all series detected in this study, just six (8%) had a second whale vocalizing in the background.

3. Infrasonic call segmentation

Call segmentation was seen in 19% of monotonic, 53% of downsweep, and 38% of hybrid infrasounds.[5] Segmented calls had 1–2 nodes, with 80% containing single nodes.

For the comparison of the nodal patterns of these calls with the generated interference pressure fields, several possible whale movements were considered: stationary, changing either range or depth, and changing both range and depth. If a resulting source position/track seemed unreasonable based on what is known about blue whales from the literature or from field circumstances at the time of the call, it was not included in the set of calls determined to be caused by interference effects. Unreasonable source depths fell into two categories: those below 50 m, which is deeper than what has been reported for blue whales [10–40 m (Thode *et al.*, 2000), 10–23 m (Oleson *et al.*, 2003)], and those that were below the seafloor depth. For whale swim speeds, those greater than 5 m/s were considered unreasonable.[6] Unreasonable source range was determined on a case-by-case basis.

The same trends were seen for all three call types when surface-interference patterns were examined. For the single-node calls, with depth limited to the deepest point of the channel (300 m), no solutions were possible past a range of 200 m. Limiting the source depth to 100 m, this range shortened to 50–60 m. Even at the unreasonably close source range of 25 m, the depths needed to create the appropriate interference patterns were still greater (55–70 m) than those reported for vocalizing blue whales. For calls containing two nodes, no possible solution exists where the interference pattern could create both nodes of these calls. Because no source locations can produce interference patterns consistent

TABLE II. Quantitative parameters for audible calls. Mean ± s.d., range, and median values given for each parameter. The variable *m* indicates the number of calls of Rank 1–3 that are attributable to blue whales.

			Frequency (Hz)			Received level (dB)		
Call type	Duration (s)	Bandwidth (Hz)	Minimum	Maximum	Peak	Avg. power (*re* 1 μPa^2)	Total energy (*re* 1 μPa^2/Hz)	Max, rms Power (*re* 1 μPa^2)
	2.0±0.6	50.6±19.2	37.8±15.9	88.3±22.8	53.9±18.1	121	124	125
Downsweep	1–4	20–286	10–104	56–242	22–158	81–134	83–137	89–139
m=115	1.9	50.4	34.7	83.8	50.4	114	117	118
	0.8±0.2	16.7±11.9	57.3±12.0	74.0±13.2	65.1±10.9	103	102	104
Blurp	0.4–1.4	0.7–60.8	26–94	41–113	36–104	81–129	80–127	62–129
m=103	0.7	11.2	57.1	71.7	64	95	94	98
	1.0±0.3	130±84.4	46.0±30.0	176±79	78.7±43.9	109	110	112
Grunt	0.5–2.1	21–377	0.4–118	75–404	0.6–183	88–130	87–130	91–133
m=40	0.9	103	49.2	152	70.9	103	103	105
	3.2±1.9	133±89.4	25.1±12.1	158±82.6	42.9±24.2	108	113	113
Bubbling	1.1–7.2	55.8–332	5.6–44.5	83.8–341	12.4–88.7	82–113	83–118	94–119
m=13	2.5	95.1	25.2	122	41.6	107.8	113.9	111.4

with the multinodal segmentation, and no reasonable source locations can produce the single-node segmentation, the surface-reflection analysis suggests that the segmentation seen in infrasonic calls is generated by the whale itself.

Results of the bottom-reflection analysis were also consistent among the infrasonic call types. Limiting the source depth to 50 m or less [based on reported blue whale calling depths (Thode *et al.*, 2000; Oleson *et al.*, 2003)], most of the single-node calls showed segmentation consistent with the interference patterns. However, over 90% of the segmentation seen in the two-node calls (maximum swim speed: 5 m/s; seafloor depth: 300 m) could still not be explained with either the surface-reflection or the bottom-reflection interference pattern analysis.

B. Audible calls

Audible calls are higher in frequency (>20 Hz) and shorter in duration (<5 s) than infrasounds. They are found singly or in multicall groupings. Unlike the strongly patterned sequences of infrasonic calls, audible calls occur as more randomly clustered groupings called *bouts*. It should be noted that this term differs in meaning from that used by Mellinger and Clark (2003). Call segmentation was examined for only one call type.

1. Audible call types

Four types of audible calls were detected during this study: downsweep (n=233), blurp (n=440), grunt (n=161), and bubbling (n=13), where n is the total number of calls of Rank 1–4 detected. Each call type is described below, with all quantitative parameters summarized in Table II.

a. Downsweep. Similar to the infrasonic variety, audible downsweep calls [Fig. 12(a) and Table II] drop in frequency over the length of the call. However, audible downsweeps have shorter durations (mean=2.0±0.6 s), broader bandwidths (mean=51±19 Hz), and higher peak frequencies (mean=54±18 Hz) than the infrasonic downsweeps. They also show more variation between individual calls (Fig. 13).

One recording made in the presence of a surface-active trio of blue whales provided the opportunity to calculate source levels for a large number (n=34) of audible downsweep calls because the source of the calls could be identified and range estimated (~100 m). Transmission loss (dB) obtained through measurement of light bulb implosions (Heard *et al.*, 1997) was 16 log R (where R is the range in m). This agreed with results obtained with the Monterey-Miami parabolic equation (Smith and Tappert, 2003) for a sandy shelf area[7] of depth 110 m. Using this transmission loss estimate, source levels were 142–166 dB *re* 1 μPa with the majority between 156–166 dB *re* 1 μPa over the bandwidth of the call.

b. Blurp. The onomatopoeic term "blurp" is used to describe this audible call type [Figs. 12(b) and 14(b)]. Blurp

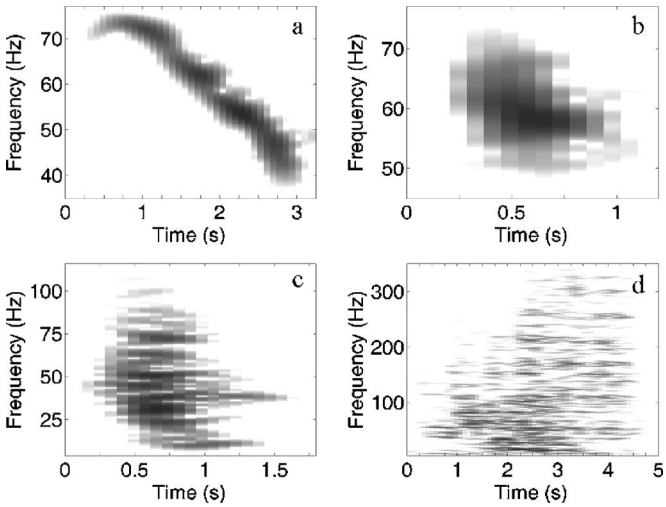


FIG. 12. Spectrograms of audible sounds made by St. Lawrence blue whales: (a) Audible downsweep, (b) blurp, (c) grunt, and (d) bubbling. All spectrograms were generated with a 1024-point FFT, 97.7% overlap, and a Hanning window with 4096 points zero padding.

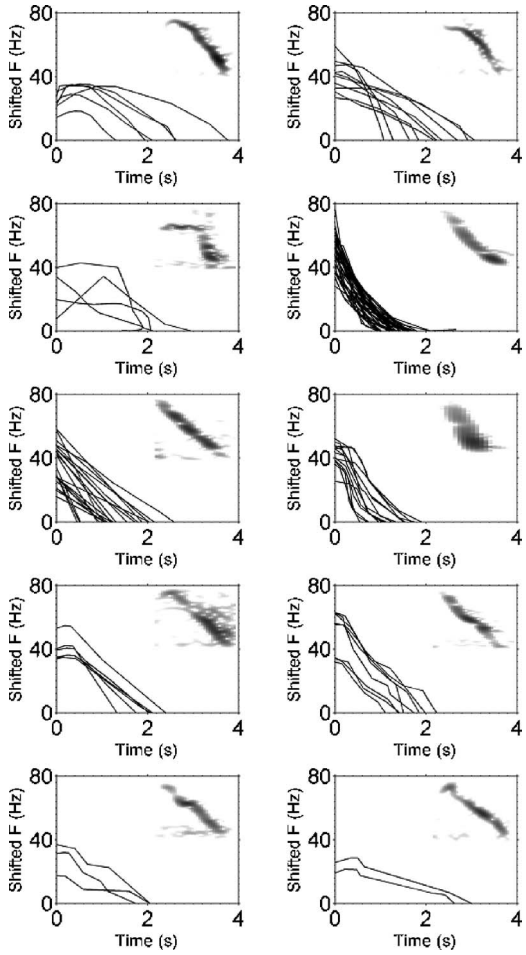


FIG. 13. Variation seen between audible downsweep call subtypes. Each plot has a superposition of trace lines plus a representative spectrogram for that call type. Call categories are (clockwise from top left) as follows: convex high arch, convex arc, concave arc, "Z," "Z" with a dropped end, tight hook, step, slight hook, straight, and other. To allow an easier comparison of call shapes, trace lines are shifted so that the minimum frequency of each call is zero. Spectrograms are not to scale.

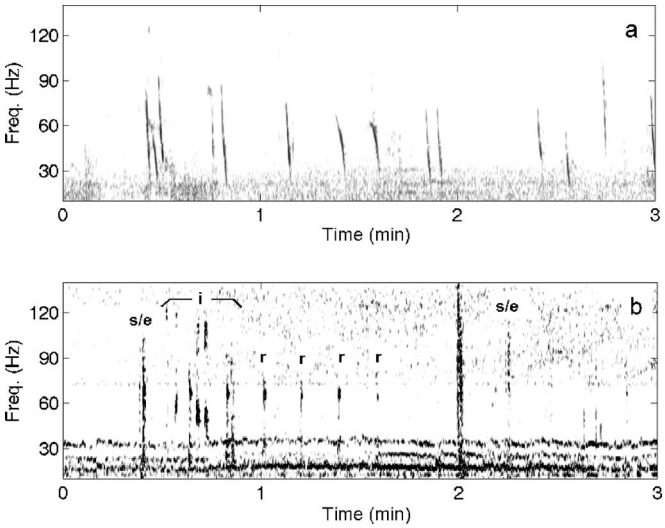


FIG. 14. In contrast to the highly structured groupings of the infrasounds, all audible call types occur in less organized groupings called *bouts*. (a) Audible downsweep calls are grouped into irregularly spaced bouts. (b) Audible blurp and grunt calls can occur together in *mixed-pattern bouts* containing both regular and irregular spacings. Letters mark spacing type: r—regularly spaced calls; i—irregularly spaced calls; s/e—start/end calls. All spectrograms generated with a 1024-pt FFT, 85% overlap, and a Hanning window with 499 points zero padding.

calls seem to be abbreviated versions of the audible downsweep call: they are shorter in duration (mean 0.8±0.2 s) and fall into the low- to mid-frequency range (mean bandwidth =17±12 Hz, mean peak frequency=65±11 Hz) of the downsweeps. Because some blurps were detected under high signal-to-noise conditions, it is unlikely they are simply downsweeps obscured by noise.

Although 239 blurp calls were of sufficient quality (Ranks 1–3) to obtain quantitative measurements, only those that could be attributed to blue whales were included in the statistics listed in Table II. This subset includes blurps found in mixed-pattern bouts occurring with a blue whale at the surface, blurps found in close association with audible downsweep calls, and blurps produced by blue whales closely approaching the research boat.

About 85% of this subset of calls were found in mixed-pattern bouts coinciding with a blue whale surfacing sequence. As described in the "Audible call spacing and sequencing" section, mixed-pattern bouts [see Fig. 14(b) for an example] typically contain some combination of regularly spaced blurp calls, irregularly spaced blurp and grunt calls, and beginning/ending grunts. Although much variation is seen among mixed-pattern bouts, they are distinctive enough to be picked out of a spectrogram. A total of 33 mixed-pattern bouts[8] were detected during this study. A comparison with field notes found that 25 of these mixed-pattern bouts occurred during a blue whale surfacing sequence (18 were blue whale pairs, six were single blue whales, and one was a blue/fin whale pair). Three bouts were detected during the surfacing sequences of unidentified (either blue or fin) whales, and five bouts occurred without any apparent whale surfacings. It does not appear that any calls made during these bouts coincide with exhalations of the whales. Thirteen blurp calls were detected in association with audible downsweep calls, the majority occurring during the surface-active trio interaction described for the audible downsweep call type. The remaining blurp calls used for the statistics were detected after dark when a blue whale approached the research boat to within 10–15 m.

c. Grunt. Qualitatively, grunt calls [Figs. 12(c) and 14(b)] generally sound more forceful and raspy than blurps. Quantitatively, they are broader in bandwidth (mean 130±84 Hz). They are short in duration (mean 1.0±0.3 s) and have a mean peak frequency of 79±44 Hz.

As was done for the blurp calls, only those grunt calls that were of sufficient quality (Ranks 1–3) and that could be attributable to blue whales were used in the statistics shown in Table II. Of the 40 grunt calls meeting these requirements, 90% were found in mixed-pattern bouts detected during blue whale surfacing sequences. Three grunt calls were detected in association with audible downsweep calls, and one grunt was detected after a blue whale made a U-turn and logged 10 m behind the research boat.

d. Bubbling. Although bubbling [Fig. 12(d) and Table II] is listed here as an audible call type, it is likely a non-voiced sound made as the whale expels bubbles through the blowhole. It is not known whether this sound is intentional, so it is included for completeness. Bubbling is heard during bouts of audible downsweeps and only in close proximity to blue whales.

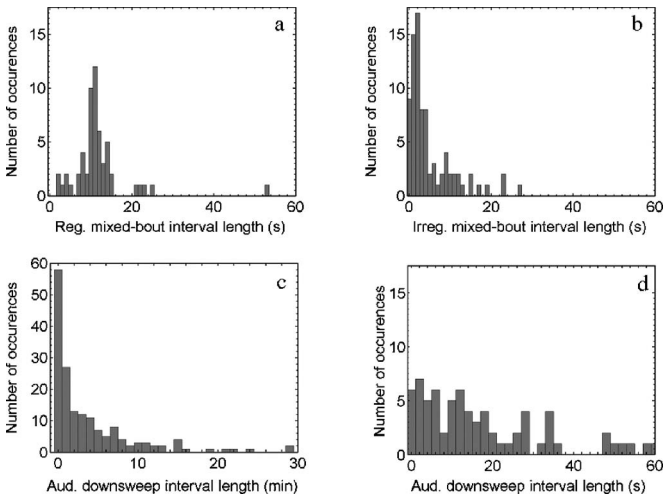


FIG. 15. Audible interval lengths. The distribution of call interval lengths in mixed-pattern bouts show: (a) intercall intervals within the regularly spaced sections strongly peaking at 11 s; and (b) calls within the irregularly spaced sections more closely spaced with interval lengths peaking at 2 s. Intercall interval lengths within bouts of audible downsweep calls have (c) a much broader distribution (note the x-axis scale is in minutes) than those within mixed-pattern bouts, and (d) no strong peaks in their distribution below 1 min, as was seen for the mixed-pattern bouts.

2. Audible call spacing and sequencing

As mentioned previously, audible calls do not occur with the unit/phrase/sequence/series hierarchy of the infrasounds. Therefore, the terminology used to describe audible call spacing and sequencing is generic: calls are separated by *intercall intervals* (measured from the end of one call to the start of the next) and grouped into bouts. As with the infrasounds, only calls that could be attributed to blue whales (either by their similarity to calls reported in the literature or by their close association with a specific blue whale behavior in the field) were used in this analysis.

The type of spacing seen in the bouts depended on the type of audible call involved. Blurps and grunts were primarily found in *mixed-pattern bouts*. The structure of this mixed-pattern bout can be seen in Fig. 14(b). There was typically some combination of three components in this pattern: regularly spaced call sections, irregularly clumped call sections, and beginning/ending calls. The size and order of these sections varied between occurrences. Blurp calls are more evenly distributed between sections, with 55% in regularly spaced and 45% in irregularly spaced sections. Grunts tend to occur more frequently in the irregularly spaced sections (65% vs 35% in regular spacings). As seen in Figs. 15(a) and 15(b), call intervals in the regularly spaced sections have a strong peak at 11 s, while those in the irregularly spaced sections peak at around 1–2 s.

Audible downsweep calls showed the most variation in interval length [Fig. 15(c)]. Although 40% of all intervals between adjacent downsweep calls were less than a minute long, no sharp peak can be seen on a finer-scale histogram of interval length [Fig. 15(d)]. Twenty-four percent of recording sessions with audible downsweep calls had only one call detection. These solitary cases produce partial intervals bordered by only one call and so were left out of the analysis, but their presence further supports the random spacing of this call type.

Occasionally, audible blurps, grunts, and bubbling were detected in close association with audible downsweep calls. The spacing among these four call types most closely resembled the irregularly spaced portion of the audible mixed-pattern bout type and had a peak value of 2 s.

3. Audible call segmentation

Audible downsweep calls have the highest percentage of call segmentation (80%) of any call type in this study. Of the 115 calls of Rank 1–3, 24 had 0 nodes, 21 had 1 node, and 70 had 2 or more nodes. A subset of these audible downsweep calls, which could be attributed to a trio of surface-active blues at ~100 m range, was used for this analysis due to the certainty of the source location. As mentioned previously, measurements made from light bulb implosions determined that the seafloor depth was 105 m at the location of the trio. A portion of the calling bout from this trio can be seen in Fig. 14(a). Of the 34 calls in this subset, 5 had 1 node and 23 had 2 or more nodes.

Again, both stationary and moving whale cases were considered when comparing the nodal structure of the calls and the interference patterns. A swim speed of 10 m/s [the maximum reported for a blue whale (Gambell, 1979)] was used because of the speed swimming observed during the interaction.

For the surface-reflection analysis, limiting source range to greater than 75 m (since the whales were estimated to be at around 100 m) and source depth to less than 150 m, the percentage of calls, whose nodal pattern could be explained by surface interference effects was 100% for single node calls, but only 48% for multinodal calls. When the source depth was restricted to 50 m or less (corresponding to reported blue whale calling depths) and the source range constrained to 75 m or more, all calls with single nodes could still be explained by surface interference effects. Under the same depth and range limitations, however, surface interference effects could explain only 17% of the multinodal calls. These results indicate that while reasonable source ranges and depths can produce interference patterns consistent with the single node segmentation, these interference patterns cannot explain all of the multinodal segmentation.

For the bottom-reflection analysis, the range was limited to 75–150 m, the whale calling depth to 50 m or less, and the seafloor depth was set to 105 m. As with the surface-reflection case, 100% of the single node calls had segmentation consistent with bottom-reflection interference. Although bottom reflections could explain the segmentation of more multinodal calls than the surface reflections, 31% of the multinodal calls detected during this interaction could not be explained by either bottom or surface interference effects. For this reason it appears that at least some of the segmentation seen in audible calls is whale generated.

IV. DISCUSSION

The recordings made during this study have shown that the St. Lawrence blue whale call repertoire is far more varied than previously thought. Not only were many different call types in both infrasonic and audible categories detected, a wide range of characteristics were seen within each call type.

A. Infrasonic calls

Because worldwide geographic differences in infrasonic blue whale vocalizations have been described in detail elsewhere (Rivers, 1997; Mellinger and Clark, 2003; McDonald *et al.*, in press), this discussion will focus on a comparison of blue whale sounds within the North Atlantic.

In addition to the St. Lawrence study by Edds (1982), two North Atlantic studies were used for this comparison. The first characterizes vocalizations recorded from 1992–1994 on U.S. Navy bottom-mounted arrays located in the western North Atlantic between 15–65° N latitude and the eastern North Atlantic between 45–65° N latitude (Mellinger and Clark, 2003); the second describes vocalizations detected from 1999–2001 on six moored hydrophones located along the Mid-Atlantic ridge between 15–35° N latitude and 33–50° W longitude (Nieukirk *et al.*, 2004). Several similarities and differences were found when results from these three studies were compared to results from this study.

First, the presence of the infrasonic downsweep call in the St. Lawrence results in a repertoire as varied as that of the North Atlantic. However, although this B call is present, the composition of St. Lawrence call phrases seems to run counter to what was found in the North Atlantic by Mellinger and Clark (2003): while they found the breakdown of phrases to be 66% AB and 34% single A, this study found the phrase composition to be 67% single A and 23% AB. The possibility that this is a coastal/open ocean difference is contradicted by the similarity of the St. Lawrence phrase composition to that of the Mid-Atlantic ridge, which has 71% single A and 29% AB phrases (Nieukirk *et al.*, 2004). If phrase type varies between North Atlantic blue whale populations, then it is possible that St. Lawrence whales migrate to the Mid-Atlantic ridge (where the most calls were detected in December and January). However, this Mid-Atlantic ridge area was also covered by the arrays used by Mellinger and Clark (2003), so unless their detections were biased toward other areas, this argument fails. On the other hand, the years of this study (1998–2001) overlap those of Nieukirk *et al.* (1999–2001), while neither occurred while the Mellinger and Clark recordings (1992–1994) were collected. Although it is possible that the composition of call phrases changes over time, it seems unlikely that a complete reversal in phrase patterning could have occurred within seven years. More data are needed to determine the reason behind these differences in call phrasing.

It appears that the interphrase spacing seen in this study may also be more similar to that reported by Nieukirk *et al.* (2004) than that of Mellinger and Clark (2003). A direct comparison between these two studies cannot be made because both reported interphrase spacing for only their predominant phrase type. Mellinger and Clark (2003) report a RPP of 73.3 ± 5.0 s ($n=4025$) for sequences made up of AB phrases, while the RPP from Nieukirk *et al.* (2004) is 71.8 s ($n=556$) for sequences made up of single-A phrases. The RPPs computed in this study are 79.4 ± 5.4 s ($n=32$) for intervals between AB phrases and 71.5 ± 10.3 s ($n=120$) for intervals between single-A phrases. The dataset from Edds (1982) gives a RPP of 69.5 ± 1.9 s ($n=4$) for one sequence made up of single-A phrases.

A comparison of frequency characteristics between studies reveals that St. Lawrence infrasounds have a much wider bandwidth than those from the North Atlantic. The monotonic call type (A) in this study started at 18.6 Hz, sweeping 2 Hz. Mellinger and Clark (2003) found this call type to start at 18.5 Hz, sweeping only 0.2 Hz. The data from Nieukirk *et al.* (2004) fell in between with 0.8 Hz of sweep (starting frequency=18.4 Hz). This latter dataset is most similar to the bandwidth of 0.9 Hz reported by Edds (1982), although her starting frequency (19.4 Hz) was much higher than in any of the more recent studies. The infrasonic downsweep call type (B) showed less variation between studies, with starting frequencies of 18.3, 18.5, and 18.3 Hz, and sweeps of 3.3, 2.8, and 2.3 Hz for this study, the North Atlantic (Mellinger and Clark, 2003), and the Mid-Atlantic ridge (Nieukirk *et al.*, 2004), respectively.

