CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 114, No. 5 November 2003

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 5, NOVEMBER 2003

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 5, NOVEMBER 2003

CONTENTS—*Continued from preceding page*

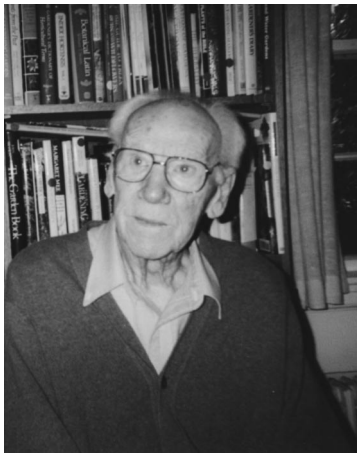
ACOUSTICAL NEWS—USA

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

Kenneth A. Cunefare—For contributions to structural acoustics and noise control.



John M. Harrison—For broad contributions to the understanding of the physiology of auditory efferent feedback.

Jeffrey A. Nystuen—For contributions to acoustical measurement of rainfall on the ocean.

Michael D. Richardson—For contributions to the acoustics of marine sediments.

Evan K. Westwood—For contributions in underwater sound propagation.

Reviewers of Manuscripts, 2002

Each year the Journal endeavors to publish a list of all the persons who reviewed manuscripts during the preceding year. The names of the following reviewers were omitted from the list of 2002 reviewers which appeared in the June 2003 issue beginning on page 2947. The Journal is very grateful to its reviewers and expresses its appreciation to these individuals who generously provide their time to help Associate Editors decide which papers should be published and to give constructive criticisms to the authors.

Abramson, Arthur S.
Horwitz, Amy R.
Tollin, Daniel J.

 0001-4966/2003/114(5)/2507/7/$19.00

Cynthia G. Clopper selected recipient of the 2003 Stetson Award

Cynthia G. Clopper

ASA member Cynthia G. Clopper was selected the recipient of the 2003–04 Raymond H. Stetson Scholarship in Phonetics and Speech Production. Ms. Clopper is a graduate student in the Department of Linguistics, conducting research in the Speech Research Laboratory at Indiana University, Bloomington, IN. She received a B.A. degree from Duke University. Ms. Clopper's current work includes development and collection of a multitalker, multidialect corpus of spoken language and perceptual and acoustic research on the categorization of American English regional dialects.

The Stetson Scholarship, which was established in 1998, honors the memory of Professor Raymond H. Stetson, a pioneer investigator in phonetics and speech science. Its purpose is to facilitate the research efforts of promising graduate students and postgraduates. The Scholarship includes a $3000 stipend for one academic year. Past recipients have been Roger Steeve (1999), Elizabeth K. Johnson (2000), Jeffery Jones (2001), and Meena Agarwal (2002).

Applications for the award are due in March each year. For further information about the award, please contact the 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; Web: http://asa.aip.org. Application information can also be found on the ASA Home Page at ⟨http://asa.aip.org/fellowships.html⟩.

ASA Celebrates the 25th Anniversary of the F. V. Hunt Fellowship in Acoustics

This year marks the 25th anniversary of the F. V. Hunt Postdoctoral Research Fellowship program of the Acoustical Society of America. The fellowship was established by the Society with ongoing support from the estate of former ASA President, Gold Medalist, and Harvard University professor Frederick V. Hunt. A scientist whose accomplishments included the training of 38 successful PhD candidates, as well as supervision of several additional post-doctoral fellows, Professor Hunt had a sincere desire to further the science of, and education in, acoustics.

At the end of their fellowship year Hunt Fellows are asked to write a report, to be published in the *Journal*, describing their activities. Some of these reports are scattered throughout JASA, with reports of several recent Fellows appearing below. The list of names of all recipients, included below, highlights the impact that this program has had on the Society itself. Many of the recipients have remained active and productive members of the Society.

F. V. Hunt Postdoctoral Research Fellows

1978–79—Steven L. Garrett
1979–80—Mary J. Osberger
1980–81—Cynthia A. Prosen
1981–82—Daniel Rugar
1982–83—Wesley N. Cobb
1983–84—Mark F. Hamilton
1984–85—Christine H. Shadle
1985–86—Anthony A. Atchley
1986–87—M. Christian Brown
1987–88—Ian M. Lindevald
1988–89—Elizabeth C. Oesterle
1989–90—E. Carr Everbach
1990–91—Kenneth A. Cunefare
1991–92—Gregory J. Sandell
1992–93—Quan Qi
1993–94—Charles E. Bradley
1994–95—T. Douglas Mast
1995–96—Robin O. Cleveland
1996–97—Mark A. Hasegawa-Johnson
1997–98—James J. Finneran
1998–99—Lily Wang
1999–00—Penelope Menounou
2000–01—James C. Lacefield
2001–02—Chao-Yang Lee
2002–03—Constantin C. Coussios
2003–04—Tryone M. Porter

Reports of Hunt Fellows

Reports of previous fellows can be found in JASA as follows: 8th Fellow—Anthony A. Atchley, Vol. 88, 2895 (1990); 9th Fellow—M. Christian Brown, Vol. 88, 2896 (1990); 12th Fellow—E. Carr Everbach, Vol. 89, 3011 (1991); 15th Fellow—Quan Qi, Vol. 97, 1341 (1995); and 18th Fellow—Robin Cleveland, Vol. 105, 1453 (1999).

Report of the 13th F. V. Hunt Postdoctoral Fellow (1990–1991)
Kenneth A. Cunefare, The Georgia Institute of Technology, School of Mechanical Engineering, Atlanta, Georgia 30332-0405

The F. V. Hunt Postdoctoral Fellowship enabled me to spend a year in residence at the Institute for Technical Acoustics, Technical University of Berlin, Germany, working with Herr Dr. Prof. Manfred Heckl. My research topic was the investigation of the concept of exterior acoustic radiation modes, the idea that exterior acoustic fields could be decomposed in an orthogonal expansion (via an eigenvalue problem or a singular value decomposition), where the eigenvalues correspond to radiation efficiencies and the corresponding eigenvectors correspond to particular vibration patterns on the surface of the body of interest.

The concept of exterior acoustic radiation modes was in its infancy. My approach to the concept developed from seeking answers to the question: What vibration pattern on the surface of a structure will be minimally efficient for acoustic radiation? Posing the question as a classical minimization problem, I was able to formulate an eigenvalue problem that yielded eigenvalues and eigenvectors that were subsequently interpreted as radiation efficiencies corresponding to particular patterns of vibration. Focusing on baffled beams and plates, I was able to develop and interpret the radiation modes as superpositions of structural modes. Using the analytical development of Davies related to the wavenumber dependency of the radiation efficiencies for individual structural modes, I was able to develop the wavenumber dependence for radiation modes below coincidence. I subsequently investigated issues related to convergence of the radiation modes, given that various functions or discretized-element-based methods could be used in the formulations of the basic problem. Further, the work naturally led to investigation of acoustic coupling between structural modes under harmonic excitation. Finally, I wrestled with the inverse problem of how to force a structure to respond in a desired vibration pattern matching a radiation mode pattern.

In the course of my work, I had occasion to use material from Cremer and Heckl's *Structure-Borne Sound* (2nd ed.). Imagine my consternation to discover therein the assertion that structural modes are uncoupled in the acoustic field! If such were true, then my work was wrong. After weeks of careful development and cross-checking, I was finally prepared to face Prof. Heckl with the audacious claim that I'd found an error in his text. As a newly minted Ph.D., I was not looking forward to the meeting. However, with characteristic grace and humor, Prof. Heckl instantly recognized that I had indeed found an error; all my anxiety was for nought.

It turns out that others were working on the concept at the same time: Borgiotti and Photiadis published original material on the radiation modes concept in 1990.

During the course of my Fellowship, I discussed and reviewed my work with Herr Dr. Prof. Manfred Heckl, as well as with Dr. Klaus Brod. I also interacted with Dr. Martin Ochmann, on matters related to acoustic radiation modeling using boundary element and multi-pole synthesis methods.

The following is a list of the publications and presentations that resulted, or followed from, the work I conducted under the Fellowship:

Journal Publications

Kenneth A. Cunefare, ''The minimum multimodal radiation efficiency of baffled finite beams,'' Journal of the Acoustical Society of America, **90**(5), pp. 2521–2529 (1991).

Kenneth A. Cunefare, ''The effect of modal interaction on sound radiation from vibrating structures,'' AIAA Journal, **30**(12), pp. 2819–2828 (1992).

Kenneth A. Cunefare and M. Noelle Currey, ''On the exterior acoustic modes of structures,'' Journal of the Acoustical Society of America, **96**(4), pp. 2302–2312 (1994).

Conference Proceedings and Presentations

Kenneth A. Cunefare, ''Minimum multimodal radiation efficiency of beams,'' in das Plenarvorträge und Kurzreferate der 17, Gemeinschaftstagung der Deutschen Arbeitsgemeinschaft für Akustik, Part A, pp. 169–172, April 1991.

Kenneth A. Cunefare, ''Active noise control and the modes of planar structures,'' AIAA 30th Aerospace Sciences Meeting and Exhibit, paper AIAA 92-0372, 6–9 January 1992, Reno, NV.

Noelle Currey and Kenneth A. Cunefare, ''On the convergence behavior of the exterior acoustic modes of structures,'' 124th Meeting of the Acoustical Society of America, New Orleans, LA, 31 October–4 November 1992

Kenneth A. Cunefare and Noelle Currey, ''The exterior acoustic modes of structures,'' 126th Meeting of the Acoustical Society of America, Denver, CO, 4–8 October 1993.

Noelle Currey and Kenneth A. Cunefare, ''The radiation modes of baffled finite plates,'' 127th Meeting of the Acoustical Society of America, Massachusetts Institute of Technology, Cambridge, MA, 6–10 June 1994.

Seminars

Kenneth A. Cunefare, ''Minimum multimodal radiation efficiency,'' Seminar, Drittes Physikalisches Institut, 28 November 1990, Göttingen, Federal Republic of Germany.

Kenneth A. Cunefare, ''The exterior acoustic modes of structures,'' Seminar, Institute of Sound and Vibration Research, The University, 25 April 1991, Southampton, United Kingdom.

Kenneth A. Cunefare, ''The acoustic velocity modes of structures,'' Seminar, Institute for Technical Acoustics, The Technical University of Berlin, 9 July 1991, Berlin, Federal Republic of Germany.

Kenneth A. Cunefare, ''New ideas on the topic of global optimization of active noise control and vibration,'' Center for Acoustics and Vibration, Seminar, The Pennsylvania State University, 19 August 1991, State College, Pennsylvania.

Kenneth A. Cunefare, ''The exterior modes of structures,'' Seminar, Georgia Chapter of the Acoustical Society of America, Georgia Tech, 17 October 1991, Atlanta, Georgia.

Report of the 17th F. V. Hunt Postdoctoral Fellow (1994–1995)

T. Douglas Mast, Ethicon Endo-Surgery, 4545 Creek Rd., Cincinnati, Ohio 45242

My tenure as a Hunt fellow was spent at the University of Rochester Diagnostic Ultrasound Research Laboratory in Rochester, New York, where I collaborated with Professor Robert Waag and others on theoretical, computational, and experimental investigations of acoustic scattering from soft tissues. These investigations had the general goal of improved methods for ultrasonic diagnosis of disease. The specific motivation for the research was to develop improved quantitative and high-resolution diagnostic techniques for use with next-generation ultrasonic apparatus.

The initial work in my fellowship included collaborating on the design, testing and implementation of a new ring transducer system.1 Part of this work involved analysis and computations applied to the design of experiments for measurement of intrinsic parameters of tissue from angle- and frequency-dependent scattering. The theoretical results of this research provided a quantitative characterization of available accuracy in ultrasonic measurements in terms of wavespace resolution.[2,3] This analysis helped to lay the groundwork for subsequent experiments that showed the feasibility of such measurements in a practical fixed transducer configuration. Experiments performed by Tomas Jansson using the ring transducer system were analyzed using a new data reduction method and showed good agreement between theoretical and measured differential scattering cross-sections for test objects.[4]

In studies of quantitative imaging using inverse scattering, I had the great pleasure of collaborating with Adrian Nachman as well as Professor Waag. These studies began with an analysis of the eigenfunctions of scattering operators that explained how these eigenfunctions focus on discrete and distributed scattering media.[5] We subsequently applied these focusing properties in a method for inverse scattering using the retransmitted fields of eigenfunctions. Combinations of retransmitted fields were used to construct optimal basis functions for expansion of unknown scattering media. The resulting inverse scattering method encompassed previous methods such as filtered backpropagation and time-reversal focusing, while also providing a means for simplification of nonlinear inverse scattering methods.[6,7]

Work with Victor Sparrow of Penn State on finite-difference modeling of propagation through inhomogeneous tissue found an application in modeling of wavefront distortion experiments performed by Laura Hinkelman. We proceeded to simulate ultrasonic pulse propagation through a number of abdominal wall cross-section models determined using a new staining technique.[8] An animation of our results was shown in the Gallery of Acoustics presented at the Fall 1995 ASA meeting in St. Louis. Visualization and analysis of the results provided some insight into the causes of ultrasonic wavefront distortion and the potential for its correction, particularly showing the importance of strong scattering from inhomogeneities within the subcutaneous fat.[9,10]

My learning experience at Rochester was also illuminated by many informal discussions on topics including physical acoustics, signal processing, inverse problems, and anatomy, with colleagues including all those mentioned above as well as Leon Metlay, Dong-Lai Liu, Jeffrey Astheimer, Ted Christopher, and David Blackstock.

In summary, the Hunt Fellowship provided me with a remarkable opportunity to study diverse areas of acoustics in a stimulating environment. I am very grateful to the Hunt family and the Acoustical Society for making this learning experience possible.

[1] R. C. Waag, D.-L. Liu, T. D. Mast, A. I. Nachman, P. Jaeger, and T. Kojima, ''An ultrasonic ring transducer system for studies of scattering and imaging,'' J. Acoust. Soc. Am. **100**, 2795 (1996).

[2] T. D. Mast, and R. C. Waag, ''Wavespace resolution in ultrasonic backscatter measurements,'' J. Acoust. Soc. Am. **95**, 2854 (1994).

[3] T. D. Mast, and R. C. Waag, ''Wave space resolution in ultrasonic scattering measurements,'' J. Acoust. Soc. Am. **98**, 3050–3058 (1995).

[4] T. T. Jansson, T. D. Mast, T. C. Heywood, and R. C. Waag, ''Measurements of differential scattering cross-section using a ring transducer,'' J. Acoust. Soc. Am. **103**, 3169–3179 (1998).

[5] T. D. Mast, and R. C. Waag, ''Eigenfunction and eigenvalue analysis of scattering operators,'' J. Acoust. Soc. Am. **96**, 3336 (1994).

[6] T. D. Mast, A. I. Nachman, and R. C. Waag, ''Inverse scattering using backpropagated eigenfunctions,'' J. Acoust. Soc. Am. **99**, 2545–2546 (1996).

[7] T. D. Mast, A. I. Nachman, and R. C. Waag, ''Focusing and imaging using eigenfunctions of the scattering operator,'' J. Acoust. Soc. Am. **102**, 715–725 (1997).

[8] L. M. Hinkelman, L. A. Metlay, C. J. Churukian, and R. C. Waag, ''Modified Gomori's trichrome stain for macroscopic tissue slices,'' J. Histotech. **19**, 321–323 (1996).

[9] T. D. Mast, L. M. Hinkelman, R. C. Waag, and V. W. Sparrow, ''Simulation of ultrasonic propagation through abdominal wall,'' J. Acoust. Soc. Am. **97**, 3325 (1995).

[10] T. D. Mast, L. M. Hinkelman, M. J. Orr, V. W. Sparrow, and R. C. Waag, ''Simulation of ultrasonic pulse propagation through the abdominal wall,'' J. Acoust. Soc. Am. **102**, 1177–1190 (1997) [Erratum: **104**, 1124–1125 (1998)].

Report of the 19th F. V. Hunt Postdoctoral Fellow (1996–1997)

Mark A. Hasegawa-Johnson, The University of Illinois, Beckman Institute, 405 N. Mathews, Urbana, Illinois 61801-0000

Although a static formant representation provides a good estimate of vowel quality, most phoneticians agree that a kinematic or dynamic repre-

sentation of some kind is necessary to fully characterize English vowels in fluent speech. As the 19th Hunt post-doctoral fellow (1996–1997), I worked with Professor Abeer Alwan at UCLA and Professor Ken Stevens at MIT to study the relationships among articulatory dynamics, vowel formant dynamics, and vowel spectral dynamics.

I began this research by testing simplified dynamic models of both tongue-surface fleshpoint data (provided by Dr. John Westbury, acquired using x-ray microbeam) and formant frequency data (obtained using the ESPS formant tracker). Both fleshpoint trajectories and formant frequency trajectories were modeled as the trajectories of point masses in a time-invariant second-order damped spring-mass system with step-wise shifting between consonant and vowel targets. Model parameters included the stiffness and damping of each point mass, target positions of the vowel and both flanking consonants, and target switching times: seven parameters per trajectory. Results showed that neither formant frequency dynamics nor fleshpoint dynamics are adequately modeled by a time-invariant second-order system. Analysis indicated that, in both cases, the assumption of constant stiffness and damping is false. Talkers either frequently adjust stiffness and damping parameters, or they smoothly adjust the target position from consonant to vowel; there is no way to differentiate these hypotheses using kinematic measurement data.