It seems unlikely that these frequency differences reflect geographical dialects, since the St. Lawrence whales migrate into the Atlantic, and this broad bandwidth does not "migrate" as well. Possibly, high local ambient noise levels[9] may force the St. Lawrence whales to use a broader frequency band so that their signals have a better chance of being detected. Also, differences between areas may be an artifact of sensor placement. The monotonic call type, which has most of its bandwidth in tails at the beginning and end of the call, shows the biggest difference among areas. These tails are lower in amplitude than the center of the calls, so they might not appear in long-range call detections, reducing the apparent bandwidth of the recorded calls. Calls detected in this study likely came from whales much closer to the sensor than in the other studies, especially with the deeply moored hydrophones in the Atlantic.

Some calls resembling those described in the "other" infrasonic call category have been reported previously. The "highly convoluted wiggle" call shows some similarity to the "rumble" call recorded in the St. Lawrence by Edds (1988, Fig. 6), which she found to occur in the presence of closely interacting fin whales. Watkins (1981) also attributed a "low-frequency rumble" call to fin whales. It is possible this call is produced only by fin whales, but given the frequency range of the call, the similarities between blue and fin whales, and the fact that no fin whales appeared to be in the immediate vicinity of the recordings, this call type might also be produced by blue whales. The "four-second chirp" sound detected off western Australia by McCauley *et al.* (2001, Fig. 36) is also similar to the "highly convoluted wiggle" call. Although not seen on the day of their recording, blue whales were sighted on the previous day, while fin whales were not mentioned at all. The "slightly kinked wiggle" seems to resemble those described by Stafford *et al.* (2001) for the Northwest Pacific. This suggests that St. Lawrence blue

whales may also be capable of making this type of sound, and so it should not be immediately dismissed. In addition, the "9 Hz sound" in the Atlantic (Mellinger and Clark, 2003) and the precursor call in the Pacific (Aburto *et al.*, 1997; McDonald *et al.*, 2001) show that blue whales are capable of making very low-frequency calls.

The results from the interference pattern analysis show that while it is unlikely that surface interference effects contributed to the single-node calls, most can be explained by interference patterns created by bottom reflections. It is possible that the amplitude of a bottom-reflection waveform is insufficient to generate a strong node in the call, but no bottom-interaction information is available at infrasonic frequencies for this study area, and so this issue cannot be addressed. For the multinodal calls, only one call had segmentation consistent with an interference pattern, so it appears that the multinodal segmentation seen in infrasonic calls is whale generated.

B. Audible calls

A great variety of audible downsweep calls were detected in this study, many during social situations. Because of this social context, it is not surprising that no audible downsweeps were detected around the solitary animal recorded by Edds (1982). Results from the Mid-Atlantic Ridge include detections of the bottom halves (the recording system had an upper limit of 50 Hz) of arch and other potential downsweep calls (Nieukirk *et al.*, 2004). Audible downsweep calls were also detected in the North Atlantic: although Mellinger and Clark (2003) report only arch call detections, many varieties of the audible downsweep call were observed during the infrasonic call processing of their dataset (Clark, 2003). Arch call detection in the St. Lawrence was limited ($n=6$) and was not associated with AB phrases as in the other Atlantic studies (Mellinger and Clark, 2003; Nieukirk *et al.*, 2004). Overall, the audible downsweep calls from this study fall within the same frequency band and time span as those previously reported worldwide (Thompson *et al.*, 1996; Aburto *et al.*, 1997; Ljungblad *et al.*, 1997; Teranishi *et al.*, 1997; Ljungblad *et al.*, 1998; Thode *et al.*, 2000; Watkins *et al.*, 2000; McDonald *et al.*, 2001; Mellinger and Clark, 2003; Oleson *et al.*, 2003; McDonald, in press). Although it is thought that audible downsweep calls do not show geographic variation, a more comprehensive and detailed comparison between regions may provide evidence to the contrary.

The "simple pulse" call detected by Ljungblad *et al.* (1997) shows similarities in both time and frequency to the "blurp" call detected in this study. They were unable to make field observations of the source of this "simple pulse" because most recordings were made at night. However, they state that blue whales were observed near more than half of the recording stations. It appears that the blurp calls were also detected off New Zealand alongside audible downsweep calls (McDonald, in press). In addition, Thompson *et al.* (1996) observed the occurrence of four narrow-band sounds with little frequency modulation in the presence of blue whales in the Gulf of California.

Grunt calls have been reported less frequently than other call types: only one study reports a vocalization with approximately the same time and frequency characteristics as the grunt call (Ljungblad *et al.*, 1997). As with the simple pulses, they were unable to observe the sound source but believed it to be a blue whale.

While there have been reports of blurp-like and grunt-like calls, no studies have mentioned them occurring in the tightly spaced groups seen in this study. The patterning and frequency range of the "short-irregular pulse series" reported for fin whales (Watkins *et al.*, 1987) most closely resembles the irregularly spaced portion of this mixed-pattern bout. In addition, the 11 s intercall interval found during the regularly spaced portions of these bouts coincides well with fin whale intercall intervals reported from many studies [summarized in Thompson *et al.*, 1992 (Table 5)]. However, the timing of the mixed-pattern bouts detected in this study showed a strong correlation with the surfacings of blue whale pairs.

The results of the surface and bottom interference pattern analysis on audible calls show that interference effects can account for all of the single-node segmentation seen. However, these interference effects cannot explain 30% of the calls with multinodal segmentation, leading to the conclusion that at least part of this segmentation is generated by the whales themselves. It should be noted that this is an extremely conservative estimate. While no bottom-interaction information is available for the infrasounds, the frequency range of these audible calls falls within that of the light bulb implosions. Measurements from these implosions indicate that the amplitude of the bottom-reflected arrival is very small compared to the amplitude of the direct path, which would reduce the number of calls with multinodal segmentation that can be explained by interference effects.

V. CONCLUSIONS

Although geographic variations in call characteristics have been found to exist among blue whale populations in the Pacific (Stafford *et al.*, 2001), no regional differences have been reported for the North Atlantic. This study has shown that St. Lawrence blue whale infrasonic call characteristics are similar to those from the North Atlantic (Mellinger and Clark, 2003, Nieukirk *et al.*, 2004). However, comparison of other vocalization parameters among these three studies suggests that North Atlantic regional dialects may exist in the form of differences in phrase composition and spacing. For audible downsweep calls, a more detailed comparison of call characteristics between regions should be made to determine whether they too may exhibit geographic variation. In addition, since the interference pattern analysis presented here indicates at least a partial control of call segmentation by the whales themselves, this segmentation should be examined for regional differences.

The dataset obtained from this study has shown that a great deal of variation exists for the vocalizations of St. Lawrence blue whales. Not only were many different call types of both infrasonic and audible categories detected, a wide range of characteristics were seen within each call type. Our focus in this paper has been on the acoustic dataset:

specifically, the characterization of the vocalizations and their classification into call types, as well as the spacing and sequencing patterns of these call types. Future papers will exploit the variety of recording locations and times used while collecting this data to examine spatial and temporal trends in calling behavior and will use the detailed biological observations made during data collection to determine possible behavioral contexts of the different vocalization types.

ACKNOWLEDGMENTS

The authors would like to thank Richard Sears and his team at the Mingan Island Cetacean Study for providing field support in the form of a research boat and other equipment, housing, field assistants, and access to their extensive biological database. C. Berchok would like to thank Sylvie Angel, Karine Aucrenaz, Alain Carpentier, Thomas Doniol-Valcroze, Joelle LeBreus, Brian Kot, Mylaine Lessard, Alex Liebschner, John Puschock, and Christian Ramp for all the long hours and hard work they put in as her field assistants, and Yvon Bélanger for sacrificing his basement to provide a research camp for the last two years of the study. Thanks to Diana McCammon for her insight on the segmentation analysis, and to John Puschock, Kate Stafford, and two anonymous reviewers for their helpful comments on previous drafts of this manuscript. Financial support for this research was provided by Graduate Fellowships from the National Science Foundation and the National Defense Industrial Association, Undersea Warfare Division; a Graduate Assistantship from the Penn State Applied Research Laboratory; and a Lerner Gray Fund for Marine Research Grant from the American Museum of Natural History.

[1]All these quantities are normalized by the acoustic impedance and so the "power" quantities have units of μPa^2 and the "energy" quantity has units of $\mu Pa^2/Hz$.

[2]An explanation of the parameters used to determine whether a source location was reasonable or not can be found in the sections on call segmentation below.

[3]To keep the results from this section concise and consistent with other studies, A and B will be used to symbolize infrasonic monotonic and downsweep calls, respectively.

[4]Interestingly, this pattern was repeated as a mirror image in time.

[5]Only calls of Rank 1–3 were used for this analysis (i.e., 187 monotonic, 62 downsweep, and 8 hybrid calls).

[6]The normal swim speed for blue whales is about 4–5 m/s, although they can reach speeds of up to 10 m/s in short (<10 min) bursts (Gambell, 1979). Dive/ascent speeds in one study of tagged blue whales ranged between 1–4 m/s (Acevedo-Gutiérrez *et al.*, 2002).

[7]A compressional attenuation coefficient of 0.8 dB/λ_p (Jensen *et al.*, 2000, Table 1.3) and a summer sound speed profile (downward refracting) were used in the model.

[8]For this comparison with field observations, the total number of mixed-pattern bouts included any that contained audible blurp and/or grunt calls of Ranks 1–4.

[9]In the 10–100 Hz frequency range, ambient noise levels in this study were found to range from 80 to 140 dB (with most above 100 dB) *re* 1 μPa.

Aburto, A., Rountry, D. J., and Danzer, J. L. (**1997**). "Behavioral response of blue whales to active signals," Techn. Rep. No. 1746. Naval Command, Control and Ocean Surveillance Center, RDT&E Division, San Diego, p. 7.

Acevedo-Gutiérrez, A., Croll, D. A., and Tershy, B. R. (**2002**). "High feeding costs limit dive time in the largest whales," J. Exp. Biol. **205**, 1747–1753.

Alling, A., Dorsey, E. M., and Gordon, J. C. D. (**1991**). "Blue whales (*Balaenoptera musculus*) off the northeast coast of Sri Lanka: Distribution, feeding, and individual identification," in *Cetaceans and Cetacean Research in the Indian Ocean Sanctuary*, edited by S. Leatherwood and G. Donovan (United Nations Environment Programme, Nairobi), pp. 247–258.

Bass, A. H., and Clark, C. W. (**2002**). "The physical acoustics of underwater sound communication," in *Acoustic Communication*, edited by A. M. Simmons, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 15–64.

Clark, C. W. (**1995**). "Application of U.S. Navy underwater hydrophone arrays for scientific research on whales," Rep. Int. Whal. Comm. **45**, 210–213.

Clark, C. (**2003**). (private communication).

Cummings, W. C., and Thompson, P. O. (**1971**). "Underwater sounds from the blue whale, *Balaenoptera musculus*," J. Acoust. Soc. Am. **50**, 1193–1198.

Cummings, W. C., and Thompson, P. O. (**1994**). "Characteristics and seasons of blue and finback whale sounds along the U.S. west coast as recorded at SOSUS stations," J. Acoust. Soc. Am. **95**, 2853(A).

Edds, P. L. (**1982**). "Vocalizations of the blue whale, *Balaenoptera musculus*, in the St. Lawrence River," J. Mammal. **63**, 345–347.

Edds, P. L. (**1988**). "Characteristics of finback *Balaenoptera physalus* vocalizations in the St. Lawrence Estuary," Bioacoustics 1, 131–149.

Gambell, R. (**1979**). "The blue whale," Biologist (London) **26**, 209–215.

Heard, G. J., McDonald, M., Chapman, N. R., and Jaschke, L. (**1997**). "Underwater light bulb implosions: a useful acoustic source," *Proc. Oceans '97*, Vol. **2**, pp. 755–762.

Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (**2000**). *Computational Ocean Acoustics* (Springer-Verlag, New York), p. 38.

Kibblewhite, A. C., Denham, R. N., and Barnes, D. J. (**1967**). "Unusual low-frequency signals observed in New Zealand waters," J. Acoust. Soc. Am. **41**, 644–655.

Ljungblad, D. K., Stafford, K. M., Shimada, H., and Matsuoka, K. (**1997**). "Sounds attributed to blue whales recorded off the southwest coast of Australia in December 1995," Rep. Int. Whal. Comm. **47**, 435–439.

Ljungblad, D. K., Clark, C. W., and Shimada, H. (**1998**). "A comparison of sounds attributed to pygmy blue whales (*Balaenoptera musculus brevicauda*) recorded south of the Madagascar Plateau and those attributed to 'true' blue whales (*Balaenoptera musculus*) recorded off Antarctica," Rep. Int. Whal. Comm. **48**, 439–442.

McCauley, R. D., Jenner, C., Bannister, J. L., Burton, C. L. K., Cato, D. H., and Duncan, A. (**2001**). "Blue whale calling in Rottnest trench—2000, Western Australia," Rep. No. R2001-6. Centre for Marine Science and Technology, Curtin University, Perth, Western Australia, p. 55.

McDonald, M. A., Calambokidis, J., Teranishi, A. M., and Hildebrand, J. A. (**2001**). "The acoustic calls of blue whales off California with gender data," J. Acoust. Soc. Am. **109**, 1728–1735.

McDonald, M. A., Hildebrand, J. A., and Mesnick, S. L. (**2005**). "Biogeographic characterization of blue whale song worldwide: using song to identify populations," J. Cetacean Res. Manage. (in press).

McDonald, M. A. (**2006**). "An acoustical survey of baleen whales off Great Barrier Island, New Zealand," New Zealand Journal of Marine and Freshwater Research **4** (in press).

Mellinger, D. K., and Clark, C. W. (**2003**). "Blue whale (*Balaenoptera musculus*) sounds from the North Atlantic," J. Acoust. Soc. Am. **114**, 1108–1119.

Moore, S. (**2005**). (private communication).

Nieukirk, S. L., Stafford, K. M., Mellinger, D. K., Dziak, R. P., and Fox, C. G. (**2004**). "Low-frequency whale and seismic airgun sounds recorded in the mid-Atlantic Ocean," J. Acoust. Soc. Am. **115**, 1832–1843.

Nishimura, C. E., and Conlon, D. M. (**1994**). "IUSS dual use: Monitoring whales and earthquakes using SOSUS," Mar. Technol. Soc. J. **27**, 13–21.

Oleson, E. M., Calambokidis, J. A., Burgess, W. C., McDonald, M. A., Wiggins, S. M., and Hildebrand, J. A. (**2003**). "Calling behavior of blue whales in the Southern California Bight," *1st International Conference on Acoustic Communication by Animals*, College Park, MD.

Rivers, J. A. (**1997**). "Blue whale, *Balaenoptera musculus*, vocalizations from the waters off central California," Marine Mammal Sci. **13**, 186–195.

Sears, R., Williamson, J. M., Wenzel, F. W., Bérubé, M., Gendron, D., and Jones, P. (**1990**). "Photographic identification of the blue whale *Balaenoptera musculus* in the Gulf of St. Lawrence, Canada," Rep. Int. Whal. Comm. **12**, 335–342.

Sears, R., and Calambokidis, J. (**2002**). "Update COSEWIC status report on

the Blue Whale *Balaenoptera musculus* in Canada," COSEWIC assessment and update status report on the blue whale *Balaenoptera musculus* in Canada. Committee on the Status of Endangered Wildlife in Canada, Ottawa, pp. 6 and 13.
Sears, R. (**2005**). (private communication).
Smith, K. B., and Tappert, F. D. (**2003**). Monterey-Miami parabolic equation. Source code obtained from the Office of Naval Research's Ocean Acoustics Library (http://oalib.saic.com/PE/mmpeintro.html).
Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**1999a**). "An acoustic link between blue whales in the eastern tropical Pacific and the northeast Pacific," Marine Mammal Sci. **15**, 1258–1268.
Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**1999b**). "Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific," J. Acoust. Soc. Am. **106**, 3687–3698.
Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**2001**). "Geographic and seasonal variation of blue whale calls in the North Pacific," J. Cetacean Res. Manage. **3**, 65–76.
Teranishi, A. M., Hildebrand, J. A., McDonald, M. A., Moore, S. E., and Stafford, K. (**1997**). "Acoustic and visual studies of blue whales near the California Channel Islands," J. Acoust. Soc. Am. **102**, 3121(A).
Thode, A. M., D'Spain, G. L., and Kuperman, W. A. (**2000**). "Matched-field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations," J. Acoust. Soc. Am. **107**, 1286–1300.
Thompson, P. O., and Friedl, W. A. (**1982**). "A long term study of low frequency sounds from several species of whales off Oahu, Hawaii," Cetology **45**, 1–19.
Thompson, P. O., Findley, L. T., and Vidal, O. (**1992**). "20-Hz pulses and other vocalizations of fin whales, *Balaenoptera physalus*, in the Gulf of California, Mexico," J. Acoust. Soc. Am. **92**, 3051–3057.
Thompson, P. O., Findley, L. T., Vidal, O., and Cummings, W. C. (**1996**). "Underwater sounds of blue whales, *Balaenoptera musculus*, in the Gulf of California, Mexico," Marine Mammal Sci. **12**, 288–293.
Urick, R. J. (**1983**). *Principles of Underwater Sound* (Peninsula Publishing, Los Altos, CA), pp. 131–135.
Watkins, W. A. (**1981**). "Activities and underwater sounds of fin whales," The Scientic Reports of the Whales Research Institute **33**, 83–117.
Watkins, W. A., Tyack, P., Moore, K. E., and Bird, J. E. (**1987**). "The 20-Hz signals of finback whales (*Balaenoptera physalus*)," J. Acoust. Soc. Am. **82**, 1901–1912.
Watkins, W. A., Daher, M. A., Reppucci, G. M., George, J. E., Martin, D. L., DiMarzio, N. A., and Gannon, D. P. (**2000**). "Seasonality and distribution of whale calls in the North Pacific," Oceanogr. **13**, 62–67.

Three-dimensional localization of sperm whales using a single hydrophone[a)]

Christopher O. Tiemann[b)]
Applied Research Laboratories, University of Texas at Austin, P. O. Box 8029, Austin, Texas 78713-8029

Aaron M. Thode
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

Janice Straley
University of Alaska Southeast, Sitka Campus, 1332 Seward Avenue, Sitka, Alaska 99835

Victoria O'Connell
Alaska Department of Fish & Game, 304 Lake Street, Room 103, Sitka, Alaska 99835

Kendall Folkert
P. O. Box 6497, Sitka, Alaska 99835

(Received 2 March 2006; revised 30 June 2006; accepted 17 July 2006)

A three-dimensional localization method for tracking sperm whales with as few as one sensor is demonstrated. Based on ray-trace acoustic propagation modeling, the technique exploits multipath arrival information from recorded sperm whale clicks and can account for waveguide propagation physics like interaction with range-dependent bathymetry and ray refraction. It also does not require ray identification (i.e., direct, surface reflected) while utilizing individual ray arrival information, simplifying automation efforts. The algorithm compares the arrival pattern from a sperm whale click to range-, depth-, and azimuth-dependent modeled arrival patterns in order to estimate whale location. With sufficient knowledge of azimuthally dependent bathymetry, a three-dimensional track of whale motion can be obtained using data from a single hydrophone. Tracking is demonstrated using data from acoustic recorders attached to fishing anchor lines off southeast Alaska as part of efforts to study sperm whale depredation of fishing operations. Several tracks of whale activity using real data from one or two hydrophones have been created, and three are provided to demonstrate the method, including one simultaneous visual and acoustic localization of a sperm whale actively clicking while surfaced. The tracks also suggest that whales' foraging is shallower in the presence of a longline haul than without. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2335577]

PACS number(s): 43.80.Ka, 43.30.Sf, 43.30.Wi [WWA] Pages: 2355–2365

I. INTRODUCTION

Passive acoustic methods for detecting and observing marine mammal activity are common in censusing, behavior studies, and mitigation efforts, provided the animals are acoustically active.[1,2] When used to localize animals, such techniques have the advantages of being unobtrusive and working when visual methods are unsuitable, such as at night or when animals are underwater. There has been a gradual advancement in the sophistication of passive acoustic localization techniques for marine mammal studies. For example, early applications used hyperbolic fixing, a geometric method that exploits the time differences of arrival of a whale call heard on multiple hydrophone pairs. However, the analytic solutions for these techniques assume that the waterborne sound speed is uniform throughout the water column,[3–8] an assumption that is invalid in many ocean regions. More recent techniques overcome inaccuracies from the isovelocity assumption through the use of acoustic propagation models that can account for ray-refractive effects, but they require improved knowledge of the sound speed profile to do so.[4,9–12]

Further advancement came with the ability to exploit multipath arrivals, i.e., echoes off the ocean surface or bottom that arrive at a hydrophone at different times, in order to use as few as one receiver to make a location estimate.[13–16] However, those localizations were limited to two-dimensional range and depth estimates of animal location. When similar analytic techniques are applied to two or more sensors, such as in a towed array, a bearing estimate can be added, but often with left/right ambiguity.[17,18] Another technique uses a history of range estimates from a single hydrophone to make a hydrophone-relative three-dimensional track of animal motion, but it does not provide an absolute measurement of azimuth.[19]

Although the use of multipath information for range-depth tracking is a standard procedure in underwater acoustics, its use in three-dimensional source localization is still a research topic. Three-dimensional passive acoustic localiza-

[a)]Portions of this work were presented in "Model-based passive acoustic tracking of sperm whale foraging behavior in the Gulf of Alaska" by C. Tiemann, A. Thode, J. Straley, K. Folkert, and V. O'Connell at the 150th ASA Meeting, Minneapolis, Minnesota, September 2005.

[b)]Electronic mail: tiemann@arlut.utexas.edu

tion was first briefly explored in the context of matched-field processing (MFP).[20,21] Although these were simulations of low-frequency noise fields on large vertical arrays, their conclusions are still relevant: a complex, azimuthally dependent environment does not degrade MFP, but rather enhances it by adding diversity of acoustic structure across the array. In other words, an azimuthally varying environment breaks azimuthal symmetry, thus providing azimuthal information. This work investigates whether the same lesson is applicable to a new geometry, where spatial diversity across a vertical array is replaced with frequency diversity in multipath arrivals at a single receiver, and in this case, azimuthal distinction comes from varied bathymetry rather than sound speed profiles.

This paper demonstrates a refined method for making three-dimensional tracks of animal motion underwater using impulsive sounds collected from just a single receiver, provided that the bottom bathymetry around the area is sufficiently diverse such that ocean bathymetry profiles referenced from the receiver location vary considerably as a function of azimuth. Real acoustic data recorded in the Gulf of Alaska in 2004 provides data to demonstrate the method. The technique described below not only exploits multipath arrival information, but also has the added advantage of not requiring explicit ray arrival identification (i.e., direct path, surface-reflected, etc.). However, in removing the assumption of a flat bathymetry to help resolve azimuthal direction, more detailed knowledge of the bathymetry around a deployment area is required.

Sperm whales are a vocally active species of deep-diving whale distributed throughout the world's oceans, the males of which are known to travel and forage at high latitudes.[22–30] Concurrent with this foraging behavior a sperm whale can make thousands of impulsive "clicks"[31–33] during a dive, which can be recorded and used to make estimates of their motion underwater. In 2004, acoustic recordings of sperm whales off the coast of Sitka, AK, were made as part of a multi-institutional effort called SEASWAP, the Southeast Alaska Sperm Whale Avoidance Project. The purpose of the study was to understand how sperm whales were depredating demersal longline fishing operations; whales have been depredating mainly sablefish off longlines since at least 1995.[34,35] The acoustic recordings were made from the longlines themselves, and the geometry of this deployment, combined with the large source levels of the clicks and the relatively shallow depth of water at the site, resulted in each recorded click event having several multipath arrivals associated with it. While such information is normally used to make estimates of a source's range and depth, the azimuthal dependence of the environment around a fixed receiver allowed for a range, depth, and bearing estimate to be made using data from a single hydrophone.

This paper describes the model-based, numerical localization technique, providing three examples of three-dimensional (3D) tracks made using data from autonomous acoustic recorders attached to fishing gear. Section II describes the equipment used and acoustic data obtained, and Sec. III explains the localization algorithm, including a way to extract and view the evolving multipath arrival patterns from many clicks in a click train. Section IV presents three tracking examples, one of which was verified with visual "ground truth" observations. When discussing the resulting tracks, the paper presumes an animal is at a foraging depth if its depth changes less than 50 m over 4 min.

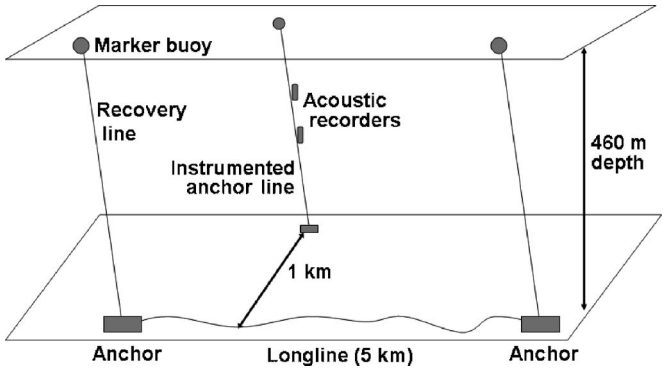


FIG. 1. Schematic of longline deployment plus instrumented anchor line, 1 km west of the longline, with acoustic recorders at depths 83 and 155 m.

II. DATA COLLECTION

There is an active longline fishery for sablefish in the eastern Gulf of Alaska, within 12 miles of shore, at the edge of a steep continental shelf. It is along this shelf edge that acoustic and visual observations of sperm whale activity have been made as part of the SEASWAP program. A demersal longline deployment typically consists of about 3 miles of hooks on a line that lies on the ocean floor, plus two buoyed vertical anchor lines at each end. For this experiment, an anchor line became a temporary vertical array by attaching two autonomous recording packages to it at depths of 83 and 155 m, as shown in Fig. 1. These recorders, built by Greeneridge Sciences, Inc.,[36] are battery powered and housed in 25 cm long acrylic pressure cases. They sampled acoustic data from a HTI-96-MIN/3V hydrophone (−172 dB re 1 V/uPa sensitivity) at 150 19 Hz, storing the data to 1 GB of onboard flash memory with 16-bit precision.

At 07:53 on May 9, 2004, the commercial fishing vessel F/V Cobra deployed an anchor line, with acoustic recorders attached, at the location shown in Fig. 2 (57.2629°N, 136.3495°W) in a local water depth of 460 m. At 09:04 the vessel started recovering longline gear that had been previously deployed the night before. Two sperm whales were first sighted next to the fishing vessel at 10:08, 930 m from the acoustic recording anchor site. By 11:00 all gear had been hauled except for the single anchorline containing the autonomous recorders. The acoustic recorders were recovered at 13:00 and the data transferred to hard disk for manual examination. Sperm whale clicks were observed throughout most of the recording period, often in long click trains up to 18 min long, with occasional overlapping of clicks from two individuals. Interclick intervals within a click train were around 1 s, a typical value for sperm whales.[37,38] Through spectrogram analysis it was noted that all click events had several associated multipath arrivals, up to eight arrivals per click.

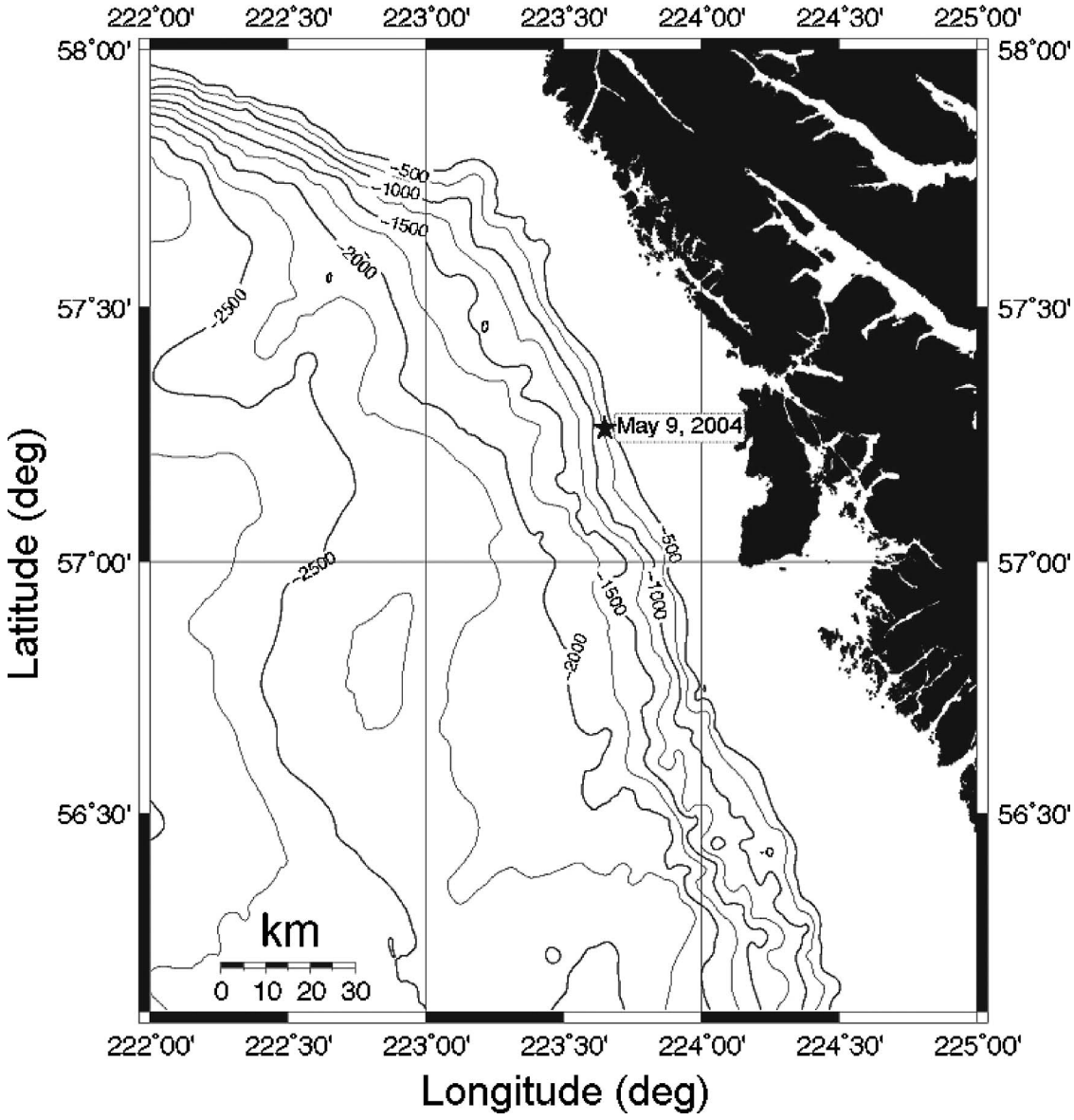


FIG. 2. Contour map (depth in meters) of the fishery near Sitka, AK, shows a steep continental shelf. The location of the array deployment on May 9, 2004, is marked with a star.

III. LOCALIZATION ALGORITHM

The model-based localization algorithm consists of three main components. First, sperm whale clicks in the acoustic data are identified and the associated multipath arrival patterns for each click event are extracted. The arrival patterns for each click in a click train evolve as the whale's position relative to the receiver changes. Next, a ray tracing propagation model predicts the multipath arrivals that would be received from hypothesized sources (whales) on a grid of several candidate ranges, depths, and bearings around a receiver. Finally, a measured arrival pattern ("data") and predicted arrivals ("replicas") are compared, and the candidate source position whose replica most closely matches the data is declared the best estimate of the whale's location for that click. By repeating this process for every click in a click train, a 3D track of the animal's motion can be made.

A. Arrival pattern extraction

A brief (20 ms), broadband sperm whale click and its associated multipath arrivals appear as vertical stripes spanning all frequencies when acoustic data are viewed as a spectrogram like that of Fig. 3(a); this spectrogram was made using a set of 256-point fast Fourier transforms with 50% window overlap. The figure shows four clicks detected by the receiver at 83 m depth taken at the initiation of a click train. Four complete multipath arrival patterns are visible, spaced approximately 0.8 s in time; the first arrivals from each pattern start at 15.3 s, 16.1, 16.9, and 17.7 s on the time axis. From the spectrogram, the relative timing and amplitude of all arrivals in an arrival pattern can be discerned.

To assist in detection of multipath arrivals, the spectral amplitudes in each time bin are integrated over frequencies from 2000 to 6600 Hz, and the resulting normalized spectral sum is shown in Fig. 3(b) for the same time period. A similar

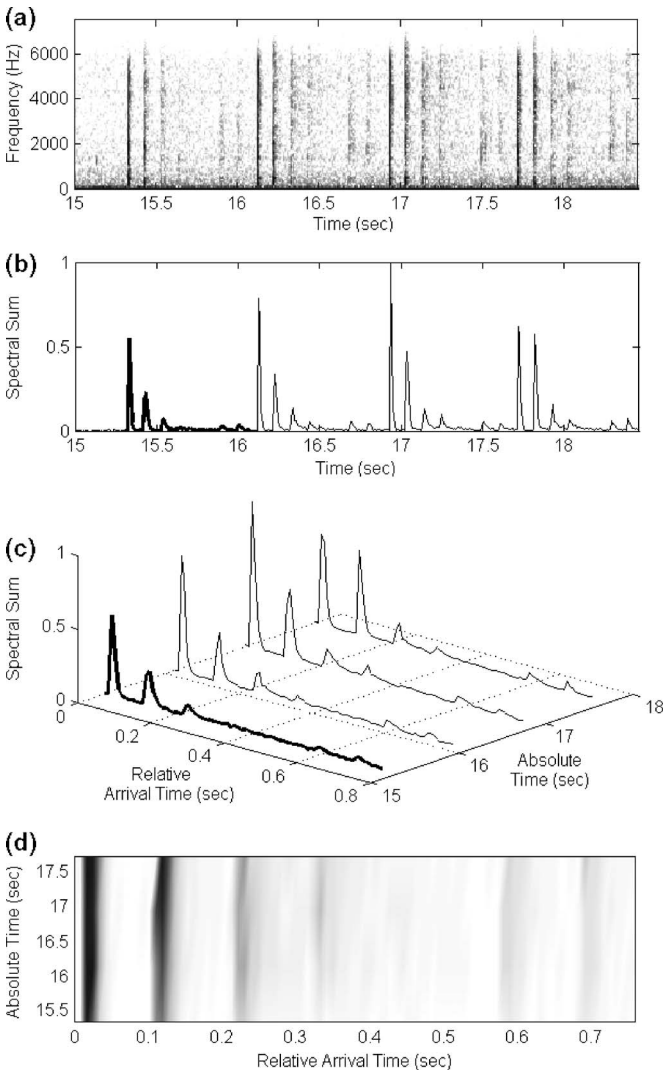


FIG. 3. (a) Spectrogram of acoustic data from receiver at 83 m depth starting at 11:42:15 on May 9, 2004. Broadband sperm whale clicks appear as vertical stripes. (b) Spectrogram summed over frequency bins; broadband clicks appear as peaks in the normalized spectral sum time series. Arrival pattern from one click event is highlighted. (c) Highlighted arrival pattern is used to find similar arrival patterns in spectral sum which are extracted and time aligned. (d) Arrival patterns from (c) represented as a 2D surface that conveys relative arrival time and amplitude of arrivals via grayscale intensity.

technique has been successfully used in the past for automatically detecting sperm whale sounds in a lower frequency band.[30,39] The integration makes broadband events appear as peaks in the spectral sum time series, and the use of frequencies over 2000 Hz ensures that low-frequency noise from the fishing vessel does not contaminate the search for click events. From the spectral sum, a pattern of six arrivals can be seen repeating every 0.8 s: two relatively loud arrivals first (probably the direct and surface-reflected ray paths) followed by four low-amplitude arrivals. The goal is to recognize and extract arrival patterns for hundreds of clicks in a click train in an automated fashion. This is further complicated in that the multipath arrival pattern will evolve over time as the whale changes position relative to the receiver, and there could be multiple clicking whales, each with different arrival patterns, being recorded at the same time.

The arrival pattern extraction process begins by identifying one complete arrival pattern, usually the first in a click train, from the spectral sum time series which will then serve as an initial guide for finding other arrival patterns. An example is shown as the highlighted segment in Fig. 3(b). This identification is done manually via a MATLAB-based graphical interface in this initial application. The assumption is made that the shapes of arrival patterns will not vary greatly between adjacent click events, so a cross correlation between the guide and 2 s of the spectral sum time series following it is performed. The peak in the cross correlation output indicates when the next click event, with a similar arrival pattern, occurred, provided the cross correlation peak exceeds some threshold (in this example, 0.8 on a normalized scale). The new arrival pattern, assumed to be the same length as the guide pattern, is saved, and the process repeats by cross correlating the guide and the next 2 s of the spectral sum time series. Because the arrival pattern is expected to change over time, the guide is updated with each iteration by taking an average of the last three arrival patterns found. The search process ends whenever no cross-correlation peak exceeds the threshold, typically at the end of the click train. This process even had some success in following one whale's arrival patterns when overlapped by a second whale's clicks.

All of the arrival patterns found using the method above are then assembled in a way that more easily allows the persistent multipath arrivals to be seen. Figure 3(c) illustrates a preliminary step in doing so by time aligning the first arrival from the click events in Fig. 3(b). Note that the absolute time axis of Fig. 3(c) indicates the start time of a given arrival pattern as read off the time axis of Fig. 3(b). The relative arrival time axis of Fig. 3(c) shows the time elapsed since the first arrival of a given click event. The contents of Fig. 3(c) are then displayed as a two-dimensional (2D) surface like that of Fig. 3(d), where each horizontal slice conveys the relative amplitude and arrival time of the arrivals for a given click event. Note that when drawing the surface, interpolation is used to fill in the gaps in absolute time between the four click events.

When several minutes of arrival pattern information are represented as a surface like that of Fig. 3(d), the gradual changes in the relative timing between multipath arrivals, and even the number of arrivals, can be seen. For example, Fig. 4(a) shows 5 minutes of arrival pattern evolution from the receiver at 83 m depth, including the data of Fig. 3(d). From such surfaces, relative arrival time information is extracted to provide one input to the localization algorithm. Note that the arrivals' amplitude information is not needed, at least in this initial version of the application, nor is there a need to identify the ray arrivals as being from direct or reflected paths. A MATLAB-based graphical interface allows a user to manually identify a few points along a persistent arrival from a surface like Fig. 4(a). Relative arrival times are then interpolated between the manually identified points for every timestamp where a click event was noted. The output of this arrival time extraction, called an "arrival map" here, is shown in Fig. 4(b) and is passed to the localization algorithm described in Sec. III C.

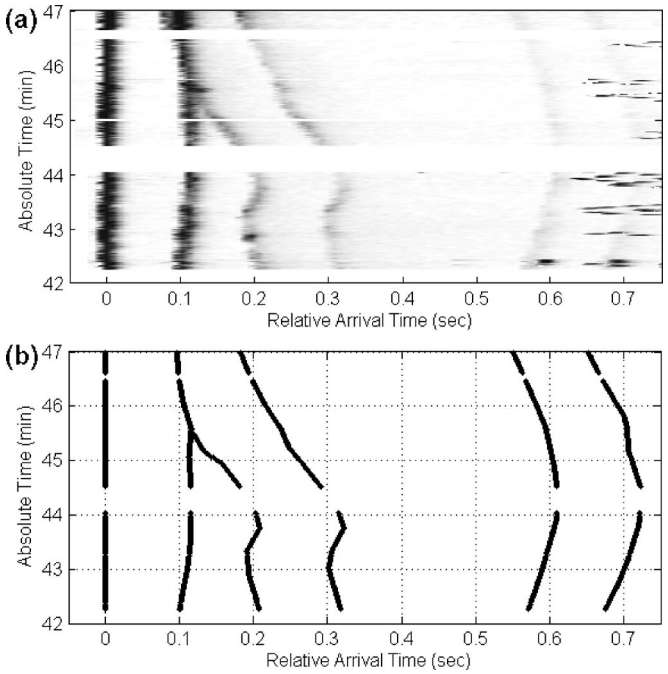


FIG. 4. (a) Arrival pattern surface for receiver at 83 m depth starting at 11:42 on May 9, 2004. (b) Relative arrival times are extracted from surface in (a) to make an arrival map. Arrival patterns evolve over time as the source (whale) changes position relative to the receiver.

B. Replica generation

The second input needed for the localization process is the replica, or predictions of relative multipath arrival times that would be recorded at a receiver from hypothesized sound sources on a grid of several ranges, depths, and bearings around the receiver. The replica only needs to be precomputed once prior to attempting a localization, but it is specific to one environment and array geometry; a new replica must be computed should the receivers be moved to a new location.

The Gaussian beam acoustic propagation model BELLHOP is used to calculate acoustic travel times as it can account for depth-dependent sound speed profiles and range-dependent bathymetry.[40,41] Simulated sources are spaced every 10 m in depth down to 800 m and every 10 m in range out to 3000 m away from the receivers. The model uses a different bathymetry profile for radials at every 5° in azimuth, making the replica range, depth, and bearing dependent. Conventional public bathymetry databases did not have the required accuracy or resolution for this application, so echosounder data from the F/V Cobra were collected to map the bathymetry around the receivers. The model assumes a source frequency of 4000 Hz and a single range-independent downward-refracting sound speed profile taken from the Levitus database of average historical sound speed profiles for the deployment location. Geoacoustic properties of the sea floor were assumed to be typical of sand: density 1.9 g/cm,^{3} compressional wave speed 1650 m/s, and compressional wave attenuation 0.8 dB/wavelength.[42,43] Use of a different bottom type would influence arrival amplitudes but not the arrival times as used here.

Figure 5 shows an example of output from the replica calculation for the receiver at 83 m depth, assuming a source at 280 m depth over a bathymetry radial on bearing 115°C. The relative arrival time and amplitude of arrivals from a source at a single range are read as a horizontal slice from the figure, thus allowing one to see how the arrival patterns evolve as a source is moved out to 3000 m from the receiver. Ignoring the amplitude information, the relative arrival time information is saved as the second input to the localization algorithm.

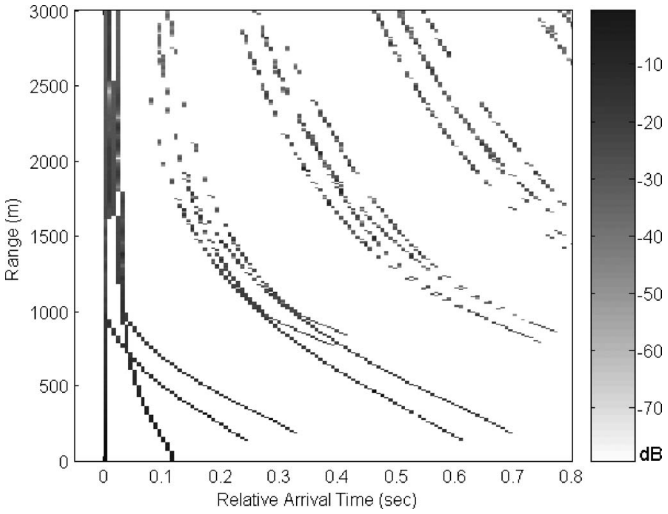


FIG. 5. Output of ray tracing model shows predicted arrival patterns at receiver at 83 m depth from sources at 280 m depth, 0–3000 m range, over a range-dependent bathymetry radial on a bearing of 115° from the receiver.

C. Ambiguity surface construction

A clicking sperm whale is localized through the construction of bearing-dependent ambiguity surfaces, or range-depth slices along a radial that graphically convey the whale's estimated location. These surfaces have the same resolution as the hypothesized source grid since each bin is assigned a score based on how closely a measured arrival pattern (data) matches the modeled arrival pattern (replica) for that source position. These scores effectively collapse all of the model/data mismatch in arrival times into a single number, and once all the scores are calculated for surfaces on all bearings, the location of the maximum overall score is declared the best estimate of source range, depth, and bearing. Note that although a ray-tracing model was used to generate the arrival pattern replicas for this example, replicas made through any technique can be used in the ambiguity construction described below.