Professor Alwan began her MRI analysis of speech production in 1993; when I arrived at UCLA in 1996, she was already finishing a series of related studies of liquid and fricative consonant production. In order to better understand vowel production, she and I decided to acquire similar MRI stacks characterizing the static 3D ''target'' positions of 11 vowels. With UCLA students Jul Cha and Shamala Pizza, I wrote MRI segmentation software (Hasegawa-Johnson, 1999; Hasegawa-Johnson *et al.*, 1999a, 1999b; http://www.icsl.ucla.edu/~spapl/CTMRedit), and manually segmented all coronal and many of the axial image stacks (database available at http://www.ifp.uiuc.edu/speech/mri). Analysis of these data demonstrated two novel correspondences between acoustic and articulatory vowel spaces. First, three-dimensional tongue shape during vowel production supports two talker-independent PARAFAC factors, roughly corresponding to tongue height and tongue fronting (Hasegawa-Johnson and Zheng, 2000; Zheng *et al.*, 2003). Second, intertalker differences in palate height are actively reduced during production of a palatal vowel (by raising the tongue as necessary to maintain talker-independent cross-sectional area), but not during production of pharygeal or uvular vowels (Hasegawa-Johnson *et al.*, in press).

These observations from the field of speech production led me to explore a new model of speech recognition, in which the dependence of acoustic spectra on phoneme identity is mediated by the dynamics of two to four continuous-valued hidden state variables. To date, the continuous state variables in this model have been initialized using formant frequency trajectories (Hasegawa-Johnson, 2000), or directly from the acoustic spectrum (Zheng and Hasegawa-Johnson, in press).

Papers resulting in whole or in part from my Hunt post-doctoral fellowship include the following.

Articles in Refereed Journals

Hasegawa-Johnson, M. (**1998**). ''Electromagnetic exposure safety of the Carstens Articulograph AG100,'' J. Acoust. Soc. Am. **104**, 2529–2532.
Zheng, Y., Hasegawa-Johnson, M., and Pizza, S. (**2003**). ''PARAFAC analysis of the three-dimensional tongue shape,'' J. Acoust. Soc. Am. **113**(1).
Hasegawa-Johnson, M., Pizza, S., Alwan, A., Cha, J. S., and Haker, K. (**in press**). ''Vowel Category Dependence of the Relationship Between Palate Height, Tongue Height, and Oral Area,'' J. Speech Language and Hearing

Research Conference Papers.

Hasegawa-Johnson, M. (**1999**). ''Combining Magnetic Resonance Image Planes in the Fourier Domain for Improved Spatial Resolution,'' Int. Conf. on Sig. Proc. Applications and Technology, pp. 81.1–81.5.
Hasegawa-Johnson, M., Cha, J. S., and Haker, K. (**1999a**). ''CTMRedit: a Matlab-Based Tool for Segmenting and Interpolating MRI and CT Images in Three Orthogonal Planes,'' 21st Annual International Conference of the IEEE/EMBS Society, p. 1170.
Hasegawa-Johnson, M., Cha, J. S., Pizza, S., and Haker, K. (**1999b**). ''CTMRedit: A Case Study in Human-Computer Interface Design,'' International Conference on Public Participation and Information Technology, pp. 575–584.
Hasegawa-Johnson, M. (**2000**). ''Multivariate State Hidden Markov Models for Simultaneous Transcription of Phones and Formants,'' Proceedings of ICASSP, pp. 1323–1327.
Zheng, Y., and Hasegawa-Johnson, M. (**in press**). ''Acoustic Segmentation using Switching State Kalman Filter,'' Proceedings of ICASSP 2003.

Published Abstracts

Hasegawa-Johnson, M., and Zheng, Y. (**2000**). ''Three Dimensional Tongue Shape Factor Analysis,'' The ASHA Leader **5**(16):144.

Software and Databases

CTMRedit: a Matlab-Based Tool for Segmenting and Interpolating MRI and CT Images in Three Orthogonal Planes. Available at http://www.icsl.ucla.edu/~spapl/CTMRedit.
The Vowels MRI Database. Available at http://www.ifp.uiuc.edu/speech/mri.

Report of the 20th F. V. Hunt Postdoctoral Fellow (1997–1998)

James J. Finneran, U.S. Navy Marine Mammal Program, Biosciences Division, Space and Naval Warfare Systems Center, San Diego, 53560 Hull St., San Diego, California 92152-5001

As the 1997–1998 F. V. Hunt Postdoctoral Fellow, I worked with Dr. Sam Ridgway at the U.S. Navy Marine Mammal Program, Space and Naval Warfare Systems Center (SSC), San Diego. My proposed research topic was to examine the motion of the dolphin melon during echolocation. Dolphin echolocation pulses are generated in the nasal cavities and projected forward through the melon, a fatty structure located on the forehead. The melon has been hypothesized to function as an impedance matching device, coupling sounds generated within the dolphin to the surrounding water, and/or as an acoustic lens, focusing the outgoing sound waves. During my doctoral work, I developed an ultrasonic system to measure the amplitude and phase of vibrating structures and used the system to map the frequency response of the goldfish auditory organs. For my Hunt Fellowship, the goal was to use this ultrasound system to measure the surface displacement of the melon while a trained dolphin performed an echolocation task. Unfortunately, the transient nature of the echolocation clicks and the gross motion of the subject made reliable measurements extremely difficult. Fortunately, there were numerous other research projects to which I was able to contribute.

Soon after arriving at SSC San Diego, I developed a computer-controlled system for archiving the facility's rather large collection of magnetic tape-based data. During my tenure I was introduced to the behavioral techniques used to measure hearing capabilities in trained marine mammals and given the opportunity to collaborate with Dr. Ridgway on his marine mammal temporary threshold shift project. During this time I worked on improving sound field measurements and calibration procedures and integrating the computer system with the measurement hardware. My tenure as Hunt Fellow also allowed me the time to complete a journal article describing the active sound control system used in my Ph.D. research. I was also able to revisit the goldfish auditory system model described in my dissertation and make substantial improvements to it.

Since my tenure as Hunt Fellow, I have remained at SSC San Diego, first as an National Research Council Postdoctoral Research Fellow (again working with Dr. Ridgway), and finally as a fulltime employee of SSC San Diego. My main areas of focus have been the hearing capabilities of and the effects of sound on marine mammals.

The following is a list of publications that resulted, at least in part, from my time as Hunt Fellow.

Refereed Journal Publications

1. Finneran, J. J., and Hastings, M. C. (**1999**). ''Active impedance control in a cylindrical waveguide for generation of low frequency, underwater plane traveling waves,'' J. Acoust. Soc. Am. **105**(6), 3035–3043.
2. Finneran, J. J., and Hastings, M. C. (**2000**). ''A mathematical analysis of the peripheral auditory system mechanics in the goldfish (*Carassius auratus),*'' J. Acoust. Soc. Am. **108**, 1308–1321.
3. Finneran, J. J., Oliver, C. W., Schaefer, K. M., and Ridgway, S. H. (**2000**). ''Source levels and estimated yellowfin tuna (*Thunnus albacares*) detection ranges for dolphin jaw pops, breaches, and tail slaps,'' J. Acoust. Soc. Am. **107**(1), 649–656.
4. Schlundt, C. E., Finneran, J. J., Carder, D. A., and Ridgway, S. H. (**2000**). ''Temporary shift in masked hearing thresholds (MTTS) of bottlenose dolphins, *Tursiops truncatus*, and white whales, *Delphinapterus leucas*, after exposure to intense tones,'' J. Acoust. Soc. Am. **107**(6), 3496–3508.

Presentations

5. Finneran, J. J., and Hastings, M. C. (**1997**). ''Bioacoustics measurement system control and coordination using LabVIEW,'' J. Acoust. Soc. Am. **102**(5), 3196(A), invited presentation at the 134th Meeting of the Acoustical Society of America, San Diego, CA, December 1997.
6. Finneran, J. J., and Hastings, M. C. (**1997**). ''Active impedance control in a water-filled waveguide for low frequency bioacoustic testing,'' J. Acoust. Soc. Am. **102**(5), 3196–3197(A), presented at the 134th Meeting of the Acoustical Society of America, San Diego, CA, December 1997.
7. Finneran, J. J., and Hastings, M. C. (**1998**). ''A cohesive lumped parameter analysis of the mechanics of the goldfish peripheral auditory system,'' in Proceedings 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America, Volume II, pp. 1021–1022, presented at the 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America, Seattle, WA, June 1998.
8. Schlundt, C. E., Kamolnick, T., Smith, R., Elsberry, W. E., Finneran, J. J., Carder, D. A., and Ridgway, S. H. (**1998**). ''Temporary effects of louder sounds on the masked hearing thresholds of dolphins (*Tursiops truncatus*) and white whales (*Delphinapterus leucas*),'' presented at the 29th Meeting of the International Association of Aquatic Animal Medicine, San Diego, CA, May 1998.
9. Oliver, C. W., Finneran, J. J., and Schaefer, K. M. (**1999**). ''Low-frequency acoustic detection of yellowfin tuna, *Thunnus albacares*, and possibly dolphins by yellowfin tuna, in the eastern tropical Pacific Ocean,'' presented at the 13th Biennial Conference on the Biology of Marine Mammals, November 28–December 3, 1999.

Report of the 21st F. V. Hunt Postdoctoral Fellow (1998–1999)

Lily M. Wang, Architectural Engineering Program, University of Nebraska–Lincoln, Omaha, Nebraska 68182-0681 (lwang4@unl.edu)

My tenure as the 21st F. V. Hunt Postdoctoral Fellow was spent overseas, working with Dr. Anders Gade in the Acoustic Technology Department at the Technical University of Denmark (DTU). We focused our efforts on investigating the phenomenon of spatial impression in concert halls. This quality is defined as the degree to which one feels enveloped in the room's sound field. Various objective quantities calculated from binaural impulse responses had been proposed over the past three decades to account for this quality, including Lateral Energy Fraction (LEF) and Interaural Cross-Correlation Coefficient (IACC). LEF and IACC are both supposed to quantify spatial impression, but they are based on different physical explanations; thus, measurements of these two quantities made in several existing halls did not correlate well. Our research sought to clarify which objective measure (and therefore physical explanation) best matched the subjective perception of spatial impression. In addition to LEF and IACC, we also tested Late Lateral Level (GLL) and Interaural Level Fluctuation (IALF).

We began by creating auralizations using the room acoustic computer modeling program, Odeon, which was developed at DTU. Binaural impulse responses from eight source–receiver scenarios were selected so that there was wide scatter in the objective quantity data. Then auralizations were created from these impulse responses, using four different musical motifs. While deciding which musical motifs to use, we hypothesized that the choice of musical motif may affect the perceived spatial impression, but that this relative effect would be the same for each scenario.

Fourteen test subjects rated the spatial impression of the auralizations, presented over headphones in paired comparison tests. Results show that the test subjects' perceptions of spatial impression were highly correlated with the specific scenario, but that there was not a significant difference of spatial impression across musical motifs. Correlations with the various objective measures showed that GLL performed best, which itself was highly correlated with IALF.

Results from our work have been presented at two Acoustical Society of America meetings [Wang, L. M., and Gade, A. C., ''Subjective relevance of objective measures for spatial impression,'' J. Acoust. Soc. Am. **107**, 2891(A) (2000); Wang, L. M., ''Perceived spatial impression from various musical motifs,'' J. Acoust. Soc. Am. **109**, 2285(A) (2001)]. A paper with more recent findings is in preparation.

I would like to express my thanks to the Acoustical Society of America for the research funding provided by the F. V. Hunt Postdoctoral Fellowship, and to everyone at DTU who welcomed me and from whom I learned an immense amount. I am grateful that I was able to pursue advanced research in architectural acoustics, an area which is typically not well-funded in the United States. Certainly my experiences in Denmark have been invaluable towards starting an architectural acoustics research group at the University of Nebraska. Particularly, the experience on working with human subjective testing and with the room acoustics modeling program Odeon have continued to be pertinent to my assorted research projects. I continue to maintain close contact with the DTU Room Acoustics Research Group, and am planning to return to Denmark in upcoming summers to collaborate on a study involving directional sources in auralizations.

Report of the 22nd F. V. Hunt Postdoctoral Fellow (1999–2000)

Penelope Menounou, The University of Texas, Department of Mechanical Engineering, Austin, Texas 78720-4090

The F. V. Hunt Postdoctoral Research Fellowship gave me the opportunity to work at the Mechanical Engineering Department of the University of Texas at Austin on propagation of finite amplitude noise. The idea that was explored was to predict the effect of nonlinear propagation distortion of the power spectral density of a given noise source, when the time waveform itself is not known, only its power spectral density. A new method was developed that predicted the combined effect of nonlinearity and thermoviscous attenuation directly on the power spectral density of a given noise source. The method was found to be valid for short propagation distances.

Besides the work on nonlinear propagation, I was able to study jagged edge noise barriers (walls built alongside the highways to protect nearby communities from the traffic noise). This type of noise barriers constitute a new design that can improve the shielding effect of the barrier with the conventional straight top edge. This work laid the foundations for a fruitful line of work and continuous funding up to these days. Two subsequent projects were funded which allowed the creation of a research group working on experiments, theory, and computer simulations of noise propagation. The group includes two distinguished members of the ASA, Professor David T. Blackstock and Professor Wayne M. Wright, as well as three students. For this opportunity, I am most grateful to F. V. Hunt and to the Acoustical Society of America.

The following is a list of publications where the F. V. Hunt Fellowship is acknowledged.

P. Menounou, ''Numerical solution of a statistical version of the Burgers equation,'' J. Acoust. Soc. Am. **106**(4), 2174 (1999).
P. Menounou and I. J. Busch-Vishniac, ''Jagged Edge Noise Barriers,'' J. Building Acoust. **7**(3), 179–200 (2000).
P. Menounou, I. J. Busch-Vishniac, and D. T. Blackstock, ''Directive line source model: A new model for sound diffraction by half planes and wedges,'' J. Acoust. Soc. Am. **107**(6), 2973–2986 (2000).
P. Menounou, ''A correction to Maekawa's curve for the insertion loss behind barriers,'' J. Acoust. Soc. Am. **110**(4), 1828–1838 (2001).

The F. V. Hunt period will also be imprinted in my memory for a very personal reason. During the fellowship period I gave birth to our first daughter, Daphne.

Report of the 23rd F. V. Hunt Postdoctoral Fellow (2000–2001)

James C. Lacefield, Departments of Electrical and Computer Engineering and Medical Biophysics, University of Western Ontario, London, Ontario N6A 5B9, Canada

With the support of the F. V. Hunt Fellowship I investigated methods for estimation and compensation of medical ultrasound focusing errors (i.e., aberration) produced by propagation through heterogeneous tissue. This research was performed at the University of Rochester in collaboration with Professor Robert C. Waag. The principal objective of the work was to characterize fundamental factors that determine the performance of aberration correction techniques. The experiments emphasized fluctuations in pulse arrival times because this form of aberration is a prominent source of focus degradation in medical imaging. We hypothesized that measurements performed using realistic models of aberration would provide quantitative results that could guide the development of more effective focusing methods. The long-term goal of this research is to improve the diagnostic utility of ultrasound imaging in technically challenging applications including breast cancer detection and examination of obese patients. Realistic experimental measurements were made possible by a fully programmable, large-aperture two-dimensional array system and novel tissue-mimicking test objects (i.e., ''phantoms''). The phantoms were jointly developed by the laboratories of

Professor Waag and Professor Ernest Madsen of the University of Wisconsin. The first- and second-order statistics of spatial variations in the energy and arrival times of ultrasound pulses distorted by these phantoms are comparable to previous measurements of aberration produced by transmission through human abdominal wall specimens (Lacefield, Pilkington, and Waag, 2002). The phantoms are distinguished from more idealized physical models of aberration by the inclusion of three-dimensionally distributed aberrating structures and by the production of statistically realistic energy fluctuations. We expect this type of phantom to become a standard model for aberration correction experiments.

One factor we investigated was the height of transducer array elements used in systems that compensate for arrival time fluctuations (Lacefield and Waag, 2001a). Adjacent two-dimensional array elements were combined electronically to synthesize rectangular elements with differing heights. Echo arrival times were estimated less accurately with larger elements due in part to spatial averaging of the pulses over the area of the elements. The largest elements that permitted effective time-shift compensation were similar in height to the elements presently used in multirow arrays. Therefore, increasing the element height to reduce the channel count and hence the cost of current multirow array systems is not a viable design option. This conclusion was confirmed in a follow-up study (Lacefield and Waag, 2002a) using additional element heights that was performed at the suggestion of colleagues who attended conference presentations of the initial results.

A separate series of experiments (Lacefield and Waag, 2002b) examined the spatial coherence of echoes acquired from randomly distributed particles that model ultrasound scattering from soft tissue. The spatial coherence of random-scattering echoes is determined by the quality of the transmit focus and the aberration experienced by the echoes on their return trip to the transducer. Arrival time differences are more readily measured when echoes have high spatial coherence because higher coherence corresponds to greater similarity in pulse shape. Compensating for arrival time distortions during transmit focusing improved the coherence of echoes received several elements apart, but had little effect on the coherence of echoes received at adjacent elements. This result indicates that time-shift compensation provides a useful first focus correction step, but opportunities remain for substantial focusing improvements if methods can be devised to correct distortions in pulse shape and amplitude. I have continued to collaborate with Professor Waag's group to address these issues in my current position at the University of Western Ontario.

The research funded by the 2000–2001 Hunt Fellowship yielded four refereed publications and contributed to six oral presentations. I also benefited from the opportunity to travel to four conferences with the assistance of Fellowship funds. I am grateful to the Hunt family and the Acoustical Society of America for the financial support that made this work possible and for the recognition provided by the Fellowship.

Refereed Publications

Lacefield, J. C., Pilkington, W. C., and Waag, R. C. (**2002**). ''Distributed aberrators for emulation of ultrasonic pulse distortion by abdominal wall,'' ARLO **3**, 47–52.

Lacefield, J. C., and Waag, R. C. (**2002a**). ''Examples of design curves for multirow arrays used with time-shift compensation,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1340–1344.

Lacefield, J. C., and Waag, R. C. (**2002b**). ''Spatial coherence analysis applied to aberration correction using a two-dimensional array system,'' J. Acoust. Soc. Am. **112**, 2558–2566.

Lacefield, J. C., and Waag, R. C. (**2001a**). ''Time-shift estimation and focusing through distributed aberration using multirow arrays,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1606–1624.

Conference Proceedings and Abstracts

Lacefield, J. C., and Waag, R. C. (**2001b**). ''Evaluation of backpropagation methods for transmit focus compensation,'' 2001 IEEE Ultrasonics Symposium Proceedings (Institute of Electrical and Electronics Engineers, Piscataway, NJ), pp. 1495–1498.

Lacefield, J. C., and Waag, R. C. (**2001c**). ''Effect of aperture size on compensation of arrival time distortion,'' J. Acoust. Soc. Am. **109**, 2396.

Lacefield, J. C., and Waag, R. C. (**2000**). ''Effect of transmit focus characteristics on estimates of aberration,'' 2000 IEEE Ultrasonics Symposium Proceedings (Institute of Electrical and Electronics Engineers, Piscataway, NJ), pp. 1665–1668.

Phillips, D. B., Lacefield, J. C., and Waag, R. C. (**2001**). ''B-scan imaging using adaptive beamformation for aberration correction,'' J. Ultrasound Med. **20**, S65–S66.

Pilkington, W. C., Lacefield, J. C., and Waag, R. C. (**2002**). ''Ultrasonic b-scan imaging with adaptive beamformation using aberration correction,'' J. Acoust. Soc. Am. **111**, 2352.

Waag, R. C., Lacefield, J. C., Phillips, D. B., and Busse, L. J. (**2000**). ''Ultrasonic wavefront coherence in the presence of aberration,'' 2000 IEEE International Ultrasonics Symposium, Abstract 2J-5.

Report of the 24th F. V. Hunt Postdoctoral Fellow (2001–2002)

Chao-Yang Lee, Speech Communication Group, 36-547 Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Mapping continuous acoustic signal onto discretely specified lexical representations constitutes the basis for auditory language comprehension. The broad goal of my research was to delineate how acoustic properties of speech are processed, based on knowledge of articulatory-acoustic-linguistic relationships, to access lexical representations. As the Hunt Fellow for the year 2001–2002 at MIT, I explored whether the approach of lexical access based on detecting acoustic landmarks and extracting phonetic features, proposed by Professor Ken Stevens at MIT, could be extended to Mandarin Chinese, a tone language in which fundamental frequency variations are used for lexical distinctions.

During my doctoral thesis research at Brown University, psycholinguistic tasks were used to evaluate the cognitive processes involved in the lexical processing of tones. The findings suggest that tones are used on-line for lexical distinctions and that tones produced in context could be identified with relatively little acoustic input, indicating the use of phonetic contextual information as suggested previously by Yi Xu. The focus of my research at MIT shifted from the cognitive processes to the acoustic properties of tones. The specific question was whether tones could be reliably estimated by extracting acoustic information around the vowel landmark, which also led to the question of whether the model of lexical access from landmarks and features would need modification to accommodate the processing of tonal information. A database was generated of six speakers producing Mandarin monosyllabic words of various syllabic structures. Preliminary findings indicate that the distribution of fundamental frequency information varies across syllable structures. In particular, some seemingly critical information such as the rise part of the rising tone does not appear until in the nasal consonant following the nucleus vowel. While cases like this might pose a challenge to the model, it was unclear whether information around the vowel landmark was all that a listener could use to uncover tones. A number of tone perception experiments are planned underway to further address the question.

During this year at MIT, I also conducted a study on the acoustic characteristics of Mandarin strident fricatives. In particular, Mandarin distinguishes between flat and palatalized postalveolar fricatives in addition to the alveolar-palatal distinction frequently seen in many languages. This study aimed to quantitatively explicate the acoustic characteristics of the fricatives and to evaluate the mapping from the acoustic properties onto phonetic features in the context of the acoustic theory of speech production. Acoustic analyses were conducted of the spectral properties of the frication noise and the following vowel. The results indicate the frication noise for each fricative is associated with distinct natural frequencies of the vocal tract, consistent with the idea that phonetic distinctions are shaped by quantal relations between articulation and acoustics. A number of measures were also identified that statistically distinguished all three fricatives. These findings will be presented in the Nashville meeting in May 2003.

I am deeply grateful for the support from the Hunt Fellowship for allowing me to study with Professor Ken Stevens and the Speech Communication Group in the Research Laboratory of Electronics at MIT. With a graduate background in the cognitive and linguistic sciences, I was intellectually challenged to understand and integrate different ways of approaching my research questions. The broader context of the Institute also exposed me to many useful computational tools and much relevant research in the speech, hearing, and cognitive sciences. The interdisciplinary stimulation proved to be an invaluable experience for me.

USA Meetings Calendar

Listed below is a summary of meetings related to acoustics to be held in the U.S. in the near future. The month/year notation refers to the issue in which a complete meeting announcement appeared.

2004

24–28 May	75th Anniversary Meeting (147th Meeting) of the Acoustical Society of America, New York, NY [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: asa.aip.org]. Deadline for receipt of abstracts: 16 January 2004.
3–7 August	8th International Conference of Music Perception and Cognition, Evanston, IL [School of Music, Northwestern University, Evanston, IL 60201; WWW: www.icmpc.org/conferences.html].
15–19 Nov.	148th Meeting of the Acoustical Society of America, San Diego, CA [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: asa.aip.org].

Cumulative Indexes to the *Journal of the Acoustical Society of America*

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

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*

Reverberant acoustic energy in auditoria that comprise systems of coupled rooms [43.55.Br, 43.55.Ka, 43.55.Gx, 43.55.Fw, 43.20.Fn, 43.20.Px]—Jason E. Summers, *Program in Architectural Acoustics, Rensselaer Polytechnic Institute, Troy, NY 12180, 29 August 2003 (Ph.D.).* A frequency-dependent model for reverberant energy in coupled rooms is developed and compared with measurements for a 1:10 scale model and for Bass Hall, Ft. Worth, TX. At high frequencies, prior statistical-acoustics models are improved by geometrical-acoustics corrections for decay within sub-rooms and for energy transfer between sub-rooms. Comparisons of computational geometrical acoustics predictions based on beam-axis tracing with scale model measurements indicate errors resulting from tail-correction assuming constant quadratic growth of reflection density. Using ray tracing in the late part corrects this error. For mid-frequencies, the models are modified to account for wave effects at coupling apertures by including power transmission coefficients. Similarly, statical-acoustics models are improved through more accurate estimates of power transmission measurements. Scale model measurements are in accord with the predicted behavior. The edge-diffraction model is adapted to study transmission through apertures. Multiple-order scattering is theoretically and experimentally shown inaccurate due to neglect of slope diffraction. At low frequencies, perturbation models qualitatively explain scale model measurements. Measurements confirm relation of coupling strength to unperturbed pressure distribution on coupling surfaces. Measurements in Bass Hall exhibit effects of the coupled stage house. High frequency predictions of statistical acoustics and geometrical acoustics models and predictions of coupling apertures all agree with measurements.

Thesis advisor: Rendell R. Torres

Electronic copies of this thesis may be obtained from the author. E-mail address: jesummers@alumni.kenyon.edu

Acoustics of native-American ceremonial sites in prehispanic America (**Acustica en los espacios escenicos rituales prehispanicos**) [43.55.Gx]—Maria Isabel Martinez, *Department of Architectural Constructions I, University of Seville, Seville, Spain, December 2002 (Ph.D.).* This thesis establishes a methodology that incorporates the latest procedures used in architectural acoustics for the study of open spaces of this general type, and definitions are given for the acoustic variables of interest. The ''Juego de Pelota'' (ball game) sites are the only ceremonial sites built specifically for the performance of a fertility ritual, and are ideal for the study of prehispanic architectural topographies. Analysis of the acoustic properties of such sites revealed that the topographical characteristics of the elevation profiles of these architectural structures determine the acoustic behavior of these spaces. Such profiles are classified into three basic types: (*i*) inclined profile, (*ii*) terraced profile, and (*iii*) mixed profile. The terraced profiles are the most efficient, and the mixed profiles are the least efficient, in regard to acoustics. The consideration of the acoustic behavior of architectural structures intended for the ''Ball Game,'' as the designs evolved over time, leads to the conclusion that acoustical sensations that contributed effectively to the characteristic mystical atmosphere of the ceremonial rituals were characteristic only of those sites constructed in the ''classical'' period.

Thesis advisors: Jaime Navarro and Juan J. Sendra

Copies of this thesis written in Spanish may be obtained by contacting the advisor Jaime Navarro, E.T.S. de Arquitectura de Sevilla, Dpto. de Construcciones Arquitectonicas I, Av. Reina Mercedes, 2, 41012 Sevilla, Spain. E-mail address: jnavarro@us.es

Sound in ecclesiastical spaces in Cordoba. Architectural projects incorporating acoustic methodology (**El sonido del espacio eclesial en Cordoba. El proyecto arquitectonico como procedimiento acustico**) [43.55.Gx, 43.55.Hy, 43.55.Ka]—Rafael Suarez, *Department of Architectural Constructions I, University of Seville, Seville, Spain, July 2002 (Ph.D.).* This thesis is concerned with the acoustic analysis of ecclesiastical spaces, and the subsequent implementation of acoustic design methodology in architectural renovations. One begins with an adequate architectural design of specific elements (shape, materials, and textures), with the intention of elimination of acoustic deficiencies that are common in such spaces. These are those deficiencies that impair good speech intelligibility and good musical audibility. The investigation is limited to churches in the province of Cordoba and to churches built after the reconquest of Spain (1236) and up until the 18th century. Selected churches are those that have undergone architectural renovations to adapt them to new uses or to make them more suitable for liturgical use. The thesis attempts to summarize the acoustic analyses and the acoustical solutions that have been implemented. The results are presented in a manner that should be useful for the adoption of a model for the functional renovation of ecclesiastical spaces. Such would allow those involved in architectural projects to specify the nature of the sound, even though somewhat intangible, within the ecclesiastical space.

Thesis advisors: Jaime Navarro and Juan J. Sendra

Copies of this thesis written in Spanish may be obtained by contacting the advisor, Jaime Navarro, E.T.S. de Arquitectura de Sevilla, Dpto. de Construcciones Arquitectonicas I, Av. Reina Mercedes, 2, 41012 Sevilla, Spain. E-mail address: jnavarro@us.es

Architectural acoustics and the heritage of theater architecture in Andalusia (**Acustica arquitectonica y patrimonio teatral en Andalucia**) [43.55.Gx, 43.55.Hy, 43.55.Ka]—Angel Luis Leon, *Department of Architectural Constructions I, University of Seville, Seville, Spain, November 2001 (Ph.D.).* This thesis reports on the study of the acoustic properties of 18 theaters belonging to the Andalusian historical and architectural heritage. These theaters have undergone recent renovations to modernize and equip them appropriately. Coincident with this work, evaluations and qualification assessments with regard to their acoustic properties have been carried out for the individual theaters and for the group as a whole. Data measurements for this purpose consisted of acoustic measurements *in situ*, both before the renovation and after the renovation. These results have been compared with computer simulations of sound fields. Variables and parameters considered include the following: reverberation time, rapid speech transition index, back-ground noise, definition, clarity, strength, lateral efficiency, interaural cross-correlation coefficient, volume/seat ratio, volume/audience-area ratio. Based on the measurements and analysis, general conclusions are given in regard to the acoustic performance of theaters whose typology and size are comparable to those that were used in this study (between 800 and 8000 cubic meters). It is noted that these properties are comparable to those of the majority of European theaters. The results and conclusions are presented so that they should be of interest to architectural acoustics practitioners and to architects who are involved in the planning of renovation projects for theaters.

Thesis advisors: Juan J. Sendra and Jaime Navarro

Copies of this thesis written in Spanish may be obtained by contacting the author, Angel L. Leon, E.T.S. de Arquitectura de Sevilla, Dpto. de Construcciones Arquitectonicas I, Av. Reina Mercedes, 2, 41012 Sevilla, Spain. E-mail address: leonr@us.es

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).]*

DAFX—Digital Audio Effects

Udo Zölzer (Editor)

John Wiley and Sons, New York, 2002.
xix+533 pp. Price: $95.00 (hardcover) ISBN: 0-471-49078-4.

The acronym ''DAFX'' stands for ''Digital Audio Effects,'' which began as the name of a European conference on that topic, held annually since 1997. This book is a collection of chapters by 13 authors from the DAFX community, many of whom are known for their past research contributions. The book spans an enormous scope of technology for digital audio effects, and is likely the first book of its scale devoted to this area. Due to the research focus of its authors, it covers many recent developments as well as more ''classical'' effects. The ''DAFX book'' is a welcome contribution in a field that has been historically characterized by trade secrets, occasional patents, and precious little open scholarly research literature.

The first five chapters (135 pp.) are by Pierre Dutilleux and Udo Zölzer, the editor. They begin with a survey of commonly used signal processing functions such as plotting signals and spectra, the discrete Fourier transform, FFT windowing, and spectrograms, with implementation examples given in Matlab. Chapter 2 is about filters used in audio, both analog and digital, including the Sallen-Key analog filter, the state variable filter (analog and digital), and prevalent equalizer components such as shelving filters, peak filters, and parametric first- and second-order sections. Time-varying filters such as the ''wah-wah'' and ''phaser'' are defined. Chapter 3 is a short (12 pp.) summary of elementary delay-based effects such as comb filters, fractional delay lines, flanging, chorus, slapback, echo, and multiband effects. Chapter 4 (30 pp.) covers modulation and demodulation, including ring modulation, amplitude and frequency modulation, elementary envelope detection, and applications. There are interesting examples of stereo phasing and a rotary loudspeaker effect using sinusoidal modulation of gain. Chapter 5 (43 pp.) is a relatively advanced chapter on nonlinear processing, including effects such as dynamic gain equalization, limiters, noise gates (with a two-page Matlab example), de-essing, electronic tube simulation (with a Matlab example), overdrive (popular for guitar processing), and a brief discussion of aural exciters/enhancers. Even relatively experienced researchers and practitioners are likely to benefit from the knowledge and expertise assimilated in this chapter.

Chapter 6 (64 pp.) is ''Spatial Effects'' by Davide Rocchesso. The coverage includes a blend of signal processing and perceptual considerations, as any treatment of spatial audio effects should. Topics include the precedence effect, Doppler effect, 3D simulation in headphones (including a Matlab function for simulating head shadowing), 3D simulation with loudspeakers (add a stereo cross-canceler to a headphone simulator), 3D panning, Ambisonics (a particular generalization of stereo to 3D), transaural stereo (stereo cross-canceling again), and an introduction to methods for artificial reverberation. The special filters known as ''head related transfer functions'' (HRTF) are not provided, but the basic concepts and considerations are well presented. The topic of reverberation is relatively well covered (22 pp.), including a summary of the use of comb/allpass filters, Schroeder reverberators, feedback delay networks, brute force convolution, and sound radiation simulation. The bibliography appears to be fairly complete (over five pages of references); not cited, however, are the Synthesis Tool Kit (STK—free software providing several Schroeder reverberators in C++), and pointers to the many on-line references and materials in this area.

Chapter 7 (36 pp.) is ''Time Segment Processing'' by Dutilleux, Giovanni De Poli, and Zölzer. Topics covered include time scale modification (TSM—speeding up or slowing down an audio recording without altering pitch), the SOLA (synchronous overlap-add) algorithm for TSM (with Matlab), PSOLA (pitch-synchronous SOLA, with Matlab), pitch shifting (dual of TSM, with SOLA and PSOLA based Matlab examples), and time-shuffling and granulation (for so-called ''granular synthesis'' of musical sound, with Matlab). This chapter is exemplary in that Matlab examples are included for every major technique discussed (the chapters are inconsistent in this regard). Furthermore, the techniques chosen are tried and true performers, yielding high-quality results for a wide variety of signal types.