Calculating a score for every candidate source position is done by first counting the number of measured arrivals that have the same relative arrival time as modeled arrivals from the replica for that position. Recognizing that each ray arrival has a finite length (about 20 ms), a tolerance is defined for declaring a match between a measured and modeled arrival. In this application, measured and modeled relative arrival times within 15 ms of each other are considered to be overlapping. All modeled arrivals that are within that same tolerance of each other are counted as just one arrival, and modeled arrivals occurring later than the latest measured

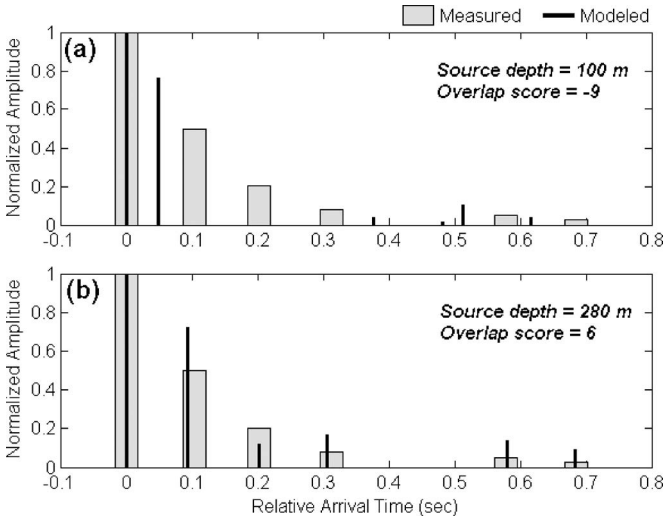


FIG. 6. (a) Arrivals from a modeled source at 200 m range, 100 m depth, bearing 115° overlaid on measured arrivals from 11:42:15 on May 9, 2004. Few arrivals overlap, so it is given a low "overlap score" indicating a low likelihood that the true source is at the modeled source location. (b) Arrivals from a modeled source at 200 m range, 280 m depth, bearing 115° overlaid on measured arrivals from 11:42:15 on May 9, 2004. All measured and modeled arrivals overlap, so this is the best estimate of true source location.

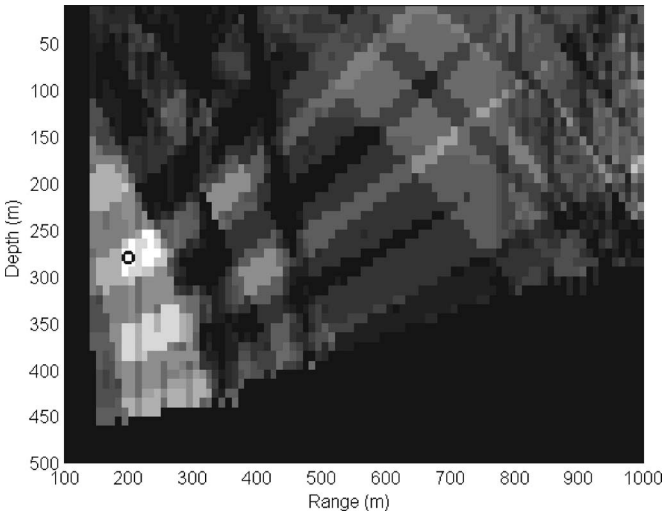


FIG. 7. An ambiguity surface showing overlap scores on a grid of ranges and depths along bearing 115° from the receiver. The peak at 200 m range, 280 m depth is marked and indicates best estimate of source location for the click event recorded at 11:42:15 on May 9, 2004.

relative arrival time are not considered. From that number of overlapping arrivals, the number of nonoverlapping modeled and measured arrivals are subtracted to give a net "overlap score." This approach for scoring was used instead of some least-squares fit between the modeled and measured arrivals primarily because there needs to be a scoring penalty assessed when there are extraneous multipaths in the modeled arrivals that are not seen in the data. Furthermore, allowing the tolerance window for matching arrivals adds robustness against environmental modeling mismatch.

A visual example of this overlap scoring process for a single click event is shown in Figs. 6(a) and 6(b). Both figures show the relative arrival time and amplitude information for the same measured arrival pattern, but different modeled arrivals from sources at two different depths (100 and 280 m depth). The first arrivals from each are time aligned at the 0 s mark, with additional arrivals spaced along the relative time axis. Note that the measured arrival times are shown as gray bars with a width of 30 ms centered around the times extracted from the arrival pattern maps; this is to represent the matching tolerance described above. The example in Fig. 6(a) shows only one overlapping arrival (the first arrival at 0 s) plus ten that do not overlap, resulting in a low overlap score and a low likelihood that the whale is at this modeled source location. Figure 6(b) shows all arrivals overlapping for a high overlap score of 6, so this becomes the best estimate of the whale location. Note that although normalized amplitude information is presented here to convey the model's good agreement to the data, it is not used in the scoring process.

The scoring process is repeated for all candidate source positions, and when concluded, the scores are assembled into ambiguity surfaces like that shown in Fig. 7. High scores are represented by bright spots on this range/depth slice, indicating likely source positions. An ambiguity surface like this is made for all search bearings, and the maximum score among all bearings is declared the best estimate of source location. This ambiguity surface construction can be repeated for all click events in a click train, allowing assembly of a track of whale location versus time, examples of which will be shown in Sec. IV.

Occasionally the overlap scoring process will result in multiple search bins that have the same score. For example, some adjacent range bins may have replicas whose arrivals are within the 15 ms tolerance of each other and thus could all overlap the same measured arrivals, resulting in the same overlap score. In those cases, a "tie-breaker" calculation is done by summing the small differences in arrival times between nearest measured and modeled arrival pairs. The replica with the smallest summed error is declared the best match.

Another situation that can cause multiple search bins to have the same score, even after the tie-breaking calculation, occurs if the bathymetry between source and receiver is the same along different bearings, thus causing their replicas to be identical. To be more specific, replicas will be identical if ray interactions with the ocean bottom occur at the same depths and ranges, regardless of the bathymetry between those reflection points. This situation happened in at least one case of trying to track an animal at close range to the southeast of the receivers where bearing resolution became limited to a 20° sector due to low resolution of the available bathymetry data. The remedy for this case is to have higher resolution bathymetry data to use in the replica generation. Another remedy used here is to look in a complete track history for times when bearing could be uniquely determined and assume that same bearing for questionable times.

D. Sensitivity analysis

There are two types of error that can negatively affect the accuracy of this localization technique, and sensitivity to both was tested. Given that this is a model-based solution,

mismatch between the assumed and truth environment is one source of error, while another comes from errors in measurements of the relative travel times for arrivals in a given click event. The combined effect of both model and measurement error is a nonlinear process, but two examples can give a sense of the relative size of their effects. To simulate the effects of an error in modeled sound speed profile, the arrival pattern from a source at 800 m range, 200 m depth, along bearing 170° is calculated using a historic sound speed profile and saved as "truth data." (This source location was chosen because it matches one estimated in Sec. IV A to follow.) Next, a replica is generated using an isovelocity sound speed profile at 1475 m/s and a localization attempted on the simulated data. The resulting location estimate is 14 m away from the "truth" source location, implying a relatively small error from sound speed mismatch.

In the next example, the same historic environment is used to make both the replica and truth data for a source at the same location: 800 m range, 200 m depth, and bearing 170°. This would normally result in a perfect localization, but noise will be added to the travel time data to simulate measurement error. Given that each real multipath arrival has a length of about 20 ms, a measurement of its arrival time within 10 ms of the truth should be possible. Therefore, a random error not to exceed 10 ms was added to each simulated arrival time; a constant error was not added to each arrival as then the relative travel times would not change, thus resulting in a perfect localization. Repeated localizations with 100 realizations of travel time measurement error had a mean error of 16 m away from the true source location, again relatively small and comparable to that of sound speed mismatch. With confidence that the localization technique should be reasonably robust against the expected errors, its application to real data is presented next.

IV. LOCALIZATION RESULTS

This section presents some tracks of sperm whale motion as estimated by the model-based tracking algorithm, including pictures of the arrival patterns that contributed to the localizations. Examples were chosen that illustrate whale behavior in the presence and absence of human fishing activity, using receivers either independently or jointly. One "ground truth" case describes an unusual whale behavior that was observed both visually and acoustically.

A. Whale dive profile absent fishing activity

The first example track comes from data of May 9, 2004, 11:18–11:33, when all fishing activity in the area had been completed. The arrival pattern extraction process was applied to this time period, and the resulting surface of aligned arrival patterns recorded from the receiver at 155 m depth is shown in Fig. 8(a). Figure 8(b) shows the resulting track estimates of range and depth over time at a bearing of 170° from the array as estimated by both receivers independently. This track follows a sperm whale diving from the surface to a depth of 340 m while closing in range to the receivers. During this time the F/V Cobra was moving toward the instrumented anchor line, and visual observations

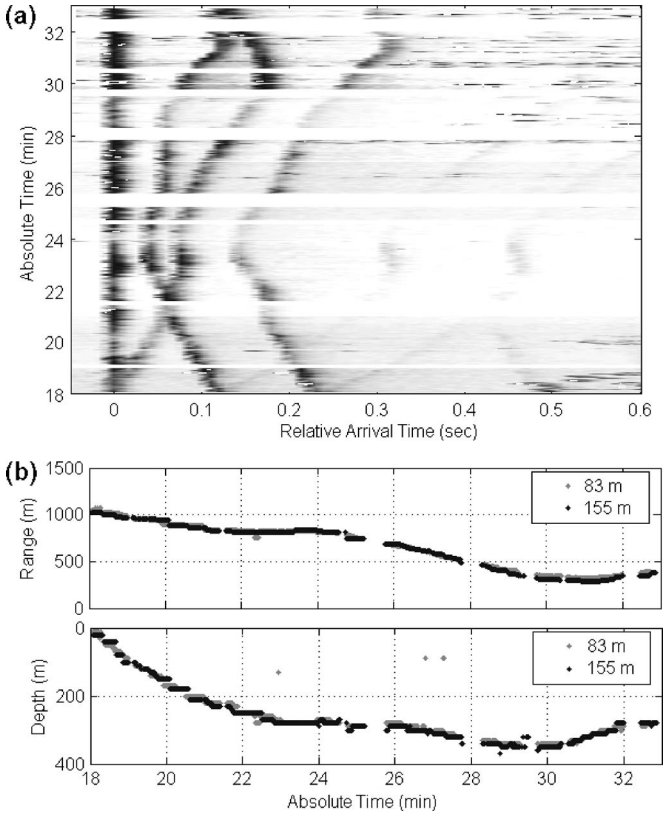


FIG. 8. (a) Arrival patterns from receiver at 155 m depth starting at 11:18 on May 9, 2004. (b) Range and depth estimates along bearing 170° as derived using arrivals data from receivers at 83 and 155 m independently. Both receivers track a whale diving from the surface in the absence of human fishing activity, consistent with divergence of direct and surface-reflected arrivals seen in (a) at 11:18.

from the vessel indicated that the animals were following in the wake, thus explaining why the range is decreasing in this example. This foraging depth is shallower than what has been reported elsewhere, but the Alaskan waters are also shallower than other areas of sperm whale observation.[16,17,44]

Not only are tracks from the two receivers in agreement in estimating a dive from the surface, but analytic solutions assuming an isovelocity sound speed profile and flat bottom bathymetry (technique described in Ref. 16) yielded the same 2D results (range and depth only) at ranges less than 300 m. The conclusion that the animal was clicking at the surface and then initiated a dive is also consistent with the arrival pattern evolution shown in Fig. 8(a). It would be difficult to distinguish between a direct ray path arrival and surface-bounce arrival as received from a very shallow source, such as a whale at the surface. However, the travel times for those two ray paths would diverge as a source moves deeper, and that is what is seen in the arrival pattern data as the direct path arrival (relative arrival time=0 s) splits into two arrivals immediately at the start of the click train. The other loud arrivals at times 0.11 s and 0.22 s, probably representing the bottom-bounce and bottom-surface-bounce ray paths, also bifurcate as the source deepens.

B. Whale dive profile during fishing activity

The next tracking example begins at 09:26 on May 9, 2004, around 2 h earlier than the previous example. The tim-

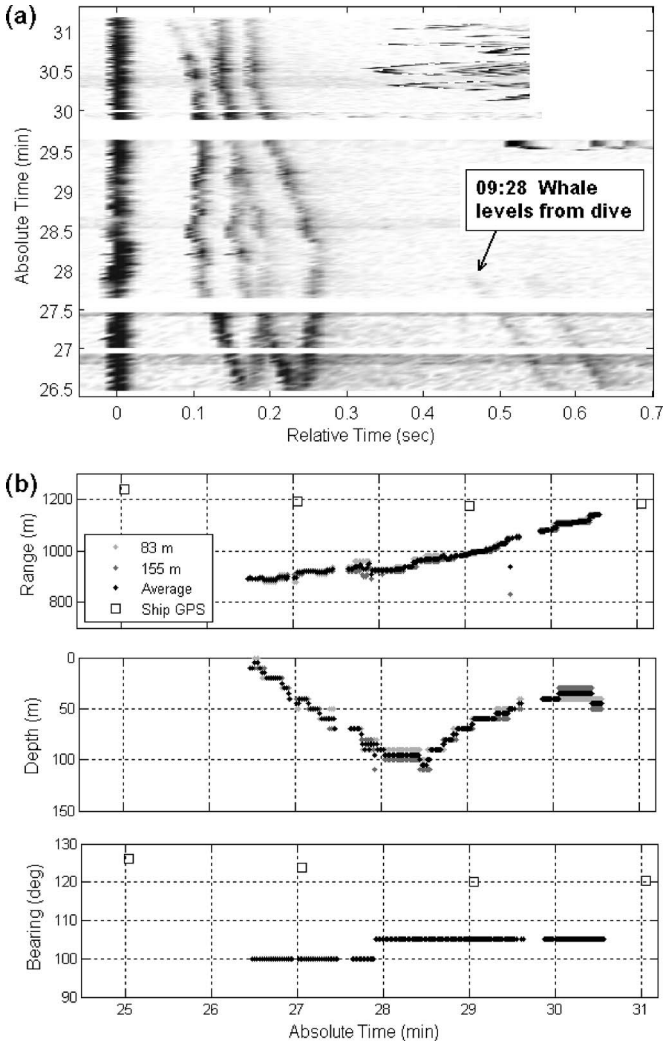


FIG. 9. (a) Arrival pattern surface for receiver at 83 m depth starting at 09:26:30 on May 9, 2004. (b) Range, depth, and bearing estimates as derived using arrivals data from receivers at 83 and 155 m jointly, plus fishing boat location per GPS data. Track follows a whale diving to an unusually shallow foraging depth as whale swims towards a fishing vessel conducting a longline recovery. Late arrivals in (a) disappear at 09:28 as depth track shows whale leveling from dive, implying directionality of sound source.

ing of this track is interesting because it begins just a few minutes after a longline recovery had begun, so the whale may be reacting to sounds unique to fishing activity. Arrival patterns recorded from the receiver at 83 m depth are shown in Fig. 9(a), and Fig. 9(b) shows the resulting track estimates in range, depth, and bearing. Per range and bearing estimates it is confirmed that the whale is swimming at a steady 1 m/s away from the receivers and converging on the F/V Cobra, whose GPS location is also indicated on Fig. 9(b); GPS location error was on the order of 5 m. This track also follows a sperm whale that began clicking at the surface, but in this case, the dive is relatively shallow at 105 m maximum. At the time where the whale is closest to the fishing vessel the dive depth does not exceed 50 m, a very shallow depth when compared with the usual depths at which these animals are known to make regular clicking sounds.

This example also illustrates how data from two or more receivers, if available, can be used jointly. Due to the whale's movement, its bearing relative to the array was changing slightly over the course of this track, and bearing estimates from the two receivers did not exactly agree. In order to make a single best estimate of whale location, the peak range and depth scores along all radials, not just from highest scoring bearing, were saved for the two receivers independently. For each click event, the radial where the two receivers' range and depth estimates agreed most closely was declared the best bearing estimate, and an average of the range and depth estimates on that radial was calculated. Thus, Fig. 9(b) adds a joint bearing estimate and another marker to the range and depth plots to indicate the average.

Last, this example also demonstrates how the arrival pattern maps indicate some information about the directivity of the sperm whale's beampattern. Figure 9(a) shows up to seven multipath arrivals being recorded for the click events just prior to 09:28, but immediately after that, the two latest arrivals at relative arrival times 0.45 and 0.55 s abruptly disappear. According to the track estimate of Fig. 9(b), the whale leveled out from a dive at that same time. When received rays disappear concurrent with a change in the elevation angle of a sound source a directional source is implied, supporting other reports of the directionality of the sperm whale click production mechanism.[16,45,46]

C. Visual verification

The last tracking example presents an unusual sperm whale behavior that also serves to verify the acoustic localization algorithm through comparison to a visual observation. Per the logs of F/V Cobra, a sperm whale was seen floating motionless on the water surface within 100 m of the ship at 10:26:05 on May 9, 2004, and through interpolation of GPS logs, the ship was 993 m away from the acoustic receivers on a bearing of 75° from true north at that time. The whale's position relative to the vessel was not recorded, but given its range, it would be somewhere within a sector from 69° to 81° from the array. The acoustic record around that time contained a sperm whale click train from 10:24 to 10:31, so a localization attempt was made by the two receivers independently. The deeper receiver estimated the whale's bearing at 75° from the array; the shallow receiver's bearing estimates varied between 75° and 85°.

The localization was repeated using the replica from just the 75° bearing—in effect "steering" the receivers to look towards the ship. Figure 10 shows the acoustic localizations made independently by both receivers in addition to markers for the visual localizations and ship positions; bars on the visual markers indicate a 100 m radius from the ship. Both receivers estimated a whale to be not only within 100 m of the ship but also located on the surface for several minutes, actively clicking. This is unusual as it is typically observed that whales are silent while resting on the surface (e.g., Ref. 37). As further evidence that the acoustic and visual contacts are the same whale, visual observations also note that the whale fluked up and dove underwater at 10:29:47, exactly the same time that the acoustic depth track shows the whale beginning a dive to 95 m depth. Beyond that time interference with a second clicking whale made further tracking difficult with the methods presented here.

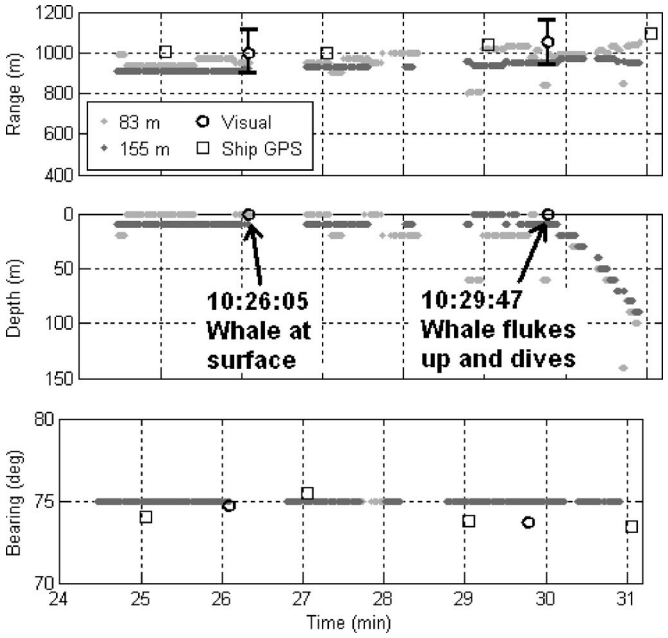


FIG. 10. Range, depth, and bearing estimates of clicking sperm whale starting at 10:24 on May 9, 2004 as derived from receivers at 83 and 155 m independently, plus locations of ship position and visual sightings of sperm whale. Acoustic track matches visual observations of whale resting at surface and diving.

D. Tracking summary

The above tracking examples are three of ten tracks total constructed from acoustic data from the morning of May 9, 2004. Metrics from all tracks made using the receivers jointly are presented in Table I, including an indicator of whether there was concurrent fishing operations and whether the track followed a dive profile beginning at the surface. The average length of a click train and its associated track was 11.6 min, with the longest being 17.7 min. The farthest estimated tracking range was 1170 m.

From this time sequence there is limited evidence that the presence of fishing boat activity may be modifying the sperm whales' dive depths. During the three tracks concurrent with fishing activity, the maximum estimated whale depth was 145 m with an average of 115 m. Absent fishing activity, all maximum depths were at least 180 m with the average foraging depth of 281 m.

TABLE I. Metrics from sperm whale tracks from May 9, 2004 derived using data from receivers at 83 and 155 m depth jointly.

Start Time 5/9/2004	Duration (min)	Max range (m)	Max depth (m)	Average bearing (deg)	During fishing?	Track starts at surface?
09:26	4.4	1170	105	105	Yes	Yes
09:45	11.6	900	145	135	Yes	No
10:24	6.4	1000	95	75	Yes	Yes
11:03	13.2	1170	205	140	No	No
11:18	14.6	1045	340	170	No	Yes
11:33	14.0	760	365	120	No	No
11:49	17.4	755	325	150	No	No
12:03	17.7	1050	350	175	No	Yes
12:57	9.7	965	205	145	No	No
13:21	7.0	880	180	145	No	No

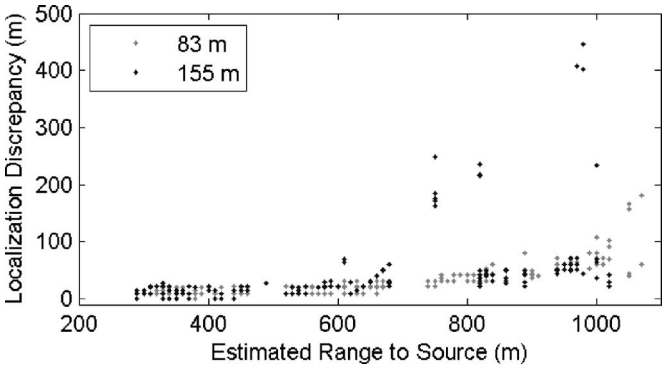


FIG. 11. Whale tracking using the data from Sec. IV A is repeated for both receivers using an isovelocity sound speed environment and compared to localizations using a historic sound speed profile. Localizations from the different replicas agree closely until the source moves beyond 700 m in range.

V. DISCUSSION AND CONCLUSION

This paper has demonstrated a passive acoustic technique for locating and tracking the movement of an impulsive sound source, which in this case are clicking sperm whales in the Gulf of Alaska. The technique uses an acoustic propagation model and a numerical solution to make a three-dimensional location estimate using data from as few as one sensor where azimuthal directivity is provided by bathymetric differences along different radials. The tracking examples provided show the algorithm being applied to data from small, rapidly deployable sensors well suited for field operations, and the examples even suggest some atypical whale behavior in the presence of human fishing activity.