Chapter 8 (61 pp.) is ''Time-Frequency Processing'' by Daniel Arfib, Florian Keiler, and Zölzer. After a brief review of the short-time Fourier transform (STFT), various vocoder techniques are summarized. The two main dual STFT interpretations, filter bank summation and overlap-add, are covered as alternate approaches to phase vocoder implementation. The Matlab coverage in this chapter is good. Further topics include phase unwrapping, elementary instantaneous frequency computation (unfortunately, more advanced methods are not even cited), time-frequency filtering, time scale modification (TSM), pitch shifting, elementary transient segmentation, morphing from one sound to another, ''robotization,'' ''whisperization,'' and elementary denoising. While this chapter does not summarize the advanced signal processing literature on the topics covered, it does present a reasonably balanced overview with a good amount of elementary starter-code in Matlab.

Chapter 9 (374 pp.) is ''Source-Filter Processing'' by Arfib, Keiler, and Zölzer. Like the previous chapter, Matlab coverage is good. ''Source-filter processing'' refers to modeling signals as a wideband source (such as noise or an impulse train) which is fed through a filter to remove energy at certain frequencies. In computer music, it is called ''subtractive synthesis'' (in contrast with ''additive synthesis'' which models sound as a sum of quasi-sinusoidal components). Topics addressed in this chapter include linear prediction (which can be regarded as a method of computing spectral envelope parameters), the cepstrum (also used to estimate spectral envelopes by smoothing the log magnitude spectrum), cross-synthesis (impressing the spectral envelope of one sound on another), spectral interpolation, pitch shifting, FFT-based spectral peak detection and pitch extraction, voiced/unvoiced discrimination, linear prediction across a pitch period, amplitude envelope estimation, and some elementary statistical feature measurements.

Chapter 10 (66 pp.) is ''Spectral Processing'' by Amatriain, Bonada, Loscos, and Serra. This is another chapter on spectral modeling of sound, but, unlike the previous two chapters, it is based on sines+noise modeling. Xavier Serra and his students and collaborators in Barcelona are very well known for their work in sines+noise modeling of musical signals, and their sound examples and demos are consistently impressive. This chapter includes quite a bit of Matlab code, and the basic elements of sinusoidal modeling are well presented. Topics covered include sinusoidal modeling, sines+noise modeling, use of the STFT, sinusoidal peak detection (with two pages of Matlab just for that), pitch estimation, formation of peak tracks

through time (with more than two pages of Matlab), noise-floor estimation by peak subtraction, feature analysis, resynthesis, and effects and transformations such as pitch shift, spectral reshaping, gender changing, harmonizing, ''hoarseness,'' spectral morphing, and more. Since there is considerably more model specificity in this framework, very interesting effects can be obtained, such as selectively altering spectral peaks belonging to one harmonic source in a mixture of sources. A seven-page Matlab listing provides the main analysis/resynthesis application.

Chapter 11 (25 pp.) is a short chapter entitled ''Time and Frequency Warping Musical Signals'' by G. Evangelista. This is a relatively specialized chapter on abscissa warping in either the time or frequency domains. An interesting section on energy-invariant frequency warping is given. Surprisingly, there is no discussion of practical methods for resampling based on bandlimited interpolation: time warping is discussed briefly only in the continuous time case, and only simple ''rounding to the nearest FFT bin'' is presented for basic spectral warping. Most of the chapter is devoted to spectral frequency warping based on dispersive delay lines and the Laguerre transform (which is never explicitly defined). The work is closely related to that of M. Karjalainen and his colleagues at HUT, but inexplicably they and others are not cited.

It is a notable omission in the book as a whole that the subject of bandlimited interpolation is not covered in any chapter. It is hard to imagine a topic more fundamental to digital audio effects. Fortunately, this hole is readily filled by a Google search.

Chapter 12 (33 pp.) is ''Control of Digital Audio Effects'' by T. Todoroff. Topics include control mapping, scaling, GUI design, real-time issues, GUI languages, parametric feature selection, gestural interfaces, a little about MIDI, and consideration of controllers for various specific instrument classes. Also discussed are force-feedback interfaces, interfaces worn on the body, and controllers which need not be touched at all (the 1919 Theremin being the classic original example). There is no software provided in this chapter, but the bibliography is extensive.

Finally, Chap. 13 (15 pp.) is ''Bitstream Signal Processing'' by Mark Sandler and Zölzer. Topics include sigma-delta modulation (SDM) (used in present day digital/analog converters), linearization of SDM, addition and multiplication of bitstream signals, and elementary filter structures for bitstream signals, such as FIR filters and first- and second-order parametric equalizer sections. While primarily of interest to VLSI chip designers, this chapter is well written and informative. Matlab listings are provided for the first- and second-order SDM converters (sample stream in, bitstream out).

Following these 13 chapters is a ten-page glossary (including a few French and German terms such as ''brassage'' and ''flatterzunge'') and a nine-page global index. It would have been nice to cross-reference glossary terms to the chapters/sections/pages which define and develop them.

An unusual feature of this book is that most chapters contain a section on ''Sound and Music'' just before the chapter's bibliography. Listed here are musical pieces exemplifying the techniques discussed.

Evidently, only 1 of the 13 authors is a native English speaker (Mark Sandler of the UK), and yet the English is uniformly excellent throughout. A valuable benefit for English-speaking readers is that some authors have summarized interesting literature written only in French and German.

In summary, the ''DAFX book'' is a comprehensive reference work for new and experienced researchers and practitioners alike. It provides cogent summaries of a large percentage of audio effects in current use, and extensive references to the research literature.

JULIUS O. SMITH III
Center for Computer Research in Music and Acoustics (CCRMA)
Stanford University
Stanford, California 94305-8180

Acoustic Communication

Andrea Megela Simmons, Arthur N. Popper, and Richard R. Fay (editors)

Springer-Verlag, New York, 2003.
404 pp. Price: $149.00 (hardcover) ISBN: 0-387-98661-8.

The seven chapters in this volume contribute to a surprisingly broad overview of acoustic communication in animals. Moreover, the authors deal not only with mechanisms but also with communicative functions in contemporary populations and broad-scale patterns of evolution. The brief introductory chapter by Simmons provides a useful overview of the field and indicates how each of the chapters contribute to an up-to-date summary of some aspect of acoustic communication as well as suggesting future research directions.

The chapter by Bass and Clark contains a detailed synopsis of general principles of underwater bioacoustics, some of which apply to their own research with sound communication in bony fishes in shallow water (Bass) and whales in relatively deep water (Clark). Whereas the signals of many whales seem well-suited for long-range communication in deep-water marine environments, the signals of bony fishes are often too low in frequency to propagate optimally in shallow water. The authors suggest several explanations for this apparent mismatch, including the need to avoid detection by predators and competitors—a functional explanation—and constraints imposed by mechanisms of sound production and (ancestral) hearing.

The chapter by Fitch and Hauser is especially lively and thought-provoking. In considering the ''honesty'' of acoustic signals, they show how mechanisms, which include the physics of efficient sound radiation and the multiple functions of sound-producing structures in many animals, generally enforce the transmission of reliable information about attributes of the signaler such as its body size. Nevertheless, such constraints can sometimes be subverted by relatively simple changes in the structure of the vocal apparatus or tract. The authors also provide a thorough treatment of perceptual and cognitive factors that shape and constrain dishonest signaling and the detection of unreliable signals by receivers.

Boughman and Moss review vocal learning and development in mammals and birds, focusing on calls, which are usually simpler in structure than songs and often function in individual and group identification. They provide both extensive tables and discussions of case studies. One organizing concept is that vocal learning takes two distinctive forms. In learned acquisition, the new acoustic signals depend on auditory experience. In social modification, existing signals are modified in response to interactions with other individuals. The authors show that social modification, including the development of group signatures, is common in both birds and mammals, whereas learned acquisition is widespread only in birds.

The chapter by Ryan and Kime is a broad treatment of the selective forces acting on long-range acoustic signals in terrestrial environments. After considering energetic and physical constraints and several forms of environmental selection such as habitat acoustics, predation, and acoustic niche partitioning, the authors focus on factors that can influence the coevolution of signals—speciation, interspecific interactions, and various forms of sexual selection. They conclude with a clear explanation of the logic and data supporting the existence of pre-existing sensory biases or sensory exploitation, which can be considered alternatives to coevolutionary models. Here, rather large changes in signals might sometimes be favored because they tap inherent biases that arose and are maintained in other behavioral contexts such as prey or predator detection.

Yamaguchi and Kelley review hormonal mechanisms that influence acoustic communication both at a basic developmental level—sex determination and the concomitant sexual differences in communication behavior—and at the level of adaptive change within individuals. The chapter begins with a review of the known and putative mechanisms of action and then proceeds to review specific studies that relate hormones and communication behavior. The authors are particularly careful to distinguish between correlations and causation and lay out many important unsolved problems ripe for future research. One take-home message is that lower and higher vertebrates differ profoundly in the hormones that have short-term effects on acoustic communication and in the reversibility of developmental changes affected by hormones.

In addition to a thorough view of neuronal selectivity in auditory and vocal pathways, Gentner and Margoliash offer a spirited defense of the neuroethological method. They argue convincingly that designs based on the processing of natural signals in animals that specialize in this respect can yield important generalizations about sensory processing and signal recognition. Although touching on studies of other organisms, most of this detailed review concerns the songbird model system, which has many advantages for testing of general models about neural development and function, such as the motor theory of speech communication. I was also favorably

impressed by the ways in which these authors relate empirical data to ideas and theories about recognition (feature detectors versus distributed processing) and research strategies.

In summary, this book provides an excellent summary of the state-of-the-art in research concerned with both mechanisms and evolution of acoustic communication. My only concern is that its high price will put it out of reach of many graduate students who would benefit from the insights and suggestions for future research that are found in every chapter.

H. CARL GERHARDT
Division of Biological Sciences
University of Missouri, Columbia
Columbia, Missouri 65211

OBITUARIES

Robert C. Bilger • 1926–2002

Robert C. Bilger, a Fellow of the Acoustical Society of America, died peacefully in his sleep 26, December 2002 in Minneapolis, MN. Bob and his wife, Lynn, had been visiting their son Peter and his family for the Christmas holiday. Bob was a member of the Acoustical Society for several decades. He avidly attended the Society's meetings, regularly made scientific presentations, and from 1955 until the present, published many important papers in the Journal.

Bob was born (28 August 1926) and raised in Indiana and received his B.S. in 1949 from Purdue University after military service in WWII. He then earned M.S. and Ph.D. degrees at Purdue in 1951 and 1954, respectively. His undergraduate and graduate majors, which spanned engineering, general science, and experimental psychology, prepared him to conduct research on a wide variety of topics in acoustics. In 1954, Bob became a research associate at the Central Institute for the Deaf (CID) in St. Louis, an affiliation that proved important for him both professionally and personally. At CID, Bob worked with many investigators, including Ira Hirsh, Bruce Deatherage, Don Eldredge, and Carl Sherrick on several phenomena concerning auditory and visual masking and temporary threshold shift.

After spending 2 years as an assistant professor at the University of Michigan (1958–1960) Bob moved to the University of Pittsburgh, where he remained until 1977, when he moved to the Department of Speech and Hearing Science at the University of Illinois. While in Pittsburgh, Bob established a laboratory that embraced students, visitors, and colleagues whose areas of concentration/expertise were truly multidisciplinary and which included electrical and biomedical engineering, mathematical statistics, signal processing, and instrumentation. A hallmark of his laboratory was the creative application of the aforementioned areas to behavioral measures of auditory and speech perception in normal and hearing-impaired listeners. Beyond providing fundamental data and unique perspectives on the research problems being investigated, Bob and his students consistently provided fresh points of view that permitted a deeper understanding of the complexities involved in understanding how we hear complex sounds, including speech. This work culminated in the first systematic and detailed acoustic and psychoacoustic evaluation of the efficacy of cochlear implants. Bob was able to garner the competitive contract from the National Institutes of Health (NIH) to conduct the evaluation both because of the sophistication of the study that he proposed, including its impeccable methods, and because of the remarkable level of trust Bob enjoyed from both the psychoacoustic and medical communities. This trust was well placed in that Bob produced an intricate set of findings that was presented in an objective and even-handed manner that permitted the data and their implications to be evaluated within each community in the positive manner that they deserved. To Bob's credit that study has stood the test of time and proved to be prophetic in many respects.

After Bob moved to Illinois, he undertook another major project involving objective measurements of speech perception in noise in normal-hearing and hearing-impaired listeners. This work capitalized on earlier award-winning research in speech perception that Bob initiated at Pittsburgh and resulted in stimulus materials specially constructed for the measurement of speech presented either with or without contextual information. This program of research illustrates the unique abilities Bob was able to bring to bear on a very difficult problem because of the diverse nature of his background and collaborators.

While Bob, the scientist, made many contributions and received many deserved awards, Bob, the person, really was a most dominant presence for his family, friends, and colleagues. Bob was warm, loving, loyal (almost to a fault), full of mirth, forgiving of people, and able to remember any event or conversation that transpired over the last 50 years at any scientific meeting, whether he actually managed to attend the presentation or not. In fact, Bob was famous for ''holding court'' in the hallway just outside meeting rooms. It should also be mentioned that Bob was unique in several ways including the relative mismatch between the near discomfort he projected while presenting his work orally and the elegant prose he routinely used in his grant proposals, book chapters, and journal articles.

Those who attended Bob's poignant memorial service in Champaign, Illinois, heard several testimonials regarding how important he was to the personal, as well as the professional, lives of his students and colleagues. Bob was loved as well as respected. This was especially evident when his children related how wonderful and accessible a grandparent he was. Although Bob will be missed by all of us, he will not be forgotten and will remain a part of the folklore that makes membership in the Acoustical Society as cherished to us as it was to him.

CONSTANTINE TRAHIOTIS
LAWRENCE L. FETH
WALT JESTEADT

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:

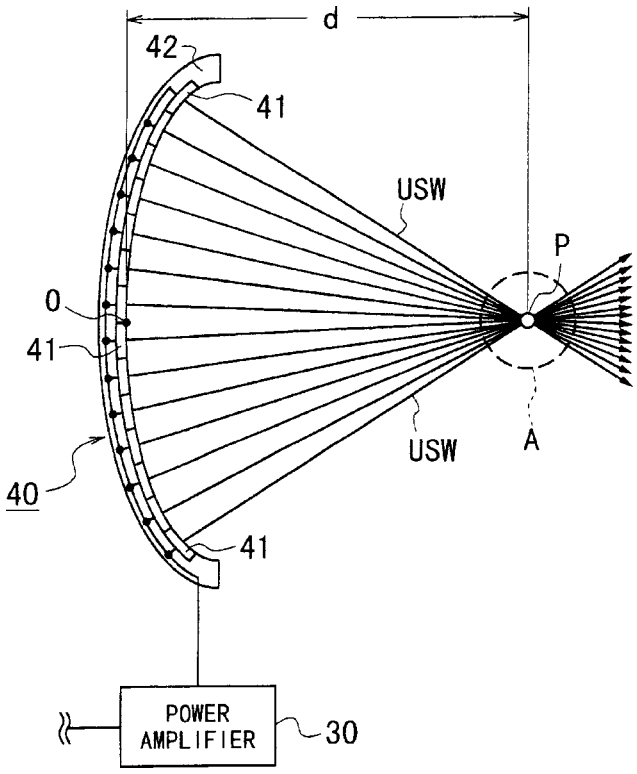
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
JOHN ERDREICH, *Ostergaard Acoustical Associates, 200 Executive Drive, West Orange, New Jersey 07052*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
HASSAN NAMARVAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
CARL J. ROSENBERG, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, Univ. of Rochester, Rochester, New York 14627*

6,542,054

43.35.Cg ACOUSTIC MIRROR AND METHOD FOR PRODUCING THE ACOUSTIC MIRROR

Robert Aigner *et al.*, assignors to Infineon Technologies AG
1 April 2003 (Class 333/187); filed in Germany 30 September 1999

A surface consisting of alternating thin layers of insulating material and metal, which is highly reflective of ultrasound, is produced by microelectronics manufacturing techniques.—EEU



6,538,360

43.35.Zc MULTIPLE FREQUENCY CLEANING SYSTEM

William L. Puskas, New London, New Hampshire
25 March 2003 (Class 310/316.01); filed 29 October 2001

Ultrasound is generated in a liquid cleaning bath by several transducers that produce sound in two or more nonoverlapping frequency bands. The frequencies are stepped very rapidly so that ultrasound from the new signals is induced in the bath before the sound from the earlier signals decays greatly. The result is said to be a combined ultrasound field with improved cleaning qualities.—EEU

6,556,687

43.38.Hz SUPER-DIRECTIONAL LOUDSPEAKER USING ULTRASONIC WAVE

Koji Manabe, assignor to NEC Corporation
29 April 2003 (Class 381/387); filed in Japan 23 February 1998

It is known that an array of ultrasonic transducers can be driven by a modulated carrier to produce audible sound from empty space. This patent suggests that if the transducers are mounted on a concave surface, then their combined energy can be focused at a specific point in space. Moreover, if the curvature is adjustable, then the focal point can be shifted. These two observations would seem to be self-evident, but they were expanded into 17 patent claims.—GLA

6,543,574

43.38.Ja METHOD OF MAKING A SPEAKER EDGE CONTAINING ISOCYANATE AND POLYOL

Sinya Mizone *et al.*, assignors to Inoac Corporation; Matsushita Electric Industrial Company, Limited
8 April 2003 (Class 181/171); filed in Japan 9 March 1999

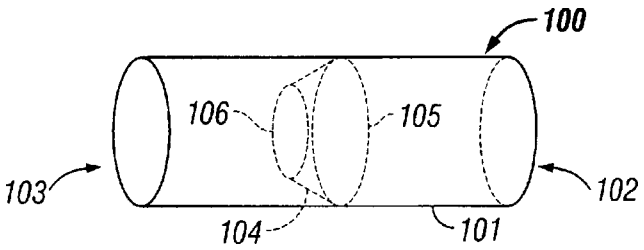
A half-roll loudspeaker cone suspension is usually formed from sheet stock molded to shape by heat and pressure. The patent describes an interesting alternative in which a liquid material is injected into a die cavity and then subjected to reaction foaming and curing—a little like baking a waffle.—GLA

6,549,637

43.38.Ja LOUDSPEAKER WITH DIFFERENTIAL FLOW VENT MEANS

Jon M. Risch, assignor to Peavey Electronics Corporation
15 April 2003 (Class 381/397); filed 24 September 1998

This patent includes just a little bit of everything, culminating in 40 fairly lengthy claims. However, the heart of the invention is the differential flow vent shown. This is an open-ended cylinder **100** fitted with funnel-shaped insert **104**. We are informed that air flowing from left to right will



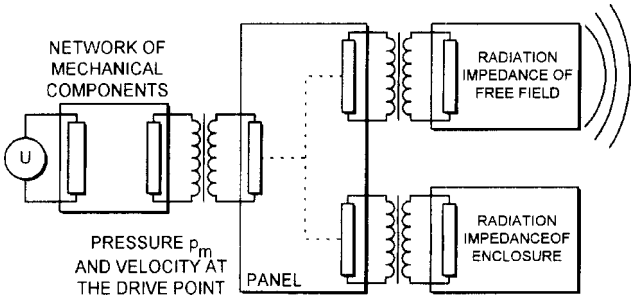
encounter more resistance than that flowing in the reverse direction. If two conventional vents in a woofer box are replaced by differential vents of opposite polarity, the result is forced air ventilation—benign turbulence, so to speak.—GLA

6,553,124

43.38.Ja ACOUSTIC DEVICE

Henry Azima and Joerg Panzer, assignors to New Transducers Limited
22 April 2003 (Class 381/345); filed in the United Kingdom 2 September 1995

This is an interesting patent that includes more than 20 pages of actual test results. The document is really a research paper followed by seven patent claims. The patent teaches that enclosing the rear radiation of a bending wave, panel-type loudspeaker results in a system that is acoustically



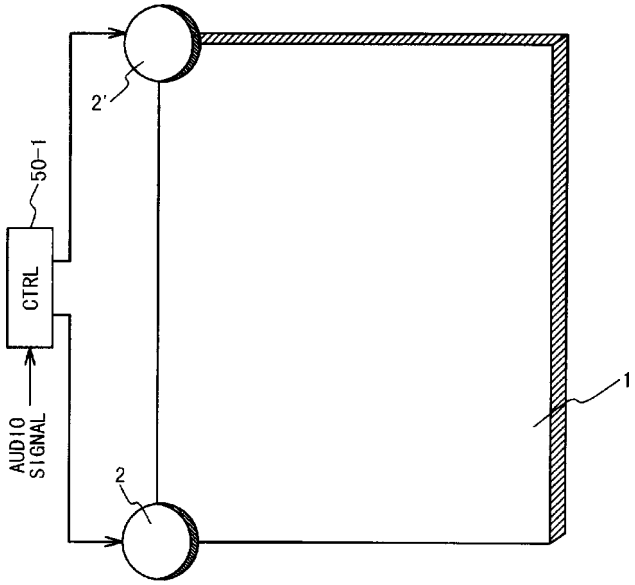
quite different from a standard closed-box loudspeaker. Some general rules for predicting and optimizing the performance of such a system are developed and explained.—GLA

6,554,098

43.38.Ja PANEL SPEAKER WITH WIDE FREE SPACE

Tatsumi Komura, assignor to NEC Corporation
29 April 2003 (Class 181/173); filed in Japan 15 June 1999

To save space, a panel-type loudspeaker diaphragm **1** can be driven at,



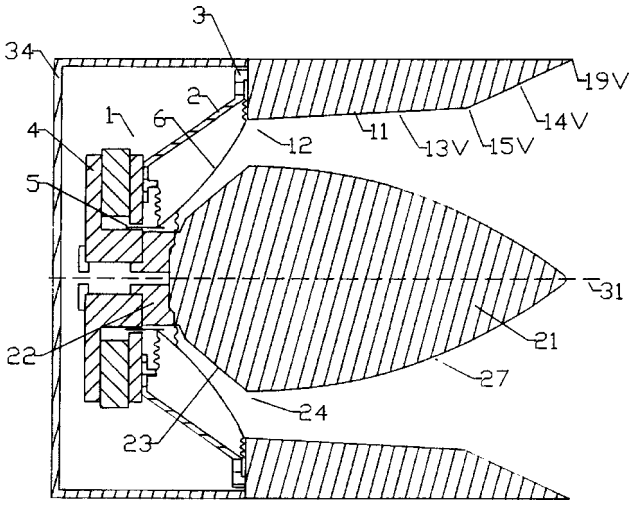
or very near, its outer edge. A single panel can be driven by more than one transducer **2**, **2′**.—GLA

6,557,664

43.38.Ja LOUDSPEAKER

Anthony John Andrews and John Newsham, both of Dorking, Surrey, the United Kingdom
6 May 2003 (Class 181/152); filed 22 February 1994

Central plug **21** is actually chisel-shaped, and surrounding horn **11** is similarly asymmetrical. The overall assembly is a close cousin to the JBL 2405 high-frequency transducer designed more than 25 years ago. In both



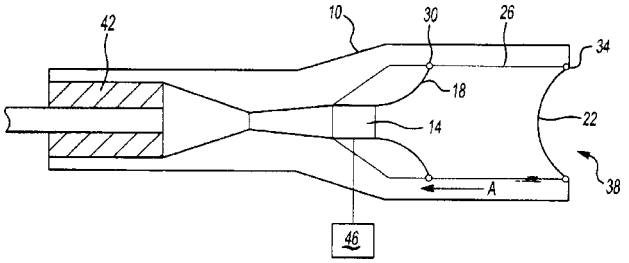
cases the objective is to create a coverage pattern that is relatively wide horizontally but vertically narrow.—GLA

6,557,665

43.38.Ja ACTIVE DIPOLE INLET USING DRONE CONE SPEAKER DRIVER

Richard D. McWilliam and Ian R. McLean, assignors to Siemens Canada Limited
6 May 2003 (Class 181/206); filed 16 May 2001

The invention is intended to provide active noise cancellation at the air intake of an internal combustion engine. Inner diaphragm **18** is electrically driven by conventional means **46**. Outer diaphragm **22** is driven acoustically to generate a noise attenuating signal. At the same time, air is somehow



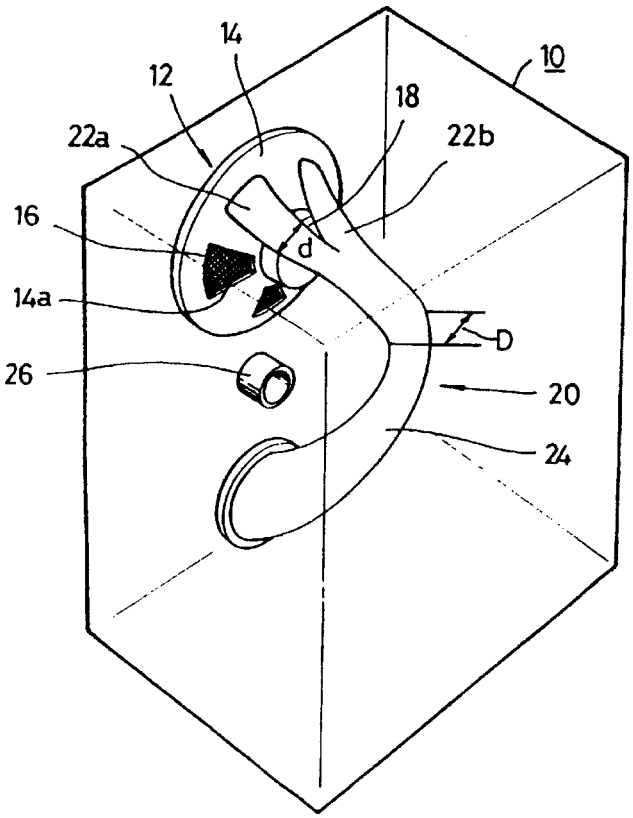
flowing from mouth **38** in direction **A** through passageway **14** into the engine even though both diaphragms are provided with seals **30**, **34**. The patent claims do not clarify the arrangement.—GLA

6,560,343

43.38.Ja SPEAKER SYSTEM

Jae-Nam Kim, assignor to Samsung Electronics Company, Limited
6 May 2003 (Class 381/349); filed in the Republic of Korea 22 April 1996

Part of the backwave energy from loudspeaker **16** is conducted through horn **24** to the face of cabinet **10**. The remainder energizes vent **26**. The patent explains that since only a portion of the rear sound waves are collected and amplified, ''...reflected waves or standing waves will not be



generated in a sound wave amplifying horn to increase amplification efficiency of bass sounds and improve the clearness of the resulting sounds.''—GLA

6,549,632

43.38.Kb MICROPHONE

Hiroshi Akino *et al.*, assignors to Kabushiki Kaisha Audio-Technica
15 April 2003 (Class 381/174); filed 19 March 1997

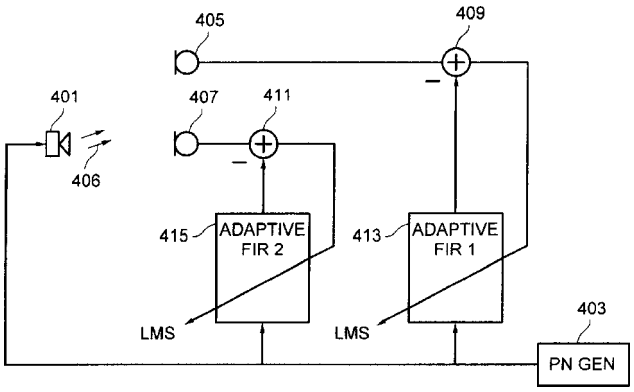
Some hand-held microphones are extremely sensitive to mechanical shocks and scrapes. This patent describes a simple, passive shock isolation system derived from mechanical analog circuit analysis. Although the patent text describes embodiments for both dynamic and capacitor microphones, all of the eight patent claims refer to capacitor microphones only.—GLA

6,549,627

43.38.Lc GENERATING CALIBRATION SIGNALS FOR AN ADAPTIVE BEAMFORMER

Jim Rasmusson *et al.*, assignors to Telefonaktiebolaget LM Ericsson
15 April 2003 (Class 381/71.11); filed 30 January 1998

Consider a hands-free communications system installed in a vehicle. The equipment includes two or more microphones **405**, **407** and a loudspeaker **401**. By introducing adaptive filters at the outputs of individual microphones it is possible to achieve in-phase summation of the signals from the direction of a talker while largely canceling the unwanted signals from the loudspeaker. Because the acoustical environment and the location



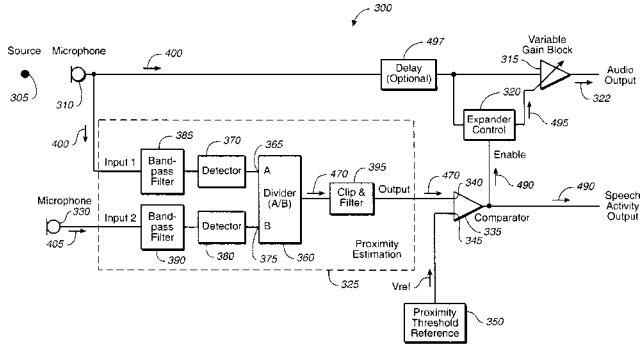
of the talker may change, the system must somehow calibrate itself. The patent describes an improved one-step calibration process that may also be augmented by utilizing the filters as fixed echo cancellers during normal operation.—GLA

6,549,630

43.38.Lc SIGNAL EXPANDER WITH DISCRIMINATION BETWEEN CLOSE AND DISTANT ACOUSTIC SOURCE

James F. Bobisuthi, assignor to Plantronics, Incorporated
15 April 2003 (Class 381/94.7); filed 4 February 2000

The objective is to reliably turn on a microphone when its user speaks and to minimize false triggering from other sound sources. A handset or headset is fitted with two microphones, one **310** near the talker's mouth and the other **330** as far as possible from the first. Their outputs are filtered,



rectified, and then divided—not compared. The patent explains in some detail how this arrangement estimates source proximity rather than signal-to-noise ratio. If the proximity estimation signal exceeds a predetermined threshold, then microphone **310** is gated on or its gain raised.—GLA

6,570,079

43.38.Md RECORDING AND REPRODUCING APPARATUS, RECORDING AND REPRODUCING METHOD, AND DATA PROCESSING APPARATUS

Shinichi Fukuda, assignor to Sony Corporation
27 May 2003 (Class 84/602); filed in Japan 19 February 1998

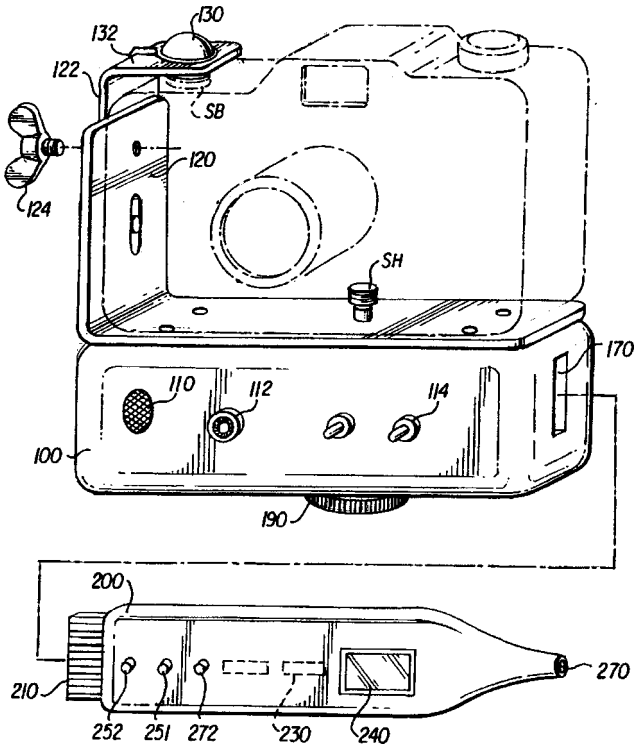
Sony let the examiner do the legwork on this one. Suppose an audio CD owner wishes to make a copy. If a lossy copy is chosen, then the user is stuck with just a bad fidelity copy but not a bill. If the high-fidelity copy is chosen, then the ''accounting system'' is involved and the owner is charged a fee. The patent doesn't address any of the wider issues regarding music distribution.—MK

6,574,441

43.38.Md SYSTEM FOR ADDING SOUND TO PICTURES

John W. McElroy, Newtown Square, Pennsylvania
3 June 2003 (Class 396/312); filed 4 June 2001

Still pictures not only lack motion—they also lack sound. The inventor



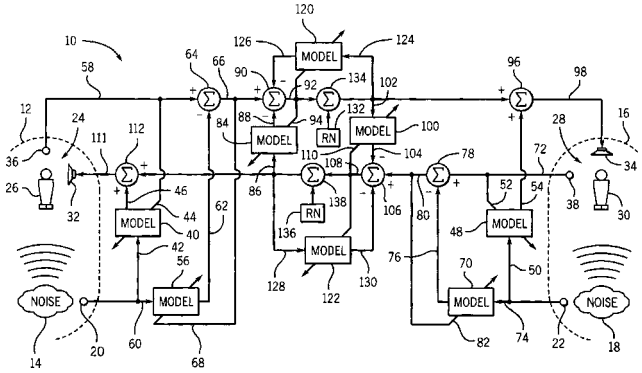
proposes adding a sound recorder **100** that eventually produces a bar coded sound strip.—MK

6,549,629

43.38.Tj DVE SYSTEM WITH NORMALIZED SELECTION

Brian M. Finn and Shawn K. Steenhagen, assignors to Digisonix LLC
15 April 2003 (Class 381/92); filed 21 February 2001

DVE stands for digital voice enhancement which, in this case, includes echo cancellation, background noise suppression, and optimal selection of multiple-zone microphones in a hands-free communications system.