It was discussed above that modeled environmental mismatch can hurt localization accuracy. One likely place for mismatch is in the sound speed profile, but such errors do not seem to have harmful effects unless localizing a source at long range from a receiver where ray refractive effects must be taken into account. For example, the localization of the whale described in Sec. IV A was repeated assuming a simple isovelocity sound speed profile at 1475 m/s, and that track was compared to the one made using the historic sound speed profile [Fig. 8(b)]. The discrepancy in the localizations arising from the use of different replicas was measured for each click event and plotted versus the range estimated using the historic sound speed profile, as shown in Fig. 11 for each of the two receivers. There is little disagreement in location estimates from the two environments until the source moves beyond 700 m in range. This is consistent with studies from other environments that have claimed that isovelocity, straight-line propagation is a suitable approximation for ranges within or one or two water depths.[11,12,17]

Another known source of environmental mismatch comes from the lack of accurate bathymetry data. This error is more critical in that it is bathymetry information that allows for bearing discrimination when using a single receiver. Early tests showed that a constant slope bathymetry assumption was not sufficient to make a unique bearing estimate; localizations over a few different slopes would score equally high but with slight shifts in their range and depth estimates. This shifting of a source location, given a mismatch in bot-

tom modeling, has been well documented in the acoustic tracking literature and has been dubbed the "mirage" effect.[47]

It was claimed above that each bearing needs a unique set of bottom interactions to do bearing discrimination, but equivalently one might say that bathymetry slices need at least two segments with different slopes to avoid the mirage effect seen in the constant slope bathymetry tests. That kind of bathymetry resolution may not exist for cases when the source is close to the receiver, so a trade off must be made between desired bearing resolution and effort spent obtaining bathymetry information. Without sufficient bathymetry information, one may have to be content with a crude bearing estimate, but this could still be considered an improvement over other two-sensor techniques that cannot resolve left-right bearing ambiguities. Furthermore, no amount of bathymetry data will yield a unique bearing estimate if an environment is genuinely flat, but range and depth estimates from a single sensor could still be obtained using this method.

Although this experiment used sperm whales as sound sources, the techniques described here should be applicable to tracking other marine mammals or manmade transient sounds. The arrival pattern extraction process can be replaced with any means for recording the impulse response of the ocean waveguide between a source and receiver. Because sperm whale clicks are brief, broadband sounds, noting their multipath arrival times gives one the impulse response, but matched-filter techniques could be used to get the same for other animal whistles or manmade sweeps, for example. Further automating the arrival pattern extraction is a goal for future refinements, as is trying to refine bearing estimation, perhaps through the use of the previously ignored arrival amplitude information or the use of multiple widely spaced, individual receivers. In principle, given enough multipath arrival information, the directivity of the source and bottom reflection coefficients might be estimated, provided that a simple model of the ocean bottom is used and the frequency-dependent amplitudes of the acoustic returns are exploited.

ACKNOWLEDGMENTS

The North Pacific Research Board funded collection of the data used in this report under Project No. F0412. Numerous individuals helped in the data collection, including Steve Weisberg, Shane Walker, Patricia Ramon, Nellie Warner, and Morgan Hartley. The Applied Research Laboratories at the University of Texas at Austin funded the data analysis and localization algorithm development. The authors also appreciate the cooperation of all the fishermen of the Alaska Longline Fisheries Association.

[1]R. Leaper, O. Chappell, and J. Gordon, "The development of practical techniques for surveying sperm whale populations acoustically," Rep. Int. Whal. Comm. **42**, 549–560 (1992).
[2]J. Barlow and B. L. Taylor, "Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey," Marine Mammal Sci. **21**(3), 429–445 (2005).
[3]C. W. Clark, W. T. Ellison, and K. Beeman, "Acoustic tracking of migrating bowhead whales," IEEE Oceans Conference Proceedings **18**, 341–346 (1986).
[4]J. L. Spiesberger and K. M. Fristrup, "Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography," Am. Nat. **135**, 107–153 (1990).
[5]A. S. Frankel, C. W. Clark, L. M. Herman, and C. M. Gabriele, "Spatial distribution, habitat utilization, and social interactions of humpback whales, *Megaptera novaeangliae*, off Hawai'i determined using acoustic and visual techniques," Can. J. Zool. **73**, 1134–1146 (1995).
[6]K. M. Stafford, C. G. Fox, and D. S. Clark, "Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean," J. Acoust. Soc. Am. **104**(6), 3616–3625 (1998).
[7]V. M. Janik, S. M. Van Parijs, and P. M. Thompson, "A two-dimensional acoustic localization system for marine mammals," Marine Mammal Sci. **16**, 437–447 (2000).
[8]C. W. Clark and W. T. Ellison, "Calibration and comparison of acoustic location methods used during the spring migration of the bowhead whale, *Balaena mysticetus*, off Pt. Barrow, Alaska, 1984–1993," J. Acoust. Soc. Am. **107**(6), 3509–3517 (2000).
[9]C. O. Tiemann, M. B. Porter, and J. A. Hildebrand, "Automated model-based localization of marine mammals near California," in *MTS/IEEE Oceans 2002 Conference Proceedings* (Holland Publications, Escondido, CA, 2002) pp. 1360–1364.
[10]C. O. Tiemann, M. B. Porter, and L. N. Frazer, "Localization of marine mammals near Hawaii using an acoustic propagation model," J. Acoust. Soc. Am. **115**, 2834–2843 (2004).
[11]A. Thode, "Three-dimensional passive acoustic tracking of sperm whales (*Physeter macrocephalus*) in ray-refracting environments," J. Acoust. Soc. Am. **118**(6), 3575–3584 (2005).
[12]E. K. Skarsoulis and M. A. Kalogerakis, "Ray-theoretic localization of an impulsive source in a stratified ocean using two hydrophones," J. Acoust. Soc. Am. **118**(5), 2934–2943 (2005).
[13]R. Aubauer, M. O. Lammers, and W. W. L. Au, "One-hydrophone method of estimating distance and depth of phonating dolphins in shallow water," J. Acoust. Soc. Am. **107**, 2744–2749 (2000).
[14]P. A. Lepper, K. Kaschner, P. R. Connelly, and A. D. Goodson, "Development of a simplified ray path model for estimating the range and depth of vocalising marine animals," in *Proc. Inst. Acoust.* (Institute of Acoustics, St. Albans, UK, 1997), pp. 227–234.
[15]W. Whitney, "Observations of sperm whale sounds from great depths," Marine Physical Laboratory, Scripps Institution of Oceanography Report MPL-U-11/68, 1968.
[16]A. Thode, D. K. Mellinger, S. Stienessen, A. Martinez, and K. Mullin, "Depth-dependent acoustic features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico," J. Acoust. Soc. Am. **112**, 308–321 (2002).
[17]A. Thode, "Tracking sperm whale (*Physeter macrocephalus*) dive profiles using a towed passive acoustic array," J. Acoust. Soc. Am. **116**, 245–253 (2004).
[18]E. K. Skarsoulis, A. Frantzis, and M. Kalogerakis, "Passive localization of pulsed sound sources with a 2-hydrophone array," *Seventh European Conference on Underwater Acoustics* (Delft, The Netherlands, 2004).
[19]C. Laplanche, O. Adam, M. Lopatka, and J. Motsch, "Male sperm whale acoustic behavior observed from multipaths at a single hydrophone," J. Acoust. Soc. Am. **118**, 2677–2687 (2005).
[20]J. S. Perkins and W. A. Kuperman, "Environmental signal processing: Three-dimensional matched-field processing with a vertical array," J. Acoust. Soc. Am. **87**, 1553–1556 (1990).
[21]C. A. Zala and J. M. Ozard, "Matched-field processing in a range-dependent environment," J. Acoust. Soc. Am. **88**, 1011–1019 (1990).
[22]W. A. Watkins, "Acoustics and the behavior of sperm whales," in *Animal Sonar Systems*, edited by R.-G. Busnel and J. F. Fish (Plenum, New York, 1980), pp. 283–290.
[23]W. A. Watkins, "Acoustic behaviors of sperm whales," Curr. Appl. Phys. **20**, 50–58 (1977).
[24]H. Whitehead, "Estimates of the current global population size and historical trajectory for sperm whales," Mar. Ecol.: Prog. Ser. **242**, 295–304 (2002).
[25]J. Barlow and B. L. Taylor, "Preliminary abundance of sperm whales in the northeastern temperate Pacific estimated from a combined visual and acoustic survey," Int. Whal. Comm., 1998.
[26]T. Lyrholm and U. Gyllensten, "Global matrilineal population structure in sperm whales as indicated by mitochondrial DNA sequences," Proc. R. Soc. London, Ser. B **265**(1406), 1679–1684 (1998).
[27]N. Jaquet, "How spatial and temporal scales influence understanding of Sperm Whale distribution: A review," Mammal Rev. **26**(1), 51–65 (1996).
[28]H. Whitehead, M. Dillon, S. Dufault, L. Weilgart, and J. Wright, "Non-geographically based population structure of south Pacific sperm whales:

dialects, fluke-markings and genetics," J. Anim. Ecol. **67**(2), 253–262 (1998).
[29]T. Lyrholm, O. Leimar, B. Johanneson, and U. Gyllensten, "Sex-biased dispersal in sperm whales: Contrasting mitochondrial and nuclear genetic structure of global populations," Proc. R. Soc. London, Ser. B **266**(1417), 347–354 (1999).
[30]D. K. Mellinger, K. M. Stafford, and C. G. Fox, "Seasonal occurrence of sperm whale (*Physeter macrocephalus*) sounds in the Gulf of Alaska, 1999-2001," Marine Mammal Sci. **20**(1), 48–62 (2004).
[31]J. C. Goold and S. E. Jones, "Time and frequency-domain characteristics of sperm whale clicks," J. Acoust. Soc. Am. **98**, 1279–1291 (1995).
[32]H. Whitehead and L. Weilgart, "Click rates from sperm whales," J. Acoust. Soc. Am. **87**, 1798–1806 (1990).
[33]L. V. Worthington and W. E. Schevill, "Underwater sounds heard from sperm whales," Nature (London) **180**, 291 (1957).
[34]A. Thode, J. Straley, C. Tiemann, V. Teloni, K. Folkert, T. O'Connell, and L. Behnken, "Sperm whale and longline fisheries interactions in the Gulf of Alaska—Passive acoustic component," North Pacific Research Board Final Report F0412, 2005, 57.
[35]P. S. Hill, J. L. Laake, and E. Mitchell, "Results of a pilot program to document interactions between sperm whales and longline vessels in Alaska waters," U.S Department of Commerce, 1999, 42.
[36]W. C. Burgess, "The bioacoustic probe: A general-purpose acoustic recording tag," J. Acoust. Soc. Am. **108**(5) Pt. 2, 2583 (2000).
[37]H. Whitehead, *Sperm Whales: Social Evolution in the Ocean* (University of Chicago Press, Chicago, IL, 2003).
[38]L. A. Douglas, S. M. Dawson, and N. Jaquet, "Click rates and silences of sperm whales at Kaikoura, New Zealand," J. Acoust. Soc. Am. **118**(1), 523–529 (2005).
[39]D. K. Mellinger, "Ishmael 1.0 User's Guide," NOAA Tech. Memo. OAR PMEL-120, 2001, 26.
[40]M. B. Porter and H. P. Bucker, "Gaussian beam tracing for computing ocean acoustic fields," J. Acoust. Soc. Am. **82**(4), 1349–1359 (1987).
[41]M. B. Porter and Y. C. Liu, "Finite-element ray tracing," in *Proceedings of the International Conference on Theoretical and Computational Acoustics*, edited by D. Lee and M. H. Schultz (World Scientific, Singapore, 1994), pp. 947–956.
[42]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, Melville, NY, 1999), p. 41.
[43]E. L. Hamilton, "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**, 1313–1340 (1980).
[44]P. J. O. Miller, M. P. Johnson, P. L. Tyack, and E. A. Terray, "Swimming gaits, passive drag and buoyancy of diving sperm whales *Physeter macrocephalus*," J. Exp. Biol. **207**(11), 1953–1967 (2004).
[45]W. M. X. Zimmer, P. L. Tyack, M. P. Johnson, and P. T. Madsen, "Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis," J. Acoust. Soc. Am. **117**, 1473–1485 (2005).
[46]B. Møhl, M. Wahlberg, P. T. Madsen, L. A. Miller, and A. Surlykke, "Sperm whale clicks: Directionality and source level revisited," J. Acoust. Soc. Am. **107**, 638–648 (2000).
[47]G. L. D'Spain, J. J. Murray, W. S. Hodgkiss, N. O. Booth, and P. W. Schey, "Mirages in shallow water matched field processing," J. Acoust. Soc. Am. **105**, 3245–3265 (1999).

Quantitative measures of air-gun pulses recorded on sperm whales (*Physeter macrocephalus*) using acoustic tags during controlled exposure experiments

P. T. Madsen[a)]
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543 and Department of Zoophysiology, Biological Institute, University of Aarhus, Aarhus, Denmark

M. Johnson
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

P. J. O. Miller
NERC Sea Mammal Research Unit, University of St. Andrews, St. Andrews, United Kingdom

N. Aguilar Soto
Department of Animal Biology, La Laguna University, La Laguna 38206, Tenerife, Spain

J. Lynch and P. Tyack
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 24 January 2006; revised 16 June 2006; accepted 26 June 2006)

The widespread use of powerful, low-frequency air-gun pulses for seismic seabed exploration has raised concern about their potential negative effects on marine wildlife. Here, we quantify the sound exposure levels recorded on acoustic tags attached to eight sperm whales at ranges between 1.4 and 12.6 km from controlled air-gun array sources operated in the Gulf of Mexico. Due to multipath propagation, the animals were exposed to multiple sound pulses during each firing of the array with received levels of analyzed pulses falling between 131–167 dB re. 1 μPa (pp) [111–147 dB re. 1 μPa (rms) and 100–135 dB re. 1 μPa^2 s] after compensation for hearing sensitivity using the *M*-weighting. Received levels varied widely with range and depth of the exposed animal precluding reliable estimation of exposure zones based on simple geometric spreading laws. When whales were close to the surface, the first arrivals of air-gun pulses contained most energy between 0.3 and 3 kHz, a frequency range well beyond the normal frequencies of interest in seismic exploration. Therefore air-gun arrays can generate significant sound energy at frequencies many octaves higher than the frequencies of interest for seismic exploration, which increases concern of the potential impact on odontocetes with poor low frequency hearing. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2229287]

PACS number(s): 43.80.Nd, 43.80.Ev, 43.80.Gx [WWA] Pages: 2366–2379

I. INTRODUCTION

A. Background

Seismic survey vessels fire towed arrays of air guns every 10–15 s to produce geo-acoustic profiles of hydrocarbon deposits in the seabed (Barger and Hamblen, 1980). The air-gun array creates a downward directed, low frequency pulse with most energy concentrated around 50 Hz and a back-calculated, broad-band source level (SL)[1] between 230 and 260 dB re. 1 μPa (0-peak) (Dragoset, 1990; Richardson *et al.*, 1995). The widespread use of this technique (Schmidt, 2004) and the ocean traversing potential of the low frequency, high powered pulses (Nieukirk *et al.*, 2004) have raised concern about the effects of air guns on marine life (Richardson *et al.*, 1995, NRC, 2000, 2005; Gordon *et al.*, 2004). A wide range of marine animal species might be affected by air-gun pulses (McCauley *et al.*, 2000, 2003) with possible negative consequences for human fisheries (Engås *et al.*, 1996). Effects of air-gun pulses on marine mammals may warrant particular concern (Richardson *et al.*, 1995; Gordon *et al.*, 2004, Caldwell, 2002), as many marine mammal species rely critically on sound for orientation, food finding, and communication (Tyack and Clark, 2000). While there are quite a few studies of the effects of seismic signals on baleen whales (Ljungblad *et al.*, 1988; Malme *et al.* 1984, 1985, 1986, 1988; Reeves *et al.*, 1984; Richardson *et al.*, 1986), data for toothed whales are limited (e.g., Goold and Fish, 1998; Stone, 2003) and have often been collected circumstantially (e.g., Madsen *et al.*, 2002).

Investigations of how noise may affect the behavior of marine mammals benefit from methods that can estimate reliably the received noise levels at the whale along with concomitant logging of relevant behavioral parameters (Richardson *et al.*, 1995; Tyack *et al.*, 2004). Although sound propagation models may be helpful in this context, animal movements in the sound field generated by directional air-gun arrays with complex radiation patterns can, together with

[a)]Electronic mail: peter.madsen@biology.au.dk

site-specific sound propagation conditions, lead to uncertainties in forward modeling of the signals received by the whale. To overcome this problem, multisensor, acoustic recording tags (e.g., Dtags; Johnson and Tyack, 2003) have been developed to record sound and behavior from the exposed animal, and these have recently proven successful in logging the three-dimensional movements of tagged whales along with recordings of animal sounds and ambient noise (Johnson and Tyack, 2003; Nowacek *et al.*, 2004).

B. Quantification of air-gun pulses for geophysical exploration

In seismic profiling, good signal to noise ratios in echoes returning from geological features below the seafloor are achieved on the transmission side by generating high sound pressure levels (Dragoset, 1984, 1990). The back-calculated source level of an air-gun array is proportional to the firing pressure and the number of guns, whereas it only increases by the cube root of the gun volume (Caldwell and Dragoset, 2000). For this reason, arrays with 30 or more guns are common. The back-calculated source level is a number of convenience for rating the sound output of an array: at close quarters the array does not act as a point source and the highest actual sound pressure levels will be considerably smaller than the back-calculated source levels (Caldwell and Dragoset, 2000). Due to the directional nature of the array, most of the sound energy will be directed toward the seabed to serve the purpose of seismic profiling. Radiated levels in the horizontal plane will be at least 20 dB lower than the back-calculated on-axis level (Barger and Hamblen, 1980; Dragoset, 1990). The industry standard acoustic unit for source sound pressure back-calculated from a known range in the far field, where the array can be viewed as a point source, is the *bar-m*, which describes the peak pressure in a specified frequency band (Dragoset, 1990). A rating in *bar -m* (i.e., the received peak or peak-to-peak sound pressure in bars times the range of the receiver in meters from the source) can readily be converted to a source level expressed in dB re. 1 μPa (0-p) or (p-p) by adding 220 dB to $20 \log_{10}$(*bar-m*) (Richardson *et al.*, 1995). The idealized signature of an air-gun pulse measured on the acoustic axis below the shallow (5–10 m deep) array is a single cycle transient with a duration on the order of 20 m followed by much weaker bubble pulses (Dragoset, 2000). The pulse is broadband with a peak-frequency around 50 Hz, and is normally characterized in a frequency band up to between 125 and 1000 Hz (Dragoset, 1990; Gausland, 2002). The distribution of power or energy in the pulse as a function of frequency (the acoustic spectral signature of the array) is typically given per 1 Hz band as power flux spectral density (conventionally described in dB re. 1 μPa^2/Hz) or energy flux spectral density (dB re. 1 J/m^2/Hz) (Fricke *et al.*, 1985). Thus, industry standards rate arrays in terms of their theoretical back-calculated, band limited on-axis signature.

C. Marine mammal hearing and sensation

Understanding the effects of air-gun sounds on an animal species requires the determination of exposure thresholds at which physiological effects and behavioral responses are elicited. Peak or peak-peak pressure units characterizing the magnitude of an acoustic signal are merely a description of the instantaneous sound pressure. While these may be useful measures from the perspective of seismic profiling, they are not meaningful as stand-alone measures of how sound is processed by an animal from a detection or sensation point of view. Most biological receivers, the mammalian ear included, are best modeled as energy detectors, integrating intensity over a frequency-dependent time window of around 200 ms (Green, 1985). Air-gun pulses will usually be received by exposed animals off the axis of the array, and at ranges for which the pulse has a much longer duration (Greene and Richardson, 1988; Madsen *et al.*, 2002) and a different frequency spectrum (Goold and Fish, 1998) than the on-axis signature described in the previous section. It is therefore important when considering the potential impact of the sound on an animal not to use the on-axis signature of the air-gun pulse but to quantify the air-gun pulses as they are likely to be received by the animals using measures that relate to sensation levels of a biological receiver.

Sound levels radiated off the axis of an air-gun array may be an order of magnitude lower than the peak pressures generated on the acoustic axis (Dragoset, 1990), but, given the high on-axis pressures, the absolute levels of these by-products may still be considerable. So what may be considered a relatively low level horizontal acoustic by-product from an operational perspective in the geophysical exploration industry could have absolute pressure levels that, when weighted by the frequency-sensitivity characteristics of the ear of an exposed animal, could lead potentially to audition-mediated physiological effects, behavioral disruption, or masking.

This study examines the sound exposures received by deep diving toothed whales diving near operating seismic survey vessels in a deep water habitat. Acoustic data were recorded by archival tags on sperm whales (*Physeter macrocephalus*) during controlled exposures to air-gun arrays in the Gulf of Mexico. Results demonstrate that for each firing of the air-gun array, sperm whales receive several versions of the primary pulse that travel on different propagation paths and which have very different temporal and spectral properties. We explore how air-gun pulses can be quantified in a way that might be relevant to sperm whale sensation levels and to their potential for interfering with sperm whale acoustic activities, and we discuss analytical problems associated with the derivation of such measures. It is demonstrated that the received levels measured from sperm whales diving up and down in the water column at variable ranges from the array cannot be predicted by simple geometric spreading laws. We show that some air-gun pulse components carry significant energy at frequencies octaves above the frequency range generally modeled by geophysicists and discuss the implications for high frequency impacts of air-gun pulses on sperm whales and other toothed whale species.

TABLE I. Tag on and tag off times are local time. "*CEE dur*" gives the duration of the CEE in minutes. "*Analyzed*" denotes the number of first/second pulses analyzed.

Tag	Date	Tag on	Tag off	CEE start	CEE stop	CEE dur (min)	Analyzed
sw02_253a	10/9/02	16:38	20:58	17:59	19:15	104	112/74
sw02_254a	11/9/02	10:13	21.45	12:16	14:20	124[a]	55/13
sw02_254b	11/9/02	10:28	22.52	12:16	14:20	124[a]	14/8
sw02_254c	11/9/02	10:34	22.56	12:16	14:20	124[a]	39/22
sw03_164a	13/6/03	9:48	23.20	18:26	19:26	60	82/34
sw03_165a	14/6/03	13:35	06:19	17:01	19:01	120	150/79
sw03_165b	14/6/03	13:38	06:05	17:01	19:01	120	175/82
sw03_173b	22/6/03	14:46	20:38	17:23	19:23	120	383/379

[a]This CEE was paused for 19 min while dolphins passed close to the source vessel.