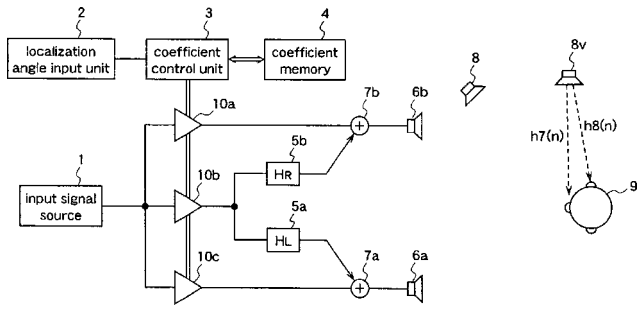
Although the procedure is too complicated to describe in a few sentences, it is clearly explained in the patent. Anyone interested in the field will find the patent informative.—GLA

6,546,105

43.38.Vk SOUND IMAGE LOCALIZATION DEVICE AND SOUND IMAGE LOCALIZATION METHOD

Takashi Katayama *et al.*, assignors to Matsushita Electric Industrial Company, Limited
8 April 2003 (Class 381/17); filed in Japan 30 October 1998

Using head-related transfer functions (HRTFs) to create virtual sound sources from a pair of loudspeakers is theoretically intriguing but messy in practice. Assuming that some kind of all-purpose HRTFs can be derived, then FIR filter coefficients can be calculated for any angular location. However, computing and/or storing filter coefficients for all possible locations is inefficient and time-consuming. This patent, like a number of earlier inven-



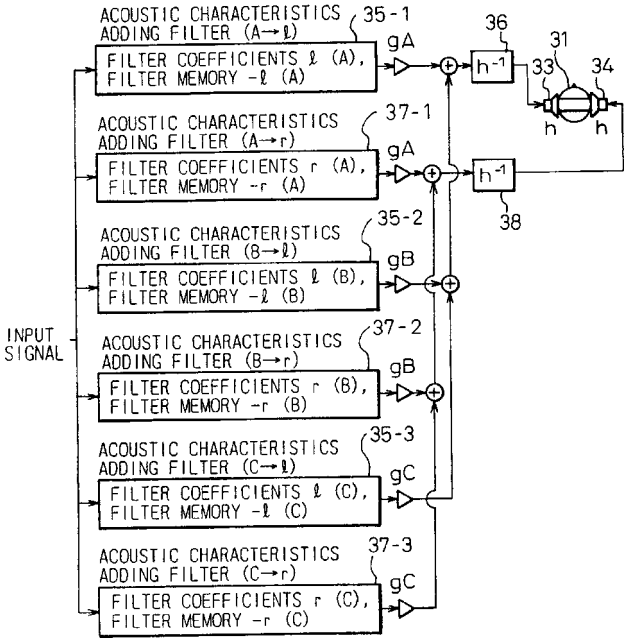
tions, attempts to find a better way. In this case, the angular location of a virtual source is fed to a coefficient control device which then performs digital mathematical operations involving only three predetermined frequency response functions. The process is said to result in a dramatic reduction in memory requirements and computational time, as compared to prior art.—GLA

6,553,121

43.38.Vk THREE-DIMENSIONAL ACOUSTIC PROCESSOR WHICH USES LINEAR PREDICTIVE COEFFICIENTS

Naoshi Matsuo and Kaori Suzuki, assignors to Fujitsu Limited
22 April 2003 (Class 381/17); filed in Japan 8 September 1995

To create convincing three-dimensional audio for computer games, a number of virtual sound sources must be controlled within a virtual sound field. At the same time, the acoustical characteristics of the actual reproduc



ing sound field must be subtracted from the sound source. A bank of large FIR filters seems to be called for, but the patent suggests a more efficient approach. By performing linear predictive analysis of the impulse response of the sound field to be added, the number of taps can be greatly reduced. A similar procedure can be applied to sound sources in motion—in effect, panning between locations rather than creating a plurality of individual sources. The patent is clearly written and includes a great many helpful illustrations.—GLA

6,526,835

43.40.At APPARATUS AND METHOD FOR CHARACTERIZING PHYSICAL PROPERTIES OF A TEST PIECE

Richard Todd Hage, assignor to Andersen Corporation
4 March 2003 (Class 73/778); filed 6 June 2001

The test piece here consists of a slender beam. The method described in the patent in essence involves observing the natural period and the logarithmic decrement of the test beam in freely decaying vibration and using these measured parameters in relations obtained from the classical expressions applicable to ideal vicously damped beams.—EEU

6,532,818

43.40.Le METHOD AND APPARATUS FOR MEASURING A VIBRATIONAL CHARACTERISTIC OF A GOLF CLUB SHAFT

Jeffrey A. Blankenship, assignor to Karsten Manufacturing Corporation
18 March 2003 (Class 73/579); filed 16 April 2001

In this apparatus for determining the stiffness of the shaft of a golf club, one end of the shaft is clamped in a rigid fixture and a two-axis accelerometer is attached to the other. The shaft is deflected and released and the resulting vibration signals are sent to a computer, which determines

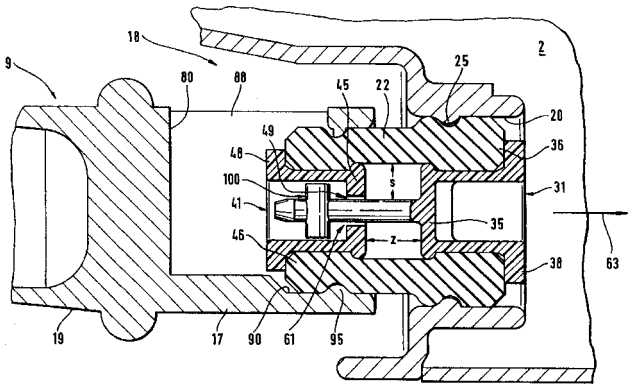
the shaft's stiffness from the measured natural frequencies. Different stiffnesses in different lateral directions are also determined from these signals.—EEU

5,699,865

43.40.Tm ANTIVIBRATION DEVICE FOR MOUNTING BETWEEN MOTOR UNIT AND A HANDLE UNIT

Karl Forderer, Schwaikheim, Germany *et al.*, assignors to Andreas Stihl
23 December 1997 (Class 173/162.2); filed in Germany 21 August 1995

This vibration isolator uses a concentric elastomeric ring **22** between the motor **2** and the handle **9** of a chain saw. The somewhat flexible ring is



held in place by plugs **38** and **48** inserted into each end of the ring and held by retainer device **100**.—JE

6,530,563

43.40.Tm MULTI-AXIS SHOCK AND VIBRATION ISOLATION SYSTEM

Thomas Joseph Miller and Daniel Charles Radice, assignors to Enidine, Incorporated
11 March 2003 (Class 267/136); filed 10 July 2001

The item that is protected is mounted in a frame via a series of wire rope isolators and a pair of vertically oriented double-acting shock absorbers. The ends of these shock absorbers are provided with resilient bushings in order to permit the item to move relatively freely in the horizontal plane.—EEU

6,533,256

43.40.Tm HIGH PRECISION VIBRATION DAMPING SYSTEM

Peter A. Warren *et al.*, assignors to Foster-Miller, Incorporated
18 March 2003 (Class 267/136); filed 23 June 2000

A vibration damping device that maintains the precision alignment of an optical system component consists of a flat plate contained in a housing with walls that are parallel to the plate's surfaces. Parallelism between the plate and housing surfaces is maintained and damping is provided by a series of buttonlike elements that are rigidly attached to one surface and free to slide along the opposing surface. The sliding surface of these elements is spherical and friction is controlled by selection of the spherical radius, the preloading, and the effective friction coefficient. The payload is attached to a member that extends through an opening in the housing.—EEU

6,535,458

43.40.Tm METHOD AND APPARATUS FOR SUPPRESSING DRILLSTRING VIBRATIONS

Richard Meehan, assignor to Schlumberger Technology Corporation
18 March 2003 (Class 367/81); filed in the United Kingdom 9 August 1997

Drilling for hydrocarbon or geothermal exploration typically is accomplished via a drillstring that consists of an assembly of tubes. In some devices it is desired to transmit information to the surface from down the hole along the walls of these tubes. In order to reduce the noise in these signals it is useful to limit them to selected frequencies. Filtering of the transmitted frequencies is accomplished by including in the drillstring a series of sections whose impedances differ from those of their adjacent sections, so that there results a series of stop bands. This patent describes an impedance element that consists of a tube with internal and external circumferential grooves of various depths and lengths.—EEU

6,536,566

43.40.Tm VIBRATION-DAMPING DEVICE FOR VEHICLES

Koichi Hasegawa *et al.*, assignors to Tokai Rubber Industries, Limited
25 March 2003 (Class 188/378); filed in Japan 1 March 2000

This patent describes a variety of dampers consisting of masses that are free to rattle inside close-fitting cavities. The masses may be coated with rubber or provided with elastomeric caps or rings.—EEU

6,536,953

43.40.Tm BEARING MOUNT SYSTEM FOR REDUCING VIBRATION

Steven A. Cope *et al.*, assignors to E. I. du Pont de Nemours and Company
25 March 2003 (Class 384/536); filed 8 November 2000

The exterior of a ball bearing's raceway is separated from the adjacent shaft via a relatively thin circumferential layer of elastomeric material, retained in a suitable groove in the shaft. The elastomeric material is thicker near its edges than in the middle, so that only the edges are compressed in the presence of small relative excursions, whereas the entire width of the material is compressed in the presence of excursions that exceed a given value, resulting in a favorable bilinear stiffness characteristic.—EEU

6,538,852

43.40.Tm BEARING DEVICE WITH DAMPING MECHANISM

Harushige Osawa and Hongbing Du, assignors to Nidec Corporation
25 March 2003 (Class 360/265.2); filed in Japan 25 December 1998

The device described here is applicable to mechanisms in which a shaft needs to be stopped rapidly in a definite position, as in hard disc drives. The device in essence combines a conventional ball bearing with an additional race that houses small masses in somewhat oversized cavities. Sudden stopping of the shaft causes these masses to traverse the free spaces in the cavities, and thus to induce damping via friction and impacts.—EEU

6,533,257

43.40.Vn COMPOSITE VIBRATION DAMPING SYSTEM

Arthur E. Clark, assignor to The United States of America as represented by the Secretary of the Navy
18 March 2003 (Class 267/140.15); filed 21 June 2001

Several damping devices involving magnetostrictive and piezoelectric materials are attached to a vibrating structure. All of these devices, which can act as generators of electricity, are connected to a distributor, which routes the current to electrical dissipators. A controller also can cause the distributor to send current to the piezoelectric devices, so that these act as actuators. The controller is programmed to minimize vibrations without the use of energy from an external source.—EEU

6,536,735

43.40.Vn VIBRATION ISOLATING APPARATUS FOR TABLE FOR MOUNTING DEVICE SENSITIVE TO VIBRATIONS AND METHOD THEREFOR

Takahide Haga *et al.*, assignors to Ebara Corporation
25 March 2003 (Class 248/550); filed in Japan 10 October 2000

Active isolation of a rigid table in six degrees of freedom is accomplished by use of a number of sensors and actuators and of a relatively simple controller. The isolation system design is based on the assumption that the center of action of the resilient supports coincides with the table's center of gravity and it takes account of the coordinates of the sensors and actuators in the plane of the table.—EEU

6,539,806

43.40.Yq FLUID-LOAD MEASUREMENT BY MAGNETIC EXCITATION AND VIBRATION SENSING OF A FLUID-LOAD-SENSITIVE DIAPHRAGM

Hubert A. Wright, assignor to Starr-Johnson
1 April 2003 (Class 73/703); filed 7 March 2001

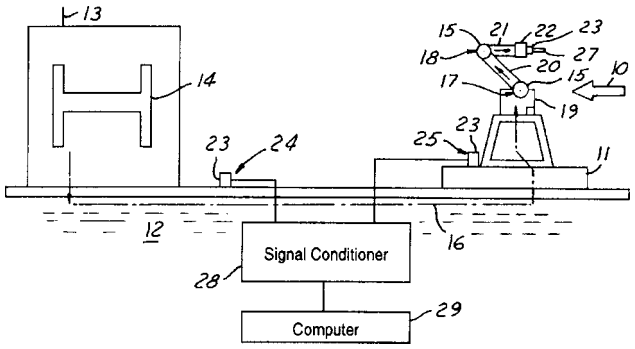
This device is intended for measuring the static pressure and/or vibration in a fluid. The flexible diaphragm that is in contact with the fluid carries a permanent magnet, which is surrounded by a coil that is attached to the device's base. The membrane's displacement is determined from measurement of the coil's inductance.—EEU

6,575,034

43.40.Yq CHARACTERIZATION OF ENVIRONMENTAL AND MACHINERY INDUCED VIBRATION TRANSMISSIVITY

Brij B. Seth *et al.*, assignors to Ford Global Technologies, L.L.C.
10 June 2003 (Class 73/579); filed 30 January 2001

This system, intended to assist in the design, building, and installation of dynamic manufacturing machinery, provides characterization of environmental and machinery vibrations by a transmitivity discriminator. Vibrational acceleration signals are monitored and captured at source site locations, which can be simply ground soil or a foundation through which environmental vibrations pass, and a target site, which can be an intended machinery foundation site or one or more sites containing interfacing joints between elements of the machinery to be placed on this foundation.



Frequency and amplitude data extracted from the signals determine displacement excitations at these locations. Transmitivity ratios along the vibration path between the source and the target site are obtained by summing the displacement excitations at different frequencies at each of the locations and then dividing the target site sum by the source site sum, thereby indicating the relative vibrational stiffness between the sites. These transmitivity ratios can then be applied to modify the design of the proposed foundation and/or the design of the interfacing machine elements.—DRR

6,108,431

43.50.Gf LOUDNESS LIMITER

Herbert Bachler, Meilen, Switzerland, assignor to Phonak, AG
22 August 2000 (Class 381/312); filed 1 October 1996

This patent describes circuitry to limit loudness instead of intensity.—JE

6,564,900

43.50.Gf METHOD AND APPARATUS FOR REDUCING ACOUSTIC NOISE IN MRI SCANNERS

David Dean *et al.*, assignors to GE Medical Systems Global Technology Company, LLC
20 May 2003 (Class 181/202); filed 22 November 2000

This technique for mitigating acoustic noise generated by a magnetic resonance imaging system involves installing energy damping elements in the scanner, principally within the gradient coil assembly, between the gradient coil assembly and the primary magnet, and around the outer and inner peripheries of the primary magnet and rf coil. The elements may be shaped to conform to the structure of the scanner and may be tilelike elements that can be retrofitted to existing equipment or installed in new scanners. The tiles may be fabricated from a range of materials and composites.—DRR

6,564,902

43.50.Gf DEVICE AND METHOD FOR A SOUND-ATTENUATING UNIT

Josef Saberi, assignor to Volvo Personvagnar AB
20 May 2003 (Class 181/237); filed in Sweden 14 November 1997

This is a muffler for reducing sound from a flowing gas stream. Two flow paths are provided within the sound-attenuating unit and a switchover device is provided to alternatively guide the gas stream along these two flow paths. A pressure detector and crossover switch constitute an adjustable throttle for blocking one flow path when the pressure falls below a predetermined value, thereby directing the flow to the other path. The throttle is adapted to open when the pressure exceeds a predetermined value. This setup is said to result in effective sound absorption, small mounting volume of the muffler system, and a low backpressure at high engine speeds.—DRR

6,575,696

43.50.Gf METHOD OF SOUND ATTENUATION IN CENTRIFUGAL BLOWERS

Leslie A. Lyons and David A. Fisher, assignors to Fasco Industries, Incorporated
10 June 2003 (Class 415/119); filed 11 September 2001

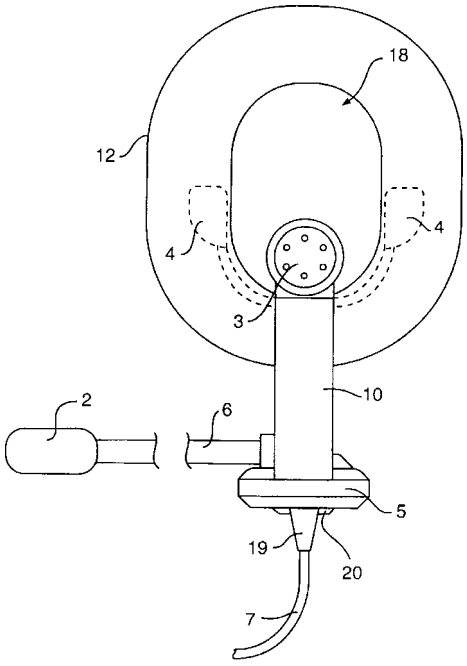
This version of a centrifuge blower, designed to attenuate the noise caused by blade pass, includes a blower housing that encloses a motor-driven impeller in order to produce a flow of air out of an outlet opening formed by the blower housing. When the impeller rotates, the blades pass by a cutoff formed along the interior of the housing between the scroll section of the housing and an integrated exhaust section. An angled cutoff portion is formed along the intersection between the tubular exhaust section and the scroll section in order to disrupt the pressure fluctuations produced by the orientation of the cutoff and the impeller blades. In addition to the angled cutoff portion, the housing incorporates at least one sound cavity that extends outward from the outer wall of the exhaust section. This sound cavity forms an open space that is said to attenuate the sound generated by the rotating impeller.—DRR

5,970,155

43.50.Hg HEADSET FOR HEARING PROTECTORS

Kal Leppalahti, Kauniainen, Finland, assignor to Kitek Oy Ab Insinooritoimisto
19 October 1999 (Class 381/72); filed 14 February 1997

This headset includes an earphone **3** and microphone **2** which may be



attached with clip **4** to a circumaural earmuff.—JE

6,068,079

43.50.Hg ACOUSTIC VALVE CAPABLE OF SELECTIVE NON-LINEAR FILTERING OF SOUND

Pascal Hamery, Mulhouse, France *et al.*, assignors to L. S. L. Institut Franco-Allemand de Recherches de Saint-Louis
30 May 2000 (Class 181/135); filed 11 August 1997

This variable acoustic filter depends on the relative orientation between two rotatable discs. The filter can be incorporated into an earplug.—JE