II. MATERIALS AND METHODS

A. Habitat and logistics

In 2002, experiments were performed from 19 August through 15 September in the Gulf of Mexico as a part of the 2002 SWSS (Sperm Whale Seismic Study) cruise. Visual and acoustic tracking was performed from the RV Gyre while the MV Rylan T., which was physically carrying a coastal survey vessel the MV Speculator, acted as the source vessel. Whales were located and tracked acoustically off the continental shelf of the northwestern Gulf of Mexico by means of a towed hydrophone array. While surfacing, the whales were tracked by visual observers with 25 magnification big-eye binoculars. In 2003, experiments were performed from 3 to 24 June 2003 in the Gulf of Mexico as a part of the 2003 SWSS cruise. In this year, the RV Maurice Ewing, operated by the Lamont-Doherty Earth Observatory, was the platform for acoustic and visual tracking and the MV Kondor Explorer was the source vessel. Procedures for localization and tracking of the whales in 2003 were the same as in 2002.

B. Controlled exposure procedure and air-gun arrays

After tagging, tagged whales were tracked acoustically and visually for at least 1–2 h before controlled exposures were initiated. The source vessel was initially positioned several kilometers from the tagged whale to ensure low initial received levels at the whale. At the beginning of the controlled exposure experiment (CEE), increasing numbers of the guns in the array were fired in a gradual ramp up following the regular procedure used by industry in an attempt to reduce the risk of a high level exposure to undetected nearby whales. The ramp-up procedure entailed starting with a single air gun, and then doubling the number of air guns firing every 5 min. The CEEs lasted between 1 and 2 h leaving the rest of the tag recording time for postexposure data logging.

In 2002, MV Rylan T. towed a small 20 gun array with 2000 psi (pounds per square inch) firing pressure and a volume of 1680 in.^{3} The far-field, vertical signature of the array had a back-calculated, wide-band (3–800 Hz) zero-to-peak SL of 41.1 *bar-m*, corresponding to 252 dB re. 1 μPa (0-peak). The array was fired every 15 s with a 30 min ramp up from 1 to 20 guns. In 2003, MV Kondor towed a larger 31 (28 in use) gun array with 2000 psi firing pressure and a volume of 2590 in.^{3} The far-field, vertical signature of the array had a back-calculated, zero-to-peak source level of 56.9 *bar-m* in the band 3–218 Hz, corresponding to 255 dB re. 1 μPa (0-peak). The array was fired every 15 s with a 30 min ramp up from 1 to 28 guns. In both years, a mitigation protocol was adopted to ensure that no animal sighted or detected acoustically in the study area was exposed to levels higher than 160 dB re 1 μPa (rms) stipulated by the federal permits under which the experiments were carried out (NMFS research permits 369-1440-01, 981-1578, and 981-1707 afforded to P. T.). Acoustic and visual watches for cetacean and turtles were performed from the seismic vessels from at least 1 h prior to the ramp-up and during the seismic emissions, with instructions to stop firing of the air guns in case of any encounter at $<$2 km range. On that basis, one CEE was paused for 19 min while two other CEEs were shortened due to lack of sufficient daylight to observe the mitigation zone (Table I).

C. Dtag specifications and deployment

A noninvasive, archival Dtag (Johnson and Tyack, 2003) was used to gather data on three-dimensional movements and sounds impinging on, or produced by, the tagged whale. Movements of the tagged whales were logged by a depth sensor and 3-axis magnetometers and accelerometers sampled at 47 Hz (2002) or 50 Hz (2003). In 2002, acoustic data were sampled at 32 kHz with a 12 bit ADC. A one-pole high pass filter (HP) at 400 Hz (−3 dB cut off) reduced flow noise and a four-pole Butterworth low-pass filter at 12 kHz countered aliasing problems. Saturation of the recorder occurred at received levels of 152 dB re. 1 μPa (0-peak). In 2003 a second version of the Dtag was used for three of the four whales tested. This tag version sampled sound with 16 bit resolution at 96 kHz again with a 400 Hz one-pole HP filter, and a saturation level of 193 dB re. 1 μPa (0-peak). This tag used a sigma-delta analog-to-digital converter with built-in anti-alias filtering and a flat (±1 dB) frequency response up to 45 kHz. The 400 Hz HP filter in both tags was corrected in postprocessing with a compensating filter yielding a well-characterized frequency response flat within ±1 dB from 0.045 to 12 or 45 kHz. Sperm whales selected for tagging were approached with a rigid hulled inflatable boat while logging at the surface. Tags were brought close to the whales with a 15 m pole cantilever mounted to the bow of the boat and were attached temporarily to the dorsal sur-

TABLE II. Columns "*Whale depth* (m)" and "*Vessel range* (km)" are the minimum-maximum values for which arriving pulses were analyzed. "*First pulse*" numbers give the received *m*-weighted levels for all first arriving pulses analyzed and "*Second pulse*" numbers give *m*-weighted received levels for second arriving pulses. "pp" means peak-peak sound pressure (dB re. 1 μPa, pp), rms is the root-mean-square sound pressure (dB re. 1 μPa, rms) and SEL is the sound exposure level (dB re. 1 μPa^{2}s).

			First pulse			Second pulse		
Tag	Whale depth	Vessel range	pp	rms	SEL	pp	rms	SEL
sw02_253a	8–658	8.4–12.6	142–162	120–144	106–127	146–159	130–146	118–129
sw02_254a	15–614	6.5–9	136–155	121–140	105–123	135–158	116–143	102–126
sw02_254b	6–611	5.7–9.5	136–152	121–135	108–118	145–158	131–142	113–128
sw02_254c	18–605	5–8.4	139–155	125–139	106–123	141–162	125–143	111–126
sw03_164a	20–500	11–12	140–157	125–146	112–129	141–164	125–140	112–124
sw03_165a	na	na	137–160	123–146	106–130	138–154	123–141	110–125
sw03_165b	na	na	135–160	119–147	105–130	135–151	119–137	104–123
sw03_173b	0–17	1.4–7.4	131–162[a]	111–147[a]	94–131[a]	131–153	114–135	104–125

[a]Some pulses were clipped in this recording.

face with suction cups. After a preprogrammed recording time, the tags released from the animals and floated to the surface for recovery by means of an attached VHF transmitter.

D. Data and analysis

1. Distance from sound source

In the 2002 and 2003 SWSS cruises, eight whales were tagged long enough under suitable conditions to perform controlled exposure experiments. The tag-on times and durations of the CEEs are summarized in Table I. In each cruise, an observation vessel, independent of the seismic source vessel, was maintained within about 2 km of the tagged whale. Whale surfacing locations were recorded by observers on this vessel whenever possible, and the whale position between sightings was later estimated using dead-reckoning (Johnson and Tyack, 2003; Zimmer *et al.*, 2005). The dead-reckoned track was computed as the time integral of an estimated velocity vector for the whale based on its orientation as a function of time, recorded by the tag, and an assumed constant swimming speed. The swim speed and a constant advection due to a net current were selected so that the dead-reckoned track would best match the visual observations during initial and final surfacings. A sample of the dead-reckoned tracks was checked against visually fixed locations for accuracy, with mean discrepancy of 370 m ± 223 m (95% CI, $N=16$ fixes). The error in the source-to-whale range estimates will only be as great as the location error when this is along the source-to-whale axis, which generally is unlikely as visual tracking was conducted from the independent observation vessel. Therefore, we conservatively consider ranges reported here to be accurate within ±0.5 km (Table II). Visual tracking was poor for two of the tested whales (sw165a and sw165b), so their position information is not included in this study. Range and depth intervals of the exposed whales are summarized in Table II along with the received sound levels (m-weighted, see section c).

2. Data offload and seismic pulse extraction

Data were offloaded from the tag via a high speed infrared port, and analyzed with custom programs in MATLAB 6.0 (*Mathworks*). Spectrogram and wave form representations of all pulses were inspected visually, and time cues for each pulse arrival were stored along with time cues for a nearby background noise segment. Seismic pulses for which sperm whale clicks occurred within a window of less than 100 ms were omitted. For each seismic pulse, three sound segments were extracted corresponding to the two maximum arrivals of the seismic pulse and a nearby noise sample. The stored window sizes were 100 ms for the first arrivals and 200 ms for second arrivals due to the longer time dispersion of the latter pulses perhaps due to reverberation in the sea floor. Some pulses had decaying tails that extended beyond the time windows, but the use of larger window sizes increased the rejection rates due to overlap with sperm whale clicks. Since the energy in the last part of the decaying tails is relatively small, we chose the indicated window sizes as a compromise between underestimating energy and rejecting too many pulses. To ensure reliable estimation of received levels, we introduced a second criterion for analysis based on signal-to-noise (SNR) ratio: pulses were only accepted for analysis if the SNR was greater than 10 dB (*sensu* McCauley *et al.*, 2000). SNR was calculated as the ratio of the broad band root-mean-square (rms) levels of each pulse and of the noise segment preceding the pulse.

3. Received level measures

Regulations for exposure of sounds to cetaceans currently specify acceptable received levels (RL) in terms of rms sound pressure (NMFS, 2003). For transients, this measure introduces the uncertainty of how to define the time window over which the squared pressure should be averaged, and it is poorly suited for predicting the level of impact of transients with high peak pressure or of long transients with high energy flux density (Finneran *et al.*, 2002; Madsen, 2005). For that reason, we have quantified the seismic pulses by three measures: peak-peak (RL_{pp}, dB re. 1 μPa, pp), rms

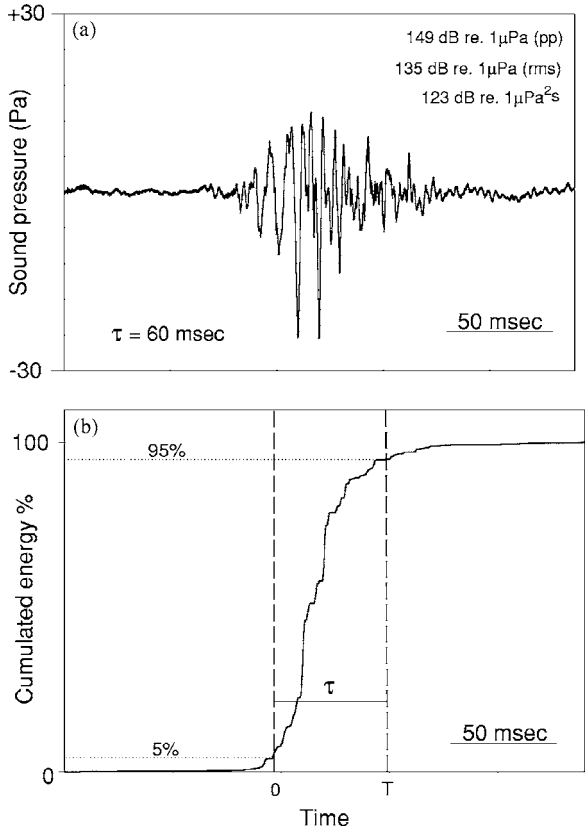


FIG. 1. (a) Wave form of the first pulse arrival from a firing airgun array showing the relationship between pp, rms, and SEL. (b) Relative cumulative energy as a function of time for the wave form shown in (a). The duration (τ) of the pulse in (a) is given by the time window containing 90% of the total relative energy of the window analyzed. τ is used as averaging time for derivation of the square root of the mean pressure-squared (rms) of the pulse and as integration time for computation of the sound exposure level (SEL).

(RL_{rms}, dB re. 1 μPa, rms) and sound exposure level (SEL, dB re. 1 μPa^2 s) (proportional to energy flux density for a plane wave propagating in an unbounded medium). This will facilitate comparison with other studies and provide measures that take into account both the peak pressure (RL_{pp}) and the sound exposure level (SEL) of the sound pulse (McCauley *et al.*, 2003). All analyses are based on the assumption of individual pressure measurements of a plane wave propagating in the far-field of the sound source.

For calculation of rms levels of a transient signal, we have adopted the 90% energy approach used by McCauley *et al.* (2000) and Blackwell *et al.* (2004). The relative energy is computed in a window around the seismic pulse [Fig. 1(a)], and the duration (τ in seconds) is defined by the smallest sample interval ($0:T$) in the analysis window containing 90% of the energy in the window [Fig. 1(b)]. This duration defines the sample interval over which the root-mean-square pressure level (RL_{rms}) is computed:

$$10\log\left(\frac{1}{T}\int_0^T p^2(t)dt\right) \quad [p(t) = \text{instantaneous pressure}]. \tag{1}$$

The SEL is given by the square of the instantaneous pressure integrated over the pulse duration T:

$$10\log\left(\int_0^T p^2(t)dt\right) \quad [p(t) = \text{instantaneous pressure}]. \tag{2}$$

Consequently, the SEL is given by rms sound pressure level (RL_{rms}) + 10 log(τ). To avoid small errors in the exposure measures due to ambient noise in the analysis window, the noise power of the preceding 100 ms noise window was subtracted from $p^2(t)$ when computing RL_{rms} and SEL.

Since the ear of most mammals, dolphins included, integrates low frequency sound over a window of around 200 ms (Johnson, 1968a; Au *et al.*, 2002), this duration was used as the maximum integration time for RL_{rms} and SEL (*sensu* Madsen *et al.*, 2002). For sound exposures high enough to generate temporary or permanent threshold shifts, the 200 ms integration window does not apply, and the entire duration of the exposure should be taken into account (Finneran *et al.*, 2002; Nachtigall *et al.*, 2004). However, the transient exposures in the present study are highly unlikely to cause TTS (Finneran *et al.*, 2002), so we feel that the use of a 200 ms maximum integration time is justified from a sensation perspective, especially considering that there is very little energy outside of this window in the pulses analyzed.

4. Frequency-weighting to approximate the frequency response of sperm whale auditory system

Seismic pulses are designed to have peak energy around 50 Hz, but there is significant energy both lower, and as will be shown, at significantly higher frequencies. The spectrally corrected Dtag recordings have a flat frequency response down to about 45 Hz, but at lower frequencies flow noise around the tag dominates the recording and received levels cannot be accurately determined. This leads to an underestimation of the low frequency energy in the pulses and thereby in the broad band sound pressure and sound exposure levels. However we argue that such low frequency components may have little relevance to sperm whales which seem to hear best at higher frequencies (Ridgway and Carder, 2001).

Mammalian auditory systems have differential spectral sensitivity with a gently sloping decrease in sensitivity toward low frequencies and a sharp cutoff in sensitivity at high frequencies (Fay and Popper, 1994). For humans it is common practice to apply spectral corrections when calculating noise exposures. The so-called A-weighting mimics the frequency dependence of sensation at moderate exposure levels while the flatter C-weighting is appropriate for transients at levels that might lead to damaging sound exposures (ANSI 1994; Harris, 1997). There is no reason to believe that marine mammals are different from humans in this respect (Finneran *et al.*, 2002) but it is a major challenge to determine suitable weighting functions for the diverse cetacean groups. Based on anatomy of the inner ear (Fleischer, 1976; Ketten, 1997, 2000), available audiograms (Au *et al.*, 1997) and the frequency ranges of vocalizations, it is believed [Richardson *et al.*, 1995; Southall *et al.* (unpublished)] that the auditory systems of toothed whales are less sensitive to low-frequency noise than those of baleen whales, who use low frequency sound for communication. In that light, different

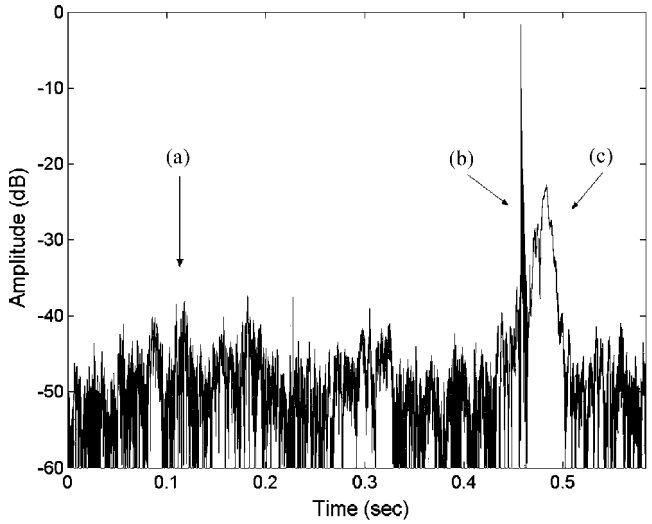


FIG. 2. Amplitude plot of the envelope (absolute value of the Hilbert transform) of the received wave form of a click and air gun pulses on a sperm whale. (a) Points to the first pulse arrival from the firing of an airgun array that is buried in noise. (b) Points to a usual click produced by the tagged whale that overlaps in time with the second pulse arrival from the firing airgun array (c). None of the pulses can be analyzed because of bad SNR (a) and overlap (c) with the sperm whale click (b).

weighting functions, called M-weighting akin to the C-weighting for human auditory systems, have been developed for different marine mammalian groups [Southall *et al.* (unpublished)].

Sperm whales have been included in the group of midfrequency odontocetes (assuming that the effective hearing range is >150 Hz) that also includes most delphinids (Ketten, 1997). While it seems highly unlikely that the hearing curve of a sperm whale (Ridgway and Carder, 2001) equals that of a dolphin (Johnson, 1968b), we will use the M-weighting function for impact of transients on midfrequency odontocetes in the present study to avoid additional confusion, accepting that we may thereby underestimate the sensation levels of the pulses impinging on the sperm whales. The weighting function was implemented in the time-domain as a filtering operation prior to exposure calculation. The filter characteristic was realized by a two pole 200 Hz HP filter with a Q factor of unity [sensu. Southall *et al.* (unpublished)].

III. RESULTS AND DISCUSSION

A. Analysis of seismic pulses recorded with onboard tags

The advantage of using onboard, calibrated sound recording tags to make exposure measures is that the sound field quantified is very close to what is received by the whale. However, quantifying low frequency sounds with sound recording tags attached to a moving animal poses the difficulty that other recorded sounds interfere with and can mask the signals of interest. The two major sources of interference are flow noise generated by water movements around the tag and sperm whale clicks that overlap with the analysis windows of the seismic pulses. Figure 2 presents an example of such interference rendering analysis futile. The first arrival (a) is low enough in amplitude to be masked by the flow noise, and the second arrival (c) overlaps in time with a click from the tagged whale (b). Sperm whales produce usual clicks at interclick intervals of 0.4–1 s with click durations, as recorded by a tag attached to the body, of up to 50 ms. The interference duty cycle of some 5% from the tagged whale's own clicks, in concert with occasional high amplitude incoming clicks from other sperm whales, leads to the rejection of a significant number of air-gun pulses that otherwise have sufficient SNR for analysis.

The design and placement of the tag probably give rise to more flow noise around the recording hydrophone than the moving whale hears. The flow noise levels vary significantly over time depending on the activity of the whale and position of the tag. To avoid saturation of the recorder by low frequency flow noise, the Dtags have a built-in one-pole prewhitening filter ($f_0=400$ Hz), but this high-pass filtering attenuates low frequency sounds of interest below 400 Hz in the same way as it does the flow noise. Likewise, the postemphasizing filter used to flatten the spectral response of the tag down to 45 Hz does not improve the broadband SNR as all low frequency noise is amplified equally. As a result, a number of seismic pulses with low received levels compared to the flow noise could not be analyzed.

A primary goal of the SWSS CEE studies [Miller *et al.*, (unpublished)] was to study the effects of seismic pulses on sperm whales with a target range of received levels from 120 to 160 dB re. 1 μPa (rms). Given the largely unknown radiation pattern of the seismic array and the uncertainties in acoustic propagation, it was necessary to take a conservative approach in positioning the seismic vessel. An additional factor was the frequent presence of dispersed groups of sperm whales near the tagged whale which sometimes prevented close approaches to the tagged whale without the risk of exceeding permitted levels of exposure to other whales. Accordingly, none of the pulses received by tagged sperm whales exceeded the 160 dB re. 1 μPa (rms) limit (maximum of 147 dB re. 1 μPa (rms)) during the CEEs (Table II). The cautious exposure approach in combination with the complex acoustic propagation conditions in the Gulf of Mexico also meant that the received levels of many of the seismic pulses impinging on the whales were low enough to be masked by the flow noise around the tags, rendering analysis impossible. Table I gives the number of pulses analyzed and it is evident that some tag recordings had a high rejection rate due to overlap with sperm whale clicks and poor SNR. This does not necessarily imply that the whales were exposed to a lot of pulses at insignificant levels, only that we were unable to quantify the exposure in these cases.

With the changing levels of flow noise between and within tag attachments, the lowest received levels of pulses that warranted analysis varied between 111 and 125 dB re 1 μPa (rms) (Table II). The quantitative properties of the weakest receptions are therefore under-represented in our assessment of the total exposure. An exception from that caveat is whale sw03_173b, which rested near the surface without clicking or moving much for the entire CEE. During this CEE almost all pulses could be analyzed except for those received when the tag was out of the water during surfacings.

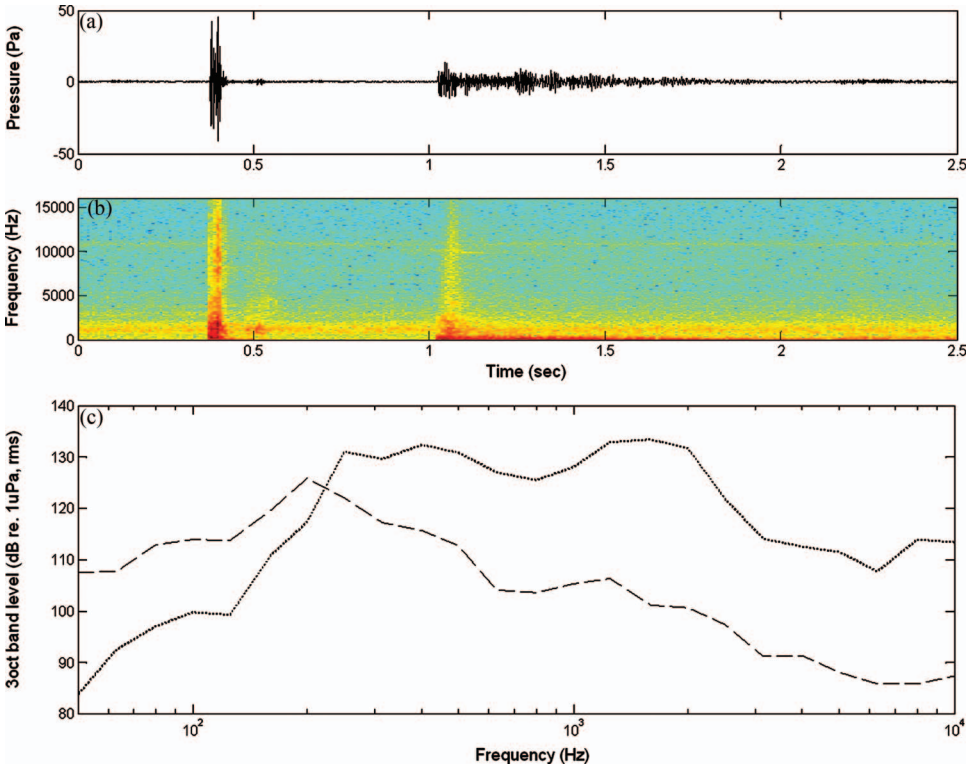


FIG. 3. (a) Wave form of an airgun exposure of whale sw03_173b at a range of 3 km and a depth of 15 m. (b) Spectrogram of the wave form in (a) (FFT=1024, 50% overlap). Note how the first pulse arrival has energy all the way up to the Nyquist frequency at 16 kHz, whereas the second arrival has little energy above 1 kHz. (c) 1/3 octave rms sound pressure levels of the two pulses displayed in (a) and (b). First arrival has a dotted line and the second arrival has a dashed line. The rms levels for these pulses can be converted to 1/3 octave SELs (dB re 1 μPa^2 s) by subtracting 13 dB (54 ms duration) for the rms levels of the first pulse and 7 dB (200 ms max integration time) for the rms levels of the second pulse.