6,070,693

43.50.Hg HEARING PROTECTOR AGAINST LOUD NOISE

Pascal Hamery, Mulhouse, France, assignor to Institut Franco-Allemand de Recherches de Saint-Louis
6 June 2000 (Class 181/135); filed 20 January 1999

This is an insert hearing protector reportedly able to attenuate noise up to 190 dB. Let's see. Bone conduction through the skull, −50 dB, leaves 190 dB−50 dB=140 dB exposure. That's not much protection.—JE

6,074,060

43.50.Hg EYESIGHT AND HEARING SAFETY APPARATUS

Joe A. Bruce, Beaumont, Texas
13 June 2000 (Class 351/158); filed 7 October 1999

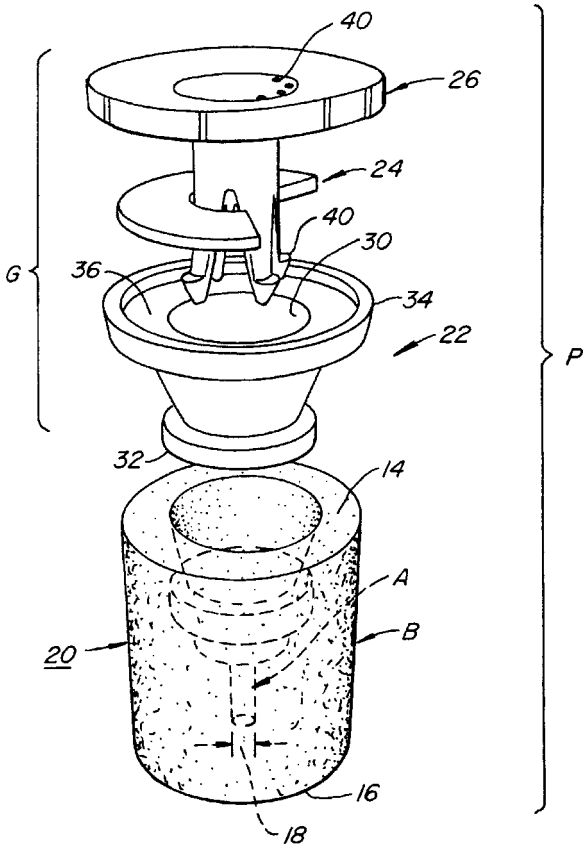
Cut the cord in half on a pair of hearing protectors and glue each end to the temples of a pair of safety glasses and you have this invention.—JE

6,082,485

43.50.Hg ADJUSTABLE EARPLUG

Eric B. Smith, San Francisco, California
4 July 2000 (Class 181/135); filed 10 August 1999

An earplug with adjustable element **26**, **24** inserted into a foam plug



provides adjustable treble attenuation.—JE

6,095,146

43.50.Hg GLOW-IN-THE-DARK HEARING PROTECTIVE DEVICES

Richard Knauer, Indianapolis, Indiana *et al.*, assignors to Aearo Company
1 August 2000 (Class 128/864); filed 20 July 1999

The title says it all except, ''Why?'' The inventors presuppose a need to see who is wearing hearing protection in the dark.—JE

6,119,805

43.50.Hg HEARING PROTECTOR ADAPTABLE TO CHAIR

Urban Eriksson, Stockholm, Sweden
19 September 2000 (Class 181/129); filed 23 March 1999

This is a ''U''-shaped hearing protector mounted on a floor stand to be used by musicians. It must do wonders for the sense of ensemble in an orchestra.—JE

6,123,168

43.50.Hg BANDED HEARING PROTECTOR

Bengt Goran Berg and Hans Peter Jorgen Høkansson, both of Tyringe, Sweden, Assignors to Dalloz Safety AB
26 September 2000 (Class 181/129); filed 30 June 1997

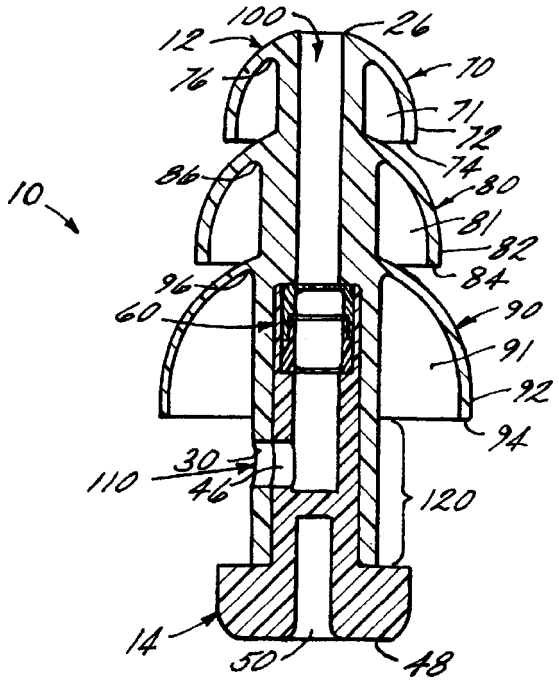
This protector incorporates a hinge in the middle of each side of the over-the-head band to permit swinging the device away from the ear.—JE

6,148,821

43.50.Hg SELECTIVE NONLINEAR ATTENUATING EARPLUG

Robert N. Falco, Indianapolis, Indiana, assignor to Cabot Safety Intermediate Corporation
21 November 2000 (Class 128/864); filed 29 April 1998

This is a variable attenuation hearing protector. The device limits acoustic energy by changing the percentage of open passageway **110**. The patent is similar to United States patent 6,286,622.—JE



6,151,717

43.50.Hg TRANSPARENT OR TRANSLUCENT EARMUFF CUP

Fredrik Lindgren, Spencer, Massachusetts *et al.*, assignors to Aearo Company
28 November 2000 (Class 2/209); filed 20 July 1999

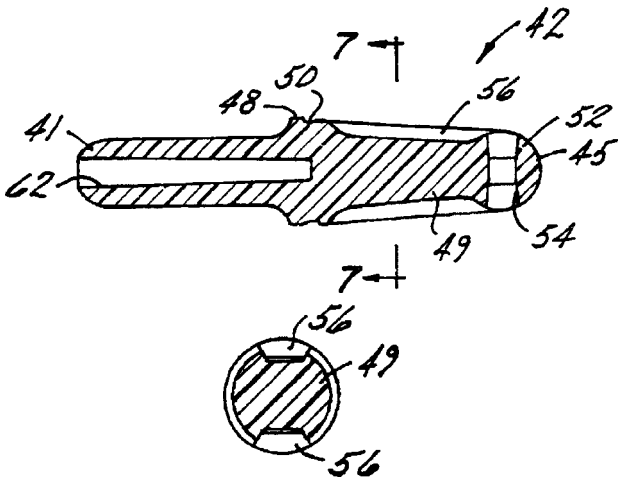
This circumaural hearing protector cup allows visual inspection of a hearing protection device simultaneously worn in the ear.—JE

6,241,042

43.50.Hg HEARING PROTECTIVE DEVICE

Robert N. Falco, assignor to Cabot Safety Intermediate Corporation
5 June 2001 (Class 181/135); filed 14 March 2000

This hearing protector is manufactured by forming foam around stem



42 during manufacture.—JE

6,256,396

43.50.Hg SELF-FITTING HEARING PROTECTION EARPLUG WITH FACILE INSERTION MECHANISM

William Bradford Cushman, Pensacola, Florida
3 July 2001 (Class 381/328); filed 26 May 1995

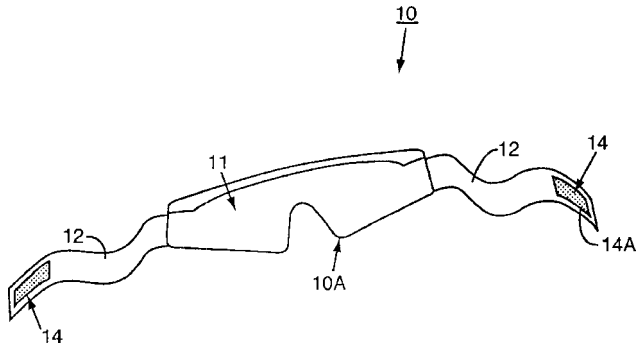
This hearing protector includes a mechanism to elongate and reduce diameter for easy insertion.—JE

6,481,846

43.50.Hg EAR AND EYE PROTECTION APPARATUS

Erik I. Mikysa, Oklahoma City, Oklahoma
19 November 2002 (Class 351/158); filed 4 August 2000

This patent describes a method of attaching safety glasses to a muff-



type hearing protector.—JE

6,237,947

43.50.Qp DEVICE AND METHOD FOR AVOIDING HEARING DAMAGE DURING ACTIVATION OF VEHICLE OCCUPANT RESTRAINT SYSTEMS

Armin Kausch, assignor to TRW Occupant Restraint Systems GmbH & Company KG
29 May 2001 (Class 280/735); filed in Germany 11 November 1996

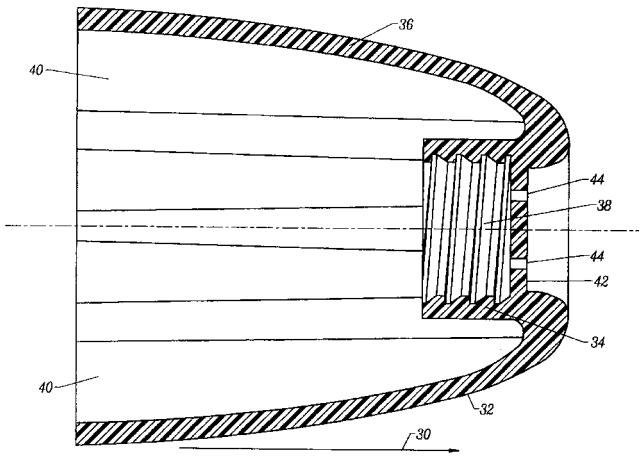
This patent describes the use of an airbag predetonation signal to elicit the acoustic reflex which mitigates sound transmitted to the cochlea.—JE

6,129,174

43.50.Yw MINIMAL CONTACT REPLACEABLE ACOUSTIC COUPLER

Joseph Brown, Valley Center, California *et al.*, assignors to Decibel Instruments, Incorporated
10 October 2000 (Class 181/135); filed 30 December 1998

This is a device for in-the-ear hearing aids. A picture is worth



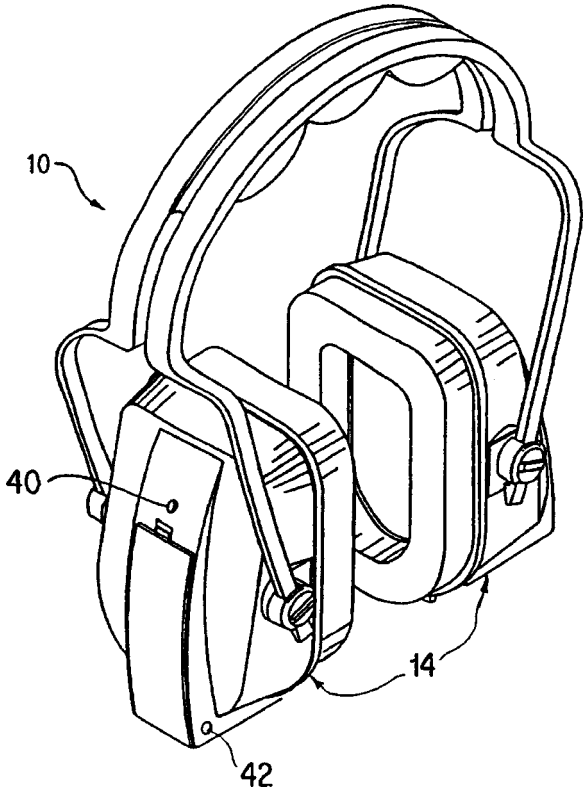
1000 words.—JE

6,456,199

43.50.Yw CONTINUOUS NOISE MONITORING AND REDUCTION SYSTEM AND METHOD

Kevin Michael, assignor to doseBusters USA
24 September 2002 (Class 340/573.1); filed 16 February 2001

This is a system for monitoring the noise dose under hearing protec-



tors. The device incorporates an indicator **40** of approaching overexposure or a digital indicator of dose.—JE

6,568,142

43.55.Ti BAMBOO FLOOR PLATE FOR SOUND INSULATION

Seiji Yoshida, assignor to Japan Blower Ind. Company, Limited
27 May 2003 (Class 52/582.1); filed in Japan 31 January 2000

Round bamboo cane is cut to lay flat and then laminated to form a plate. Grooves are cut into the plate to improve sound isolation performance.—CJR

6,565,107

43.58.Wc MOTORBIKE SOUND SIMULATOR

Lawrence B. Hartman, Bellaire, Texas
20 May 2003 (Class 280/288.4); filed 4 December 2001

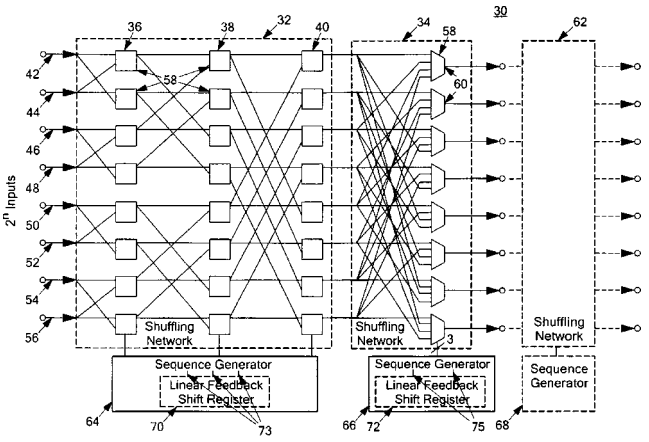
Like United States Patent 6,394,875 [reviewed in J. Acoust. Soc. Am. **112**(6), 2518 (2002)], playing cards aren't good enough for faux motobicycle sounds. The inventor proposes replacing a playing card with a stiff plastic strap. This ruins all the fun.—MK

6,570,521

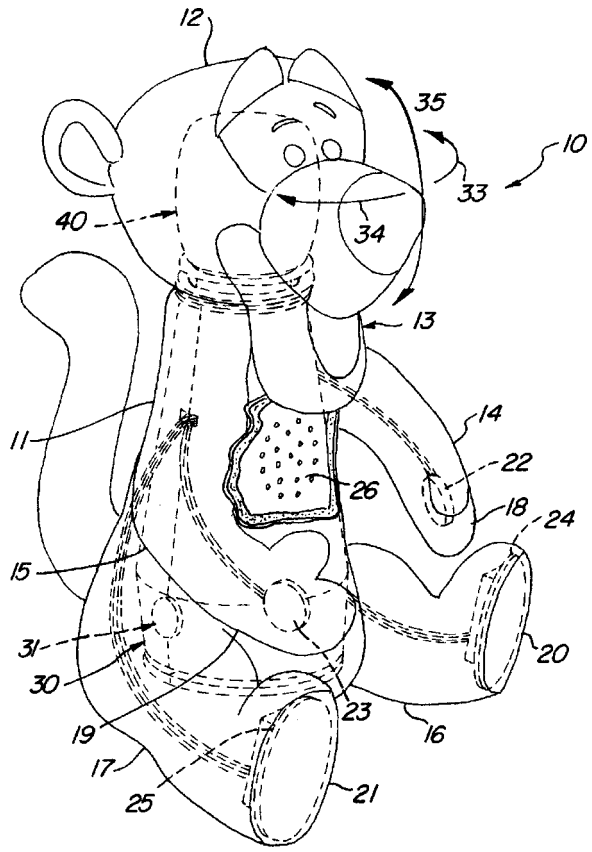
43.60.Cg MULTISTAGE SCRAMBLER FOR A DIGITAL TO ANALOG CONVERTER

William G. J. Schofield, assignor to Analog Devices, Incorporated
27 May 2003 (Class 341/144); filed 27 December 2001

It is well known that the nonlinear transfer function of digital to analog converters leads to harmonic distortion. Here, the inventor proposes using a



multistage network that scrambles data bits with a linear feedback shift register so that the error is whitened.—MK



Depending on the signal, a motor located in the body swivels the head from left to right to face the speaker. Bouncing is not specified.—MK

6,574,596

43.60.Qv VOICE RECOGNITION REJECTION SCHEME

Ning Bi *et al.*, assignors to Qualcomm Incorporated
3 June 2003 (Class 704/249); filed 8 February 1999

A voice recognition rejection scheme is described which applies an *N*-best algorithm to compare the input words with stored words and which tests predefined relationships between one or more closest comparison results.—HHN

6,565,407

43.66.Qp TALKING DOLL HAVING HEAD MOVEMENT RESPONSIVE TO EXTERNAL SOUND

John Woolington and Virgil Wulff, assignors to Mattel, Incorporated
20 May 2003 (Class 446/175); filed 2 February 2000

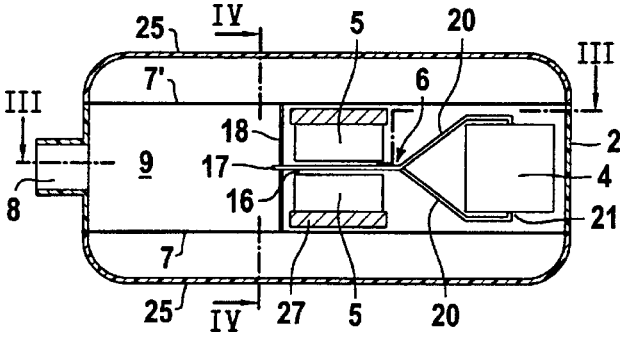
A plush toy can exhibit binaural hearing. The left microphone **31** is connected to a comparator with the right microphone (not shown).