Some of the pulses in this exposure actually clipped the tag. In general, however, it is not meaningful to provide an average exposure level for a CEE or to argue that the range of received levels that could be calculated is representative of the overall acoustic exposure, since pulses of low amplitude are rejected more often than are high level receptions. Nonetheless, the absolute levels that we are able to provide are indisputably close to those received by the whale at a given depth and range from the array and so provide a point characterization of the exposure. It should be recalled that the actual broadband exposure will in some cases be significantly underestimated compared to the true broadband received levels due to the exclusion of energy below 45 Hz and the M-weighting filter that starts at 200 Hz.

A second potential way in which received sound levels may be underestimated relates to the effect of body shading. The mismatch in acoustic impedance between seawater and sperm whale tissue, bone and airways in particular, means that the sound levels at the tag will be attenuated whenever the body of the whale is placed between the source and the tag. This effect will be most severe at high frequencies for which the wavelengths are small compared to the size of sperm whale body parts (Medwin and Clay, 1998). Body shading will therefore lead to a frequency-dependent underestimation of some received levels. The constantly changing geometry between the moving source vessel and the diving whale renders body shading effects inseparable from the effects of changing propagation conditions, but body shading is likely to be range independent and will if anything only lead to an underestimation of the received levels. Resonances in air volumes held within a diving animal might also influence the spectral emphasis of pulses recorded on the animals, but we have no means of testing such a conjecture or of assessing if such resonances would also effect the sound heard by the whale. Nonetheless, this effect can be considered second order compared to changes in whale position and multipath propagation.

B. Received levels and the effects of depth and range

Air-gun arrays are designed to produce a single downward-directed impulse that propagates through the water column and into the seabed. Unavoidably, some sound energy also radiates horizontally from the array creating a complex radiation pattern. The presence of multiple propagation paths involving surface and bottom bounces as well as the re-radiation of sound reverberating within subbottom layers increases the complexity of the received signal and can give rise to long reverberant wave forms of several seconds at long ranges (Greene and Richardson, 1988; Madsen *et al.*, 2002). At the shorter ranges of interest here (<13 km), the question is whether sperm whales are effectively exposed to a single impulse with properties akin to those of the on-axis pulse from the array or if a more complex multipulsed exposure is occurring [DeRuiter *et al.* (unpublished)].

To answer this question with an example, Fig. 3(a) shows the wave forms of pulses received by a sperm whale at a range of 3 km and at a depth of 15 m. The first arrival consists of a short, well-defined transient followed some 500 ms later by a reverberant second arrival with a long decaying tail. This pattern of multiple arrivals was observed in all exposures that could be analyzed and it is evident that, with each firing of the array, the whale may receive several pulses with differing temporal and spectral properties. An immediate consequence is that acoustic exposure of animals by air-gun arrays should not be modeled by a single well-defined pulse arrival for each firing of the array. Detailed modeling of the acoustic propagation that leads to this arrival

pattern is beyond the scope of the present paper and is treated in detail by DeRuiter *et al.* (unpublished). Here we focus only on the two strongest arrivals, since the remaining pulses carry relatively little energy.

The CEE involving whale sw03_173b which rested near the surface throughout the exposure provides an opportunity to examine the relationship between range and RL at shallow depths. Figure 4 displays how the received levels in the first and second arrivals changed over time as the source vessel approached the whale and moved away from it again. The first arrivals actually saturated the recorder during the closest approach at a range of 1–2 km and the levels in that period of time are therefore underestimated (Fig. 4). The first arrivals have short durations from 15 to 30 ms whereas the second arrivals have durations up to and beyond the 200 ms maximum analysis window due to multipath spreading. The second arrivals therefore have higher SELs and lower rms levels for a given peak pressure due to the longer integration (SEL) and averaging times (rms) [Figs. 3(a) and 4]. While the levels of both arrivals increase with reducing range to the source vessel, the variation in received levels of the first arrivals is much larger than for the second which, given their time delay with respect to the first arrivals, are likely bottom reflections. While the received levels are highest for the second arrivals at the beginning of the CEE when the source was about 5 km from the whale, the received levels of the first arrival quickly dominate in terms of p-p and rms pressure as the source approached. There is less variation over the course of the exposure when the received levels are quantified as SEL because the longer durations of the weaker second arrivals tend to compensate for their reduced peak pressure (Fig. 5).

Since almost all of the pulses received by whale sw03_173b could be analyzed, the entire exposure history of this animal throughout the CEE could be estimated. We use SEL as the measure of acoustic exposure. The cumulated SEL experienced by sw03_173b as a function of time is displayed in Fig. 5. It is seen that the second pulse arrivals dominate the exposure during the first 15 min, and that the first pulse arrivals contribute little to the overall exposure in this time interval. As the source vessel approaches, the SEL of the first arriving pulse increases rapidly and the first pulse arrivals become the determining factor for the overall exposure. The implication is that the second pulse arrivals may be more important for a near-surface whale that is distant from the source whereas direct arrivals will dominate at shorter ranges but both pulse arrivals must be considered to avoid underestimation of the combined acoustic exposure.

We argue that the overall acoustic exposure, calculated as in Fig. 5, should be considered as an exposure metric along with the maximum received sound pressure levels and SELs of individual pulses, when assessing the impact of transient noise sources on marine mammals. However, cumulated energy cannot serve as a stand alone measure for mitigation since an overall exposure of 151 dB re. 1 μPa^2 s like the one displayed in Fig. 5 could be achieved with a single 200 ms tone with a rms sound pressure level of 158 dB re. 1 μPa (rms) [$151=158+10\log(0.2\ \text{s})$] or with a single or a few ultrashort transients of very high sound pres-

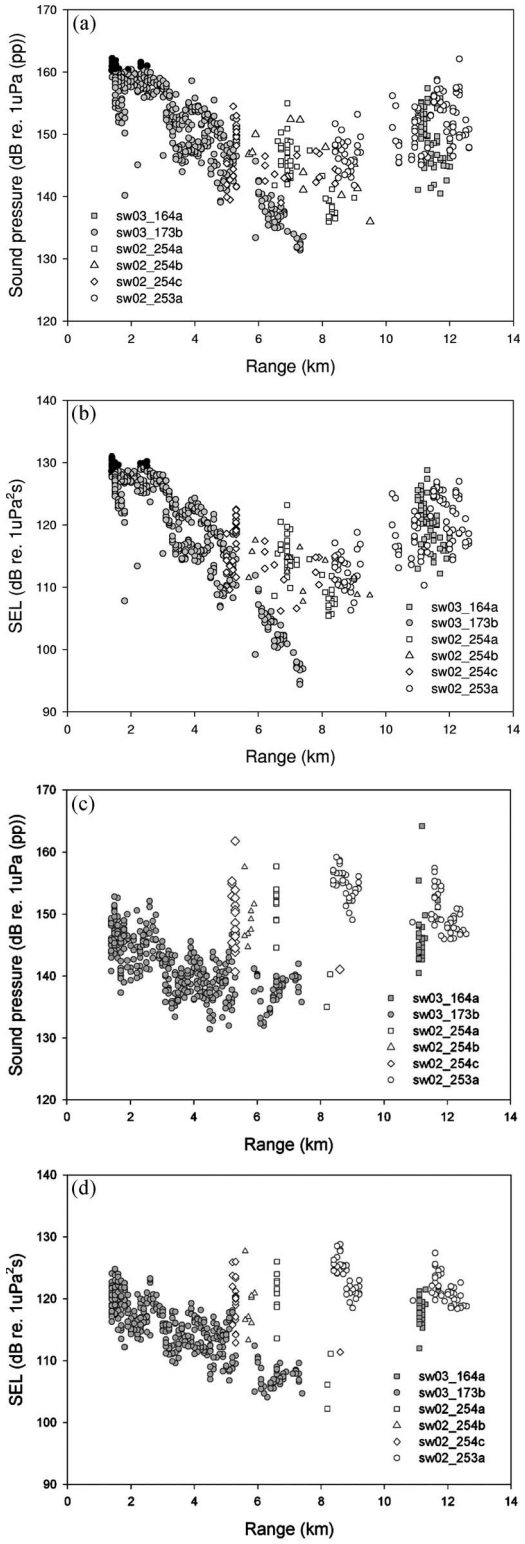


FIG. 4. (a) Received peak-peak sound pressure levels of the first arrival for each airgun pulse that could be analyzed as a function of range from all CEEs where range to the whale could be derived. The highest levels closest to the source were clipped (closed dark circles). The data are from six different whales during two seasons using two different seismic arrays. Note how the received levels reach a minimum between 5 and 9 km, after which the received levels increase again with range. (b) sound exposure levels (SEL, dB re. 1 μPa^2 s) for the same pulses as displayed in (a). (c) Received peak-peak sound pressure levels of the second arrival for each airgun pulse that could be analyzed as a function of range from all CEEs where range to the whale could be derived. Note how the received levels of this pulse component actually increase with range beyond 5 km. (d) Sound exposure levels (SEL, dB re. 1 μPa^2 s) for the same pulses as displayed in (c).

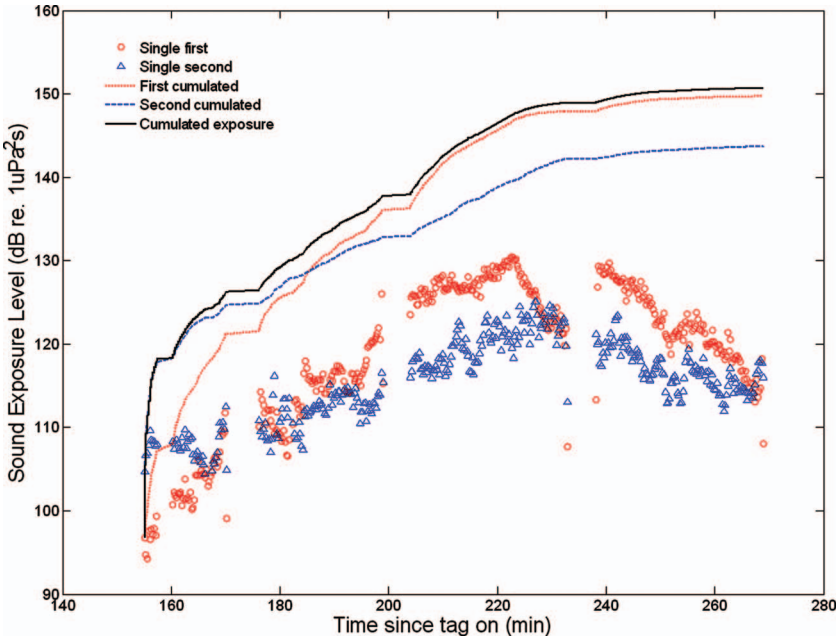


FIG. 5. Acoustic exposure of whale sw03_173b during the entire CEE with the exception of pulses impinging on the whale when surfaced. Circles denote received sound exposure levels (SEL, dB re. 1 μPa^2 s) of first arriving pulses as a function of time (min). Triangles denote received SELs of second arrivals as a function of time. Dotted line shows cumulative SEL of the first arrivals and the dashed line shows cumulative SEL of the second arrivals over course of time. The solid line shows the overall accumulating exposure as a function of time. The full acoustic M-weighted exposure during this CEE amounted to 150.7 dB re. 1 μPa^2 s. Some pulses received at the shortest distance between the array and the whale were clipped, leading to an underestimation of the overall sound exposure level.

sures above 200 dB re 1 μPa (pp) (or any combination in between). While the former may only lead to short term behavioral disruption, the high level transient exposure could perhaps induce stronger effects. Similarly, a few sound pulses may only evoke little or no behavioral response, whereas a long sequence of the same pulses could have negative effects through sensitization (Richardson *et al.*, 1995).

In contrast to sw03_173b, most of the exposed whales continued to perform foraging dives throughout the exposure. Data from these animals provide an indication of the way that the sound exposure varies with distance over the normal range of depths traversed by sperm whales. Figure 4 plots all received levels for whales for which the range to the array could be determined. It is evident that there is no simple relationship between received levels of the first pulse arrivals and the range to the seismic array no matter whether RLpp (4a) or SEL (4b) are considered. Rather, the received levels fall to a minimum between 5 and 9 km and then start increasing again at ranges between 9 and 13 km. It must be emphasized that these received levels as a function of range are generated from six different whales during two field seasons with different seismic arrays. Nevertheless, the pattern is consistent across both seasons, and within individual experiments. It must be concluded that the received level of first pulse arrivals can be just as high (160 dB re. 1 μPa, pp) at 12 km as at a range of 2 km from the array. When looking at the secondary arrivals [Figs. 4(c) and 4(d)], it is seen that they have higher received levels at 5–12.6 km than at ranges closer to the seismic sources. It is therefore clear that sperm whale exposure to different pulse components at ranges from 1 to 13 km from the seismic sources does not necessarily attenuate with increasing range. Rather, both received sound pressures and SELs may actually increase if the whales move from say 7 to 12 km (Fig. 4). Similar results have also been reported from recordings of air-gun arrays operating in the Gulf of Mexico and off California in 1995, where shallow hydrophones (10–100 m) received sound pressure levels with several range-dependent local maxima up to 170 dB re 1 μPa (pp) from 3 to 10 km (Lepage *et al.*, 1995; 1996).

These exposure patterns emerge because the received levels from different pulse components in this range interval and in this location do not conform to simple geometric spreading laws such as 20 log(Range) (Urick, 1983). In fact whales diving in a stratified water column at variable ranges are exposed to a much more complicated sound field due to multipath propagation [Lepage *et al.*, 1995; 1996; DeRuiter *et al.* (unpublished)].

Acoustic shadow and convergence zones are generated by downward refracting sound speed profiles such as are found in the summer months in the Gulf of Mexico. Since such situations are common, and this type of profile results in distinct, robust "shadow zone-convergence zone" characteristics of the acoustic waveguide, it can be useful to make some general approximations for the extent of these zones, both for scientific and regulatory purposes. We derive such rules here, using standard results from ocean acoustics. See DeRuiter *et al.* (unpublished) for in-depth analysis and modeling of the sound propagation leading to the observed exposures.

Consider the simplified geometric situation depicted in Fig. 6. A near-surface source transmits sound in a downward

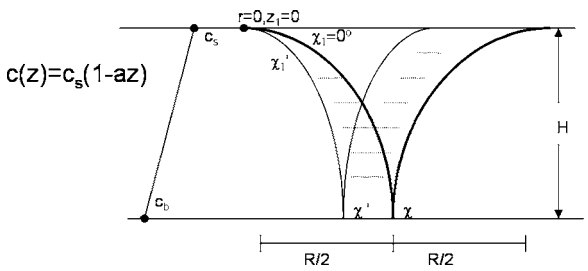


FIG. 6. Geometry for downward refracting rays and their convergence/shadow zones.

refracting waveguide of depth H with a linearly decreasing sound speed profile:

$$c(z) = c_s(1 - az). \tag{3}$$

The ray with a launch angle of zero degrees ($\chi_1 = 0$) with respect to the water surface will travel the farthest before being refracted down toward the bottom, and so defines the upper limit of the ensonified zone surrounding the source. The ray that hits the bottom at the critical grazing angle ($\chi' = \theta_{\text{crit}}^{\text{grazing}}$) will define the nearer limit of the second ensonified zone following the shadow zone, as rays hitting the bottom at steeper angles will transfer most of their energy into the substrate. As depicted in Fig. 6, rays with launch angles between χ_1 and χ_1' define the extent of the subsequent ensonified zone. Here we will focus on the zero degree launch angle ray, since it loses the least amount of energy due to having the lowest bottom grazing angle (bottom loss generally increases with increasing angle of interaction with the seabed).

Following the method of Brekhovskikh and Lysanov (2003), we define r and z as the horizontal range and depth, respectively, and place the source at $r_1 = 0$ and $z_1 = 0$. The reception point, i.e., the whale, is at position r, z. For an infinitesimal segment of the raypath, $dr = |dz/\tan \chi|$ where χ is the local grazing angle of the ray. Integrating this between source and receiver depths gives the relation:

$$r = \left| \int_{z1}^{z} \frac{dz}{\tan \chi} \right|. \tag{4}$$

Applying Snell's law, the grazing angle $\chi(z)$ can be expressed in terms of the launch angle, χ_1, as

$$\cos \chi = (1/n)\cos \chi_1, \tag{5}$$

where $n = n(z) = c_1/c(z)$ is the index of refraction. Using these equations, Eq. (4) can be rewritten:

$$r = \cos \chi_1 \left| \int_{z_1}^{z} [n^2(z) - \cos^2 \chi_1]^{-1/2} dz \right|. \tag{6}$$

Equation (4) can be used to define the ensonified zones by substituting the launch angles $\chi_1 = 0$ for the longest range ray or $\chi_1 = \chi_1^{\text{crit}}$, the solution of Eq. (5) for $\chi = \theta_{\text{crit}}^{\text{grazing}}$, for the shortest ray. For $\chi_1 = 0$ and the linear sound speed profile, $c(z)$, defined in Eq. (3), we get from Eq. (6):

$$r = \left| \int_0^H [(1 - az)^{-2} - 1]^{-1/2} dz \right| \tag{7}$$

For the locations and times of year for which data are reported here, the difference between the surface and bottom sound speed at 800 m depth was 50 m/s, so that $a \approx 0.00004$ m^{-1} from Eq. (3), i.e., a is a very small number. Thus the integrand of Eq. (7) can be well-approximated by a binomial expansion truncated at the linear term. Since $(1 - az)^{-2} \approx 1 + 2az$, we obtain

$$r \approx \left| \int_0^H (2az)^{-1/2} dz \right|, \tag{8}$$

which can be readily evaluated to give the solution for r:

$$r = \sqrt{\frac{2H}{a}}, \tag{9}$$

which is the horizontal distance traversed by the longest traveling ray before reaching the bottom, a distance denoted as $R/2$ in Fig. 6 (R is the total ray cycle distance, so the bottom reflection point is one half of this). For the CEEs examined here, $R/2 \sim 6.3$ km, in reasonable agreement with the received level patterns over two years from different whales (Fig. 4).

An even simpler estimate of the ray half cycle distance can be derived by replacing the grazing angle $\chi(z)$ by a constant angle defined as one half the difference between the launch angle and the bottom grazing angle of impact. Likewise, the sound-speed profile can be replaced by the average sound-speed over the waveguide height in an "isovelocity" approximation. With these approximations, $R/2$ can be estimated with straight-line geometry as

$$R/2 \approx \frac{H}{\tan \theta^{\text{average}}}. \tag{10}$$

Using even this crude estimate gives $R/2 \sim 8.1$ km, which is also in reasonable agreement with our data (Fig. 4). Straightforward ray tracing models therefore seem to be

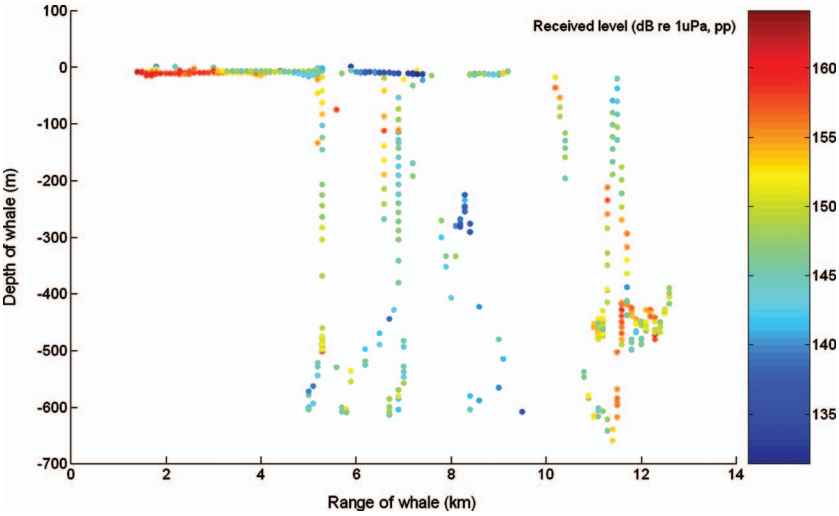


FIG. 7. Received peak-peak sound pressure levels (dB re. 1 μPa, pp) for all pulses that could be analyzed and for which depth and range could be derived. A few pulses close to the source were clipped. Note that the received levels can be as high at 12 km range (450 m depth) as at a range of 2 km. There is an important gap in the data set with no measurements on whales with depth greater than 20 m at ranges shorter than 4 km to the air gun source.

useful for predicting sound exposure levels as a function of receiver depth and range from the source. Moreover, these results appear to be fairly robust between years in the Gulf of Mexico (Fig. 4).

The variability in received levels of both pulse arrivals to the tagged whales as a function of both depth and range is summarized in Fig. 7. It is seen that whales, while ascending or descending during foraging dives, move in and out of the acoustic convergence and shadow zones, where the received sound pressure levels and SELs can change rapidly. It is also clear that at times during the CEE, the whales would have been exposed to much lower levels than those depicted in Fig. 4, which only includes pulses with a SNR of more than 10 dB. The collected data show that none of the sperm whales were exposed to sound levels higher than 162 dB re. 1 μPa (pp) (147 rms and 131 SEL) when diving more than 2 km from the seismic sources (Figs. 4 and 7). These received levels match fairly well with the findings of Tolstoy *et al.* (2004),[2] but the findings of Lepage *et al.* (1996) reporting broadband received level maxima of more than 170 dB re 1 μPa (pp) out to ranges of 7 km emphasize that broadband received levels are likely higher than reported for the M-weighted pulses in the present study. It is important to note that no data were obtained from deep-diving whales within 4 km of the source, as the only whale that was approached closely made shallow dives while the array was firing. Clearly the highest received levels will be experienced within the downward-projected beam of the airgun array where we were unable to collect data.

It is not the intention of this paper to evaluate the possible effects of sound on the behavior of exposed whales [see Miller *et al.*, (unpublished)]. However, if pulses with received levels in the range 140–165 dB re. 1 μPa (pp) (115 to 135 SEL) are found to have negative effects on sperm whales (as seen for bowhead whales (Richardson *et al.*, 1986; Ljungblad *et al.*, 1988), then animals in the Gulf of Mexico could be impacted at ranges of more than 10 km from seismic survey vessels, well beyond the ranges predicted by geometric spreading laws and beyond where visual observers on the source vessel can monitor effectively. We have shown that if whales wish to reduce their exposure then horizontal displacement away from the seismic survey vessel may not be the correct strategy. Rather, in the 5–9 km range, they may, depending on the acoustic propagation conditions, reduce their exposure by moving closer to the array or by vertical rather than horizontal displacement. Such movements, while reducing received levels in the short-term, might end up prolonging the overall exposure time and the accumulated SEL. This observation is of particular relevance when employing ramp-up procedures under the untested assumption that these will lead to horizontal displacement of animals away from the array.

The propagation conditions over ranges of kilometers in a deep water habitat like the Gulf of Mexico are incompatible with the zones of exposure method for rating potential impacts outlined by Richardson *et al.* (1995), which is based on the assumption that received levels decrease with range in a simple fashion with less and less impact on the exposed animals further from the noise source (NRC, 2005). If the received levels measured here in the range from 1 to 13 km have no significant effect on marine life, the effects of multipath propagation can be ignored from an environmental mitigation perspective. However, if received levels in this dB range are impactful as seen for some baleen whale species (Malme *et al.*, 1985; Richardson *et al.*, 1995), we face the challenge of how to mitigate under such conditions, where animals can dive in and out of high exposure levels at considerable ranges from the air-gun array (see also Lepage *et al.*, 1996).

C. Spectral properties and high frequency by-products

Noise transients produced by seismic survey vessels are designed to have maximum sound energy at around 50 Hz (Dragoset, 1990; Barger and Hamblen, 1980) and their impact on toothed whales, which are considered to be less sensitive to low frequency sound (Au *et al.*, 1997), could accordingly be assumed small (NRC, 2005). Dissenting data, reported by Goold and Fish (1998), showed that dolphins can be exposed to noise above ambient levels from air-gun pulses at frequencies of up to 8 kHz and at ranges up to 8 km. Resolution of this issue is important since the impact in terms of masking, physical damage and sensation levels on different cetaceans will relate to the frequency content of the noise pulses in question (Harris, 1997). The present data set provides an opportunity to test these contentions by measuring the absolute band levels of pulses impinging on sperm whales.

The spectral distribution of noise is often defined in terms of power spectral density (dB re. 1 μPa^2/Hz) or energy flux spectral density (dB re. 1 μPa^{2}s/Hz) with both measures using an analysis bandwidth of 1 Hz. In contrast, the mammalian ear integrates energy over much broader bandwidths. It is common practice in describing noise exposure to approximate the way noise is integrated by the auditory system of mammals by measuring the rms noise power in 1/3 octave bands (third octave levels (TOL), dB re. 1 μPa, rms) (Richardson *et al.*, 1995). As an example, Fig. 3(b) presents the spectrogram of a seismic signal [Fig. 3(a)] received on sw03_173b while the whale was resting close to the surface and so well away from the axis of the sound source. The energy in the first arrival extends to the Nyquist cut-off frequency at 15 kHz, whereas the energy in the second arrival is concentrated below 500 Hz. To quantify the frequency distributions of these two arrivals, we performed a TOL analysis on the wave forms correcting for the HP filter of the tag but without the M-weighting. The result displayed in Fig. 3(c) shows that, despite the flat recording response to 45 Hz, the first arrival carries little energy below 300 Hz where the energy of an on-axis airgun pulse would be concentrated (Dragoset, 1990; Caldwell and Dragoset, 2000). Instead the energy is concentrated in the third octaves from 300 Hz to 3 kHz where the TOLs experienced by the whale 3 km from the array are around 130 dB re. 1 μPa rms. In contrast, the energy in the second arrival is concentrated at low frequencies around 200 Hz. This pattern was consistent over the entire CEE involving sw03_173b and was observed with less regularity in all other CEEs.

The observation that sperm whales are exposed to significant levels of high frequency energy from air-gun arrays is consistent with the report of Goold and Fish (1998). This high frequency energy is a by-product of the air-gun sound source and is beyond the frequency range within which air guns are usually characterized. Although its relative contribution to the overall output of the source is likely small (Tolstoy *et al.*, 2004), the relevant parameter for assessing the impact of seismic pulses on marine mammals is the absolute received levels at frequencies where the exposed animals have good hearing sensitivity. While it may not be surprising that a powerful impulsive sound source like an air gun could generate by-products with considerable energy at higher frequencies, this has not been addressed quantitatively before and the observation warrants further measurements and modeling as started by Tolstoy *et al.* (2004).

High SL emission of energy at midfrequencies has the potential to affect animals with apparent less sensitive low frequency hearing such as dolphins and beaked whales which are normally not considered in assessments of impacts from seismic surveying (Tolstoy *et al.*, 2004). Here we document that a whale more than 1400 m from the air-gun array received more than 162 dB re. 1 μPa (pp) (147 rms, 131 SEL) from pulses with essentially no energy below 300 Hz emphasizing that the potential for negative effects of this high frequency by-product on marine mammals should not be dismissed lightly. That predominantly high frequency pulses were received by whales near the surface is consistent with their radiation from grating lobes in the source array beam pattern. The combination of the Lloyd's mirror effect attenuating low frequency energy by destructive interference, and a high-frequency surface duct in the warm stratified summer water of the Gulf of Mexico will lead to high pass filtering of the signal [Deruiter *et al.*, (unpublished)]. All air-breathing mammals are forced to spend significant time near the sea surface to ventilate their lungs between dives. Deep diving marine mammals, such as sperm whales and beaked whales, will enter the high frequency exposure zone from air-gun arrays, when oceanographic conditions support it, whenever returning to the surface to recover from deep dives. Some species, such as pelagic dolphins, will likely be more exposed to the high frequency components, because they spend more time traveling and socializing near the surface.

The presence of significant energy at high frequencies in some air-gun pulses not only implies that the sensation levels are likely higher than previously expected for toothed whales, but also that the potential for masking should be considered. Masking occurs when the noise power is increased in one or more critical bands that overlap in the frequency domain with a signal of interest (Richardson *et al.*, 1995). Some sperm whale click types (e.g., coda, slow, and calf clicks) that are believed to serve a communicative purpose have most of their energy below 5 kHz (Madsen, 2002) and so overlap in frequency with the high frequency energy in some air-gun pulses. While masking of sperm whale communication sounds accordingly could occur, the short duration and low duty cycle of the high frequency air-gun transients renders the masking power very small as compared to comparable continuous noise, for example, from ship traffic (Aguilar *et al.*, 2006) So despite the presence of high frequency energy in some air-gun pulses, the low duty cycle of air-gun noise suggests that the pulses are not likely to pose a significant masking problem for sperm whale acoustic communication or echolocation.

IV. CONCLUSION

Onboard acoustic recording tags have been used to quantify the sound field impinging on sperm whales from air-gun arrays during a series of controlled exposure experiments in the Gulf of Mexico. We have demonstrated that, due to multipath propagation, sperm whales are exposed to several pulses for each firing of the array with very different temporal and spectral properties. Noise exposure estimates should consider these different pulse components and their potential combined impacts. Sperm whales diving at ranges between 4 and 13 km were exposed to pulses with received levels of up to 162 dB re 1 μPa (pp) (127 dB re. 1 μPa^2 s). The relative strength of pulses arriving on different paths vary with range and depth of the diving whales, but the absolute received levels can be as high at 12 km as they are at 2 km. We conclude that simple geometric spreading models cannot be used to establish impact zones when assessing potential effects on marine mammals in a deep water habitat like the Gulf of Mexico. We have also shown that air-gun arrays can generate high absolute levels of sound energy at frequencies octaves higher than that used for seismic profiling. Some pulse components have the bulk of their energy at frequencies above 300 Hz, and the relatively high received levels of such pulses at ranges of kilometers from the operating array is a cause for concern for toothed whales, including smaller species such as dolphins and beaked whales, not normally considered when assessing the impact of seismic surveys on marine life. The current study did not provide exposure measurements for sperm whales diving deep closer than 4 km from the array, and this lack of data should be addressed in further experiments. The different exposures experienced by whales while diving as compared to resting near the surface emphasize that sound exposure as a function of depth and range should not be extrapolated between habitats with varying sound velocity profiles and bottom properties.

ACKNOWLEDGMENTS

We thank the field parties and ships crews of the SWSS cruises for logistical and practical support. A. Bocconcelli, T. Hurst, and K. Shorter were instrumental in tag development and deployment. D. Cato, S. DeRuiter, C. Greene, Y.T. Lin, A.E. Newhall, B.K. Nielsen, M. Wahlberg, and two anonymous reviewers are thanked for helpful discussions and/or constructive critique on earlier versions of the manuscript. We thank A. Hansen for technical support and literature search. B. Southall and coauthors kindly provided access to unpublished information on M-weighting. Funding was provided under Minerals Management Service Cooperative Agreement Nos. 1435-01-02-CA-85186 and NA87RJ0445, the Office of Naval Research Grant Nos. N00014-99-1-0819

and N00014-02-10187, and the Strategic Environmental Research and Development Program Grant No. D8CA7201C0011. PJOM was supported by a Royal Society Fellowship and P.T.M. is currently supported by the Danish Natural Science Research Council via a Steno Fellowship. Sperm whale tagging was performed following the conditions of NMFS research permits 369-1440-01, 981-1578 and 981-1707 afforded to P.L.T. This research was approved by the Woods Hole Oceanographic Institution Animal Care and Use Committee.

[1]Sound level back-calculated to 1 m range on the acoustic axis of the source (Urick, 1983).

[2]Tolstoy *et al.* used much longer fixed averaging times for derivation of rms levels and their levels were not M-weighted.

Aguilar Soto, N., Johnson, P., Madsen, P. T., Tyack, P. L., Bocconcelli, A., and Borsani, J. F. (**2006**). "Does intense ship noise disrupt foraging in deep-diving Cuvier's beaked whales (*Ziphius cavirostris*)?," Marine Mammal Sci. **22**(3), 690–699.

ANSI (**1994**). "American National Standard Acoustical Terminology, 1994," ANSI S1.1–1994, Acoustical Society of America, New York, p. 9.

Au, W. W. L., Nachtigall, P. E., and Pawloski, J. L. (**1997**). "Acoustic effects of the ATOC signal (75 Hz, 195 dB) on dolphins and whales," J. Acoust. Soc. Am. **101**, 2973–2977.

Barger, J. E., and Hamblen, W. R. (**1980**). "The air gun impulsive underwater transducer," J. Acoust. Soc. Am. **68**, 1038–1045.

Brekhovskikh, L. M., and Lysanov, Y. P. (**2003**). *Fundamentals of Ocean Acoustics* (Springer, Berlin).

Blackwell, S. B., Lawson, J. W., and Williams, J. T. (**2004**). "Tolerance by ringed seals (*Phoca hispida*) to impact pipe-driving and construction sounds at an oil production island," J. Acoust. Soc. Am. **115**, 2346–2357.

Caldwell, J. (**2002**). "An introduction to the special cases: Effects of air guns on marine mammals," The Leading Edge **19**, 860–861.

Caldwell, J., and Dragoset, W. (**2000**). "A brief overview of seismic air-gun arrays," The Leading Edge **19**(8), 898–902.

De Ruiter *et al.* (unpublished).

Dragoset, W. (**1984**). "A comprehensive method for evaluating the design of airguns and airgun arrays," Geophysics **3**, 52–61.

Dragoset, W. (**1990**). "Airgun array specs: A tutorial," Geophysics **9**, 24–32.

Dragoset, W. (**2000**). "Introduction to airguns and airgun arrays," The Leading Edge **19**, 892–897.

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.

Fay, R. R., and Popper, A. N. (**1994**). *Comparative Hearing: Mammals* (Springer, New York).

Finneran, J. J., Schlundt, C. E., Dear, R., Carder, D. A., and Ridgway, S. H. (**2002**). "Temporary shift in masked hearing thresholds in odontocetes after exposure to single underwater impulses from a seismic watergun," J. Acoust. Soc. Am. **111**, 2929–2940.

Fleischer, G. (**1976**). "Hearing in extinct cetaceans as determined by cochlear structure," Journal of Palaeontology **50**, 133–152.

Fricke, J. R., Davis, J. M., and Reed, D. H. (**1985**). "A standard quantitative calibration procedure for marine seismic sources," Geophysics **50**, 1525–1532.

Gausland, I. (**2000**). "The impact of seismic surveys on marine life," The Leading Edge **19**, 903–905.

Goold, J. C., and Fish, P. J. (**1998**). "Broadband spectra of seismic survey air-gun emissions, with reference to dolphin auditory thresholds," J. Acoust. Soc. Am. **103**, 2177–2184.

Gordon, J. C., Gillespie, D., Potter, J. R., Frantzis, A., Simmonds, M. P., Swift, R., and Thompson, D. (**2004**). "A review of the effects of seismic surveys on marine mammals," Mar. Technol. Soc. J. **37**, 16–34.

Green, D. M. (**1985**). *Temporal Factors in Psychoacoustics*, in *Time Resolution in Auditory Systems*, edited by A. Michelson (Springer, New York), pp. 120–140.

Greene, C. R., and Richardson, W., J. (**1988**). "Characteristics of marine seismic survey sounds in the Beaufort Sea," J. Acoust. Soc. Am. **83**, 2246–2254.

Harris, C. M. (**1997**). *Handbook of Acoustical Measurements and Noise Control*, 3rd ed. (McGraw-Hill, New York).

Johnson, C. S. (**1968a**). "Relation between absolute threshold and duration-of-tone pulses in the bottlenosed porpoise," J. Acoust. Soc. Am. **43**, 757–763.

Johnson, C. S. (**1968b**). "Masked tonal thresholds in the bottlenosed porpoise," J. Acoust. Soc. Am. **44**, 965–967.

Johnson, M., and Tyack, P. L. (**2003**). "A digital acoustic recording tag for measuring the response of wild marine mammals to sound," IEEE J. Ocean. Eng. **28**, 3–12.

Ketten, D. R. (**1997**). "Structure and function in whale ears," Bioacoustics 8, 103–135.

Ketten, D. R. (**2000**). "Cetacean ears," in *Hearing in Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay, (Springer, New York), pp. 43–108.

Lepage, K., Malme, C., Mlawski, R., and Krumhansel, P. (**1995**). "Exon SYU Sound Propagation Study," BBN Report No. 8120, BBN Acoustic Technologies.

Lepage, K., Malme, C., Mlawski, R., and Krumhansel, P. (**1996**). "Mississippi Canyon Sound Propagation Study," BBN Report No. 8139, BBN Acoustic Technologies.

Ljungblad, D. K., Würsig, B., Swartz, S. L. and Keene, J. M. (**1988**). "Observations on the behavioral responses of bowhead whales (*Balaena mysticetus*) to active geophysical vessels in the Alaskan Beaufort Sea," Arctic **41**, 183–194.

Madsen, P. T. (**2002**). "Sperm whale sound production—in the acoustic realm of the biggest nose on record in Sperm whale sound production," PhD. dissertation, University of Aarhus, Denmark.

Madsen, P. T. (**2005**). "Marine mammals and noise: What is a safety level of 180 dB re. 1 uPa (rms) for transients?," J. Acoust. Soc. Am. **117**, 3952–3957.

Madsen, P. T., Møhl, B., Nielsen, B. K., and Wahlberg, M. (**2002**). "Male sperm whale behavior during exposures to distant seismic survey pulses," Aquat. Mamm. **28**, 231–240.

Malme, C. I., and Miles, P. R. (**1985**). "Behavioral responses of marine mammals (gray whales) to seismic discharges," in *Proceedings of the Workshop on Effects of Explosives Use in the Marine Environment, January 1985*, edited by G. D. Greene, F. R. Engelhardt, and R. J. Paterson, Tech. Rep. No. 5. Can. Oil & Gas Lands Adm., Environ. Prot. Br., pp. 253–280.

Malme, C. I., Miles, P. R., Clark C. W., Tyack, P., and Bird, J. E. (**1984**). "Investigations on the potential effects of underwater noise from petroleum industry activities on migrating gray whale behavior. Phase II, January 1984 migration," BBN Laboratories Inc., Cambridge, MA for U.S. Minerals Management Service, Washington, DC, BBN Report No. 5586, NTIS PB86–218377.

Malme, C. I., Miles, P. R., Tyack, D. P. L., Clark, C. W., and Bird, J. E. (**1985**). "Investigation of the potential effects of underwater noise from petroleum industry activities on feeding humpback whale behavior," BBN Laboratories Inc., Cambridge MA.

Malme, C. I., Smith, P. W., and Miles, P. R. (**1986**). "Characterization of geophysical acoustic survey sounds," OCS Study. Prepared by BBN Laboratories Inc., Cambridge, for Battelle Memorial Institute to the Department of the Interior-Mineral Management Service, Pacific Outer Continental Shelf Region, Los Angeles, CA.

Malme, C. I., Wursig, B., Bird, J. E., and Tyack, P. L. (**1986**). "Behavioral observations of gray whales to industrial noise: Feeding observations and predictive modeling," BBN Laboratories Inc., Cambridge, MA.

Malme, C. I., Wursig, B., Bird, J. E., and Tyack, P. L. (**1988**). "Observations of feeding gray whale responses to controlled industrial noise exposure," in *Port and Ocean Engineering under Arctic Conditions*, edited by W. M. Sackinger, M. O. Jefferies, J. L. Imm, and S. D. Treacy (University of Alaska, Fairbanks) Vol. **II**, pp. 55–73.

McCauley, R. D., Fewtrell, J., Duncan, A. J., Jenner, C., Jenner, M. N., Penrose, J. D., Prince, R. I. T., Adhitya, A., Murdoch, J., and McCabe, K. (**2000**). "Marine seismic surveys—A study of environmental implications," Australian Petroleum Production Exploration, Association, 692–708.

McCauley, R. D., Fewtrell, J., Duncan, A. J., Jenner, C., Jenner, M. J., Penrose, J. T., Prince, R. I. T., Adhitya, A., Murdoch, J., and McCabe, K. (**2003**). "Marine seismic surveys: Analysis and propagation of air-gun signals; and effects of exposure on humpback whales, sea turtles, fishes and squid. Environmental implications of offshore oil and gas development in Australia: further research," Australian Petroleum Production Exploration,

Association, Canberra, pp. 364–521.
Medwin, H., and Clay, C. S. (**1998**). *Acoustical Oceanography* (Academic, Boston).
Miller *et al.* (unpublished).
Nachtigall, P. E., Supin, A. Ya., Pawloski, J. L., and Au, W. W. L. (**2004**). "Temporary threshold shifts after noise exposure in a bottlenosed dolphin (*Tursiops truncatus*) measured using evoked auditory potentials," Marine Mammal Sci. **20**, 673–68.
Nieukirk, S. L., Stafford, K. M., Mellinger, D. K., Dziak, R. P., and Fox, G. F. (**2004**). "Low frequency whale and seismic airgun sounds recorded in the mid-Atlantic Ocean," J. Acoust. Soc. Am. **115**, 1832–1843.
NMFS (**2003**). "Taking marine mammals incidental to conducting oil and gas exploration activities in the Gulf of Mexico," Federal Register Vol. **68**, 9991–9996.
Nowacek, D. P., Johnson, M. P., and Tyack, P. L. (**2004**). "North Atlantic right whales (*Eubalaena glacialis*) ignore ships but respond to alerting stimuli," Proc. R. Soc. London **271**, 227–231.
NRC (**2000**). *Marine Mammals and Low-Frequency Sound* (National Academy Press, Washington, DC).
NRC (**2005**). *Marine Mammal Populations and Ocean Noise* (National Academic Press, Washington, DC), 126 pp.
Reeves, R. R., Ljungblad, D. K., and Clarke, J. T. (**1984**). "Bowhead whales and acoustic seismic surveys in the Beaufort Sea," Polar Record **22**, 271–280.
Richardson, W. J., Würsig, B., and Greene, C. R. (**1986**). "Reactions of Bowhead whales, *Balaena mysticetus*, to seismic exploration in the Canadian Beaufort Sea," J. Acoust. Soc. Am. **79**, 1117–1128.
Richardson, W. J., Greene, C. R., Malme, C. I., and Thompson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).
Ridgway, S. H., and Carder, D. (**2001**). "Assessing hearing and sound production in cetacean species not available for behavioral audiograms: Experience with Physeter, Kogia, and Eschrichtius," Aquat. Mamm. **27**(3), 267–276.
Schmidt, V. (**2004**). "Seismic contractors realign equipment for industry's needs," Offshore **64**, 36–44.
Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Greene, C. R., Jr., Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E., Richardson, W. J., Thomas, J. A., and Tyack, P. L. (**2006**). "Marine mammal noise exposure criteria; Exposure of single individuals to single sources" (unpublished).
Stone, C. J. (**2003**). *The Effects of Seismic Activity on Marine Mammals in UK Waters, 1998–2000* (JNCC, Peterborough).
Tolstoy, M., Diebold, J. B., Webb, S. C., Bohnenstiehl, D. R., Chapp, E., Holmes, R. C., and Rawson, M. (**2004**). "Broadband calibration of R/V Ewing seismic sources," Geophys. Res. Lett. **31**, doi:10.1029/2004GL020234.
Tyack, P. L., and Clark, C. W. (**2000**). "Communication and acoustic behavior of dolphins and whales," in *Hearing by Whales and Dolphins* edited by W. W. L., Au, A. N. Popper and R. R. Fay (Springer, New York), pp. 156–224.
Tyack, P. L., Gordon, J., and Thompson, D. (**2004**). "Controlled exposure experiments to determine the effects of noise on marine mammals," Mar. Technol. Soc. J. **37**, 41–53.
Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York).
Zimmer, W. M. X., Tyack, P. L., Johnson, M. P., and Madsen, P. T. (**2005**). "Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis," J. Acoust. Soc. Am. **117**, 1473–1485.