6,563,933

43.66.Ts ELECTROMAGNETIC TRANSDUCER FOR GENERATING SOUND IN HEARING AIDS, PARTICULARLY ELECTRONIC HEARING AIDS

Torsten Niederdraenk, assignor to Siemens Audiologische Technik GmbH
13 May 2003 (Class 381/417); filed in Germany 15 November 1999

High-gain hearing aids, normally using receivers (speakers) with one diaphragm, are frequently unable to achieve desired output levels due to instability caused by mechanical vibrations produced by the receiver getting back to the microphone. To reduce this mechanical feedback proglem



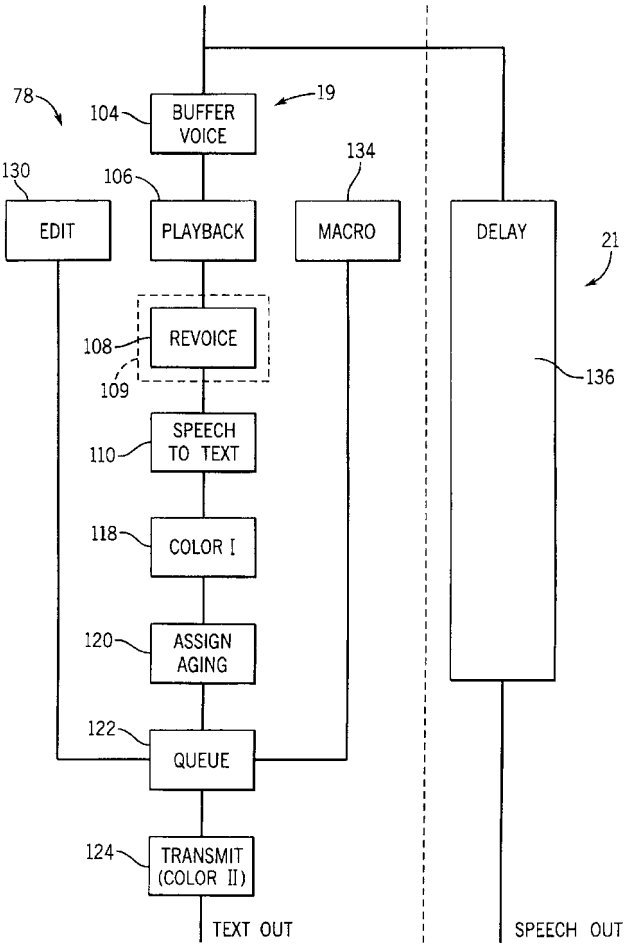
and produce more usable acoustic gain, a hearing aid receiver contains two identical diaphragms placed symmetrically on opposite sides of the drive mechanism. The diaphragms are driven equally but in opposite directions so as to at least partially cancel the mechanical vibrations caused during the drive mechanism movements.—DAP

6,567,503

43.66.Ts REAL-TIME TRANSCRIPTION CORRECTION SYSTEM

Robert M. Engelke *et al.*, assignors to Ultratec, Incorporated
20 May 2003 (Class 379/52); filed 20 February 2001

The present device relates to systems for transcribing voice communications into text and, more specifically, to means of facilitating higher accuracy, real-time editing of a transcribed text stream by a human call assistant. The device permits the call assistant to select words for editing based on their screen location, most simply by touching the word on the screen.



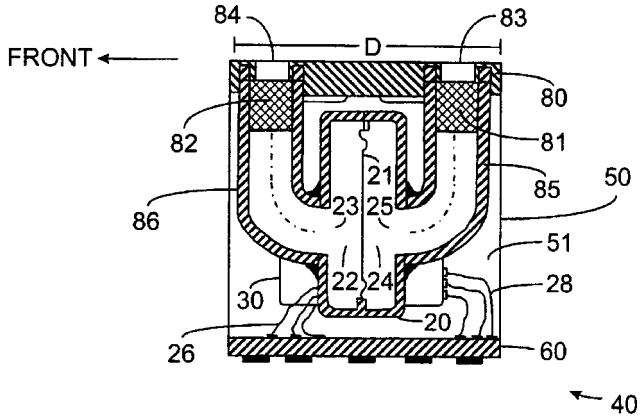
Lines of text are preserved intact as they scroll off the screen to assist in tracking individual words. Words on the screen change color to indicate their status for editing and transmission. The delay before transmission of transcribed text may be adjusted, for example, based on dynamical error rates, perceptual rules, or call assistant or user preference.—DRR

6,567,526

43.66.Ts DIRECTIONAL MICROPHONE ASSEMBLY

Mead C. Killion *et al.*, assignors to Etymotic Research, Incorporated
20 May 2003 (Class 381/313); filed 13 October 2000

A hearing aid microphone capsule contains both a first-order directional element and an omnidirectional element, their acoustical coupling tubes and electrical equalization to provide the same frequency response in directional and omnidirectional modes. Acoustic resistors and inertances are utilized in the front and rear sound passages, respectively, for the directional



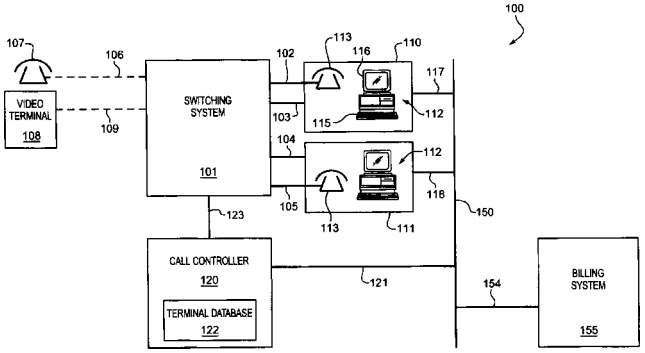
microphone to provide the proper acoustic delays for good directivity across the frequency range while leveling the frequency response. The electrical equalization components provide about the same sensitivity when a hearing aid wearer switches between directional and omnidirectional modes.—DAP

6,570,963

43.66.Ts CALL CENTER FOR HANDLING VIDEO CALLS FROM THE HEARING IMPAIRED

Thomas Michael Watson and Paul W. Ludwick, assignors to Sprint Communications Company L.P.
27 May 2003 (Class 379/52); filed 30 October 2000

This call center for speech and hearing impaired subscribers provides video communication so the subscriber can use sign language to communicate with the call controller. The call center features a number of terminals connected to a switching system to handle calls. Each of the terminals includes a computer system that handles video calls from the speech and



hearing impaired users and a telephone station that handles voice calls to parties communicating with the speech and hearing impaired subscribers. A call controller connected to the switching system and the terminals assigns a terminal to handle an incoming call.—DRR

6,574,342

43.66.Ts HEARING AID FITTING SYSTEM

Keith L. Davis *et al.*, assignors to Sonic Innovations, Incorporated
3 June 2003 (Class 381/314); filed 8 February 2000

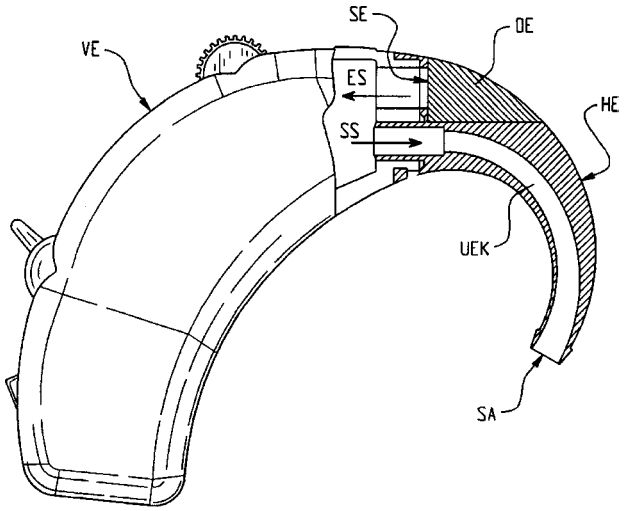
A fitting method is described in which the amount of gain across the hearing aid frequency range is selected such that loudness levels at all frequencies of interest are perceived as being the same by the wearer. Test signals are generated by the hearing aid for the wearer to determine the desired gain versus frequency characteristic. The system is interactive based on responses from the hearing aid wearer via a graphical user interface.—DAP

6,574,343

43.66.Ts HEARING AID

Hilmar Meier, assignor to Phonak AG
3 June 2003 (Class 381/322); filed in the European Patent Office 2 March 1998

A cover over the microphone inlet of a behind-the-ear hearing aid provides smoother exterior surfaces of the case than are normally achieved.



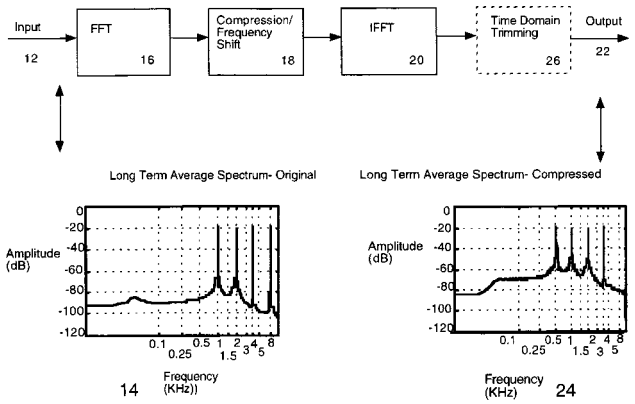
The smooth surfaces of the cover prevent air turbulence from creating acoustical artifacts in the hearing aid. The cover is typically made with an open-pore material such as polyethylene.—DAP

6,577,739

43.66.Ts APPARATUS AND METHODS FOR PROPORTIONAL AUDIO COMPRESSION AND FREQUENCY SHIFTING

Richard Ray Hurtig and Christopher William Turner, assignors to University of Iowa Research Foundation
10 June 2003 (Class 381/316); filed 16 September 1998

A method is described for frequency compression and frequency shifting in which the ratios of the vocal tract resonant frequencies are kept constant. Each frequency is shifted upward or downward by a fixed multiplicative factor while maintaining the spectral shape and duration of the audio signal. A FFT is generated to perform frequency analysis on short time



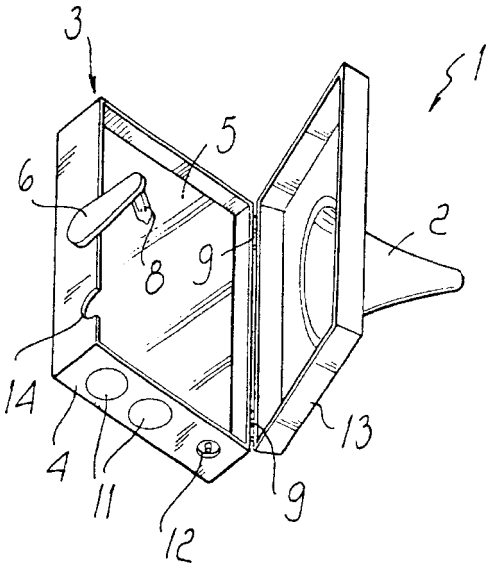
segments of the input signal. After the proportional frequency compression and spectrum shifting, an inverse FFT is performed. The resultant output signal is amplified appropriately to ensure that speech is audible across the usable frequency range.—DAP

6,569,090

43.66.Yw APPARATUS FOR SELF-INSPECTING THE EAR

Giorgio Mezzoli and Marco Vanzi, assignors to Giorgio Mezzoli
27 May 2003 (Class 600/200); filed in Italy 12 January 1999

Do you want to be able to look into your own ear? Here is a gadget to self-inspect your external auditory meatus and/or the tympanic membrane. The device consists of an ear insert **2**, reflective mirror **3**, and an illuminator



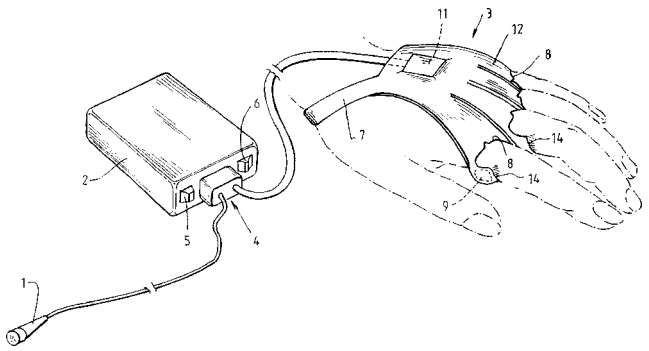
6 to light up the dark recess of the middle ear. It appears that another mirror and some deft maneuvering on the user's part would be needed to see the reflected image of the ear.—DRR

6,466,911

43.71.Ky ELECTROTACTILE VOCODER USING HANDSET WITH STIMULATING ELECTRODES

Robert S. C. Cowan *et al.*, assignors to The University of Melbourne
15 October 2002 (Class 704/271); filed in Australia 30 May 1997

This wired glovelike device receives impulses from a coded speech-stream and generates mild electrical impulses to various parts of the wearer's hand according to the spoken input material. The coding is essentially



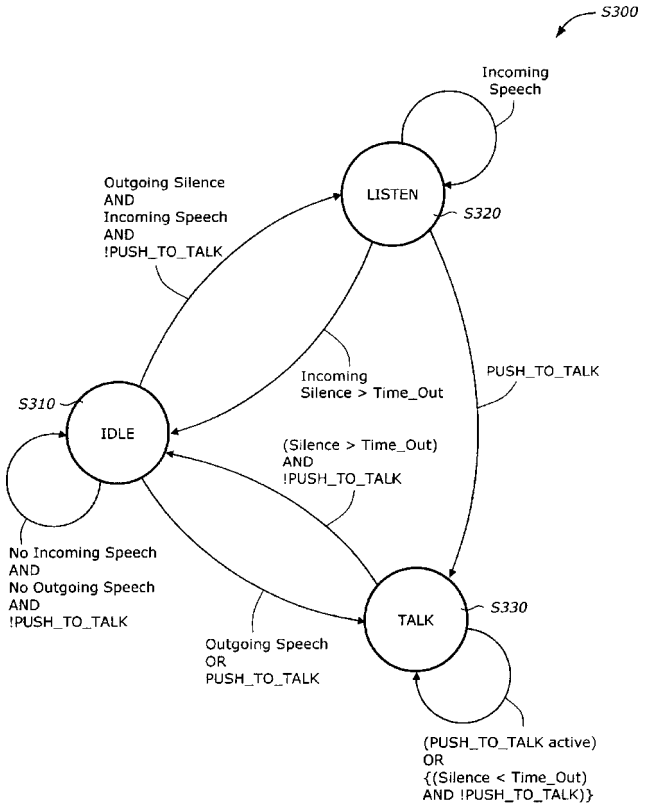
the old channel vocoder arrangement, but limited to pitch, second formant, signal amplitude, voiced/unvoiced, and high-frequency amplitude information. Each of these features controls the pulse rate applied to a particular area of the glove.—DLR

6,563,804

43.72.Gy SYSTEM AND METHOD FOR PROVIDING FULL-DUPLEX AUDIO COMMUNICATION USING A HALF-DUPLEX AUDIO CIRCUIT

Prakash Iyer *et al.*, assignors to Intel Corporation
13 May 2003 (Class 370/296); filed 5 January 1998

A method is described to provide full-duplex audio communication, as in a two-way conversation over a telephone line, with a half-duplex audio circuit. The audio circuit interfaces typically to microphones, CD players, headphones, speakers, and stereo amplifiers. System operation occurs



in either an idle state, a listen state, or a talk state. Transition to a full-duplex event from one of the states is initiated via incoming speech, outgoing speech, or a talk request from the half-duplex audio circuit via a push-to-talk switch.—DAP

6,466,909

43.72.Ja SHARED TEXT-TO-SPEECH RESOURCE

Cliff Didcock, assignor to Avaya Technology Corporation
15 October 2002 (Class 704/260); filed 28 June 1999

What this patent describes is little more than a buffering scheme whereby a speech synthesizer is started and stopped according to the amount of waveform data currently awaiting audio conversion. This allows a more or less continuous output stream with little waiting for new phrases to be composed by the synthesizer.—DLR

6,470,316

43.72.Ja SPEECH SYNTHESIS APPARATUS HAVING PROSODY GENERATOR WITH USER-SET SPEECH-RATE- OR ADJUSTED PHONEME-DURATION-DEPENDENT SELECTIVE VOWEL DEVOICING

Keiichi Chihara, assignor to Oki Electric Industry Company, Limited
22 October 2002 (Class 704/267); filed in Japan 23 April 1999

This fairly typical speech synthesizer performs an analysis of the input text and assigns segment quality and durations according to a fairly conventional set of rules. The speech waveform is generated using an overlap-and-add concatenation system, again traditional. Designed for output in the Japanese language, there is a strong emphasis on the control of vowel devoicing, with the degree of devoicing influenced by the current speech rate setting.—DLR

6,473,979

43.72.Ja ELECTRONIC TALKING COMPASS

Clive S. Lu, Hicksville, New York
5 November 2002 (Class 33/363 K); filed 5 December 2000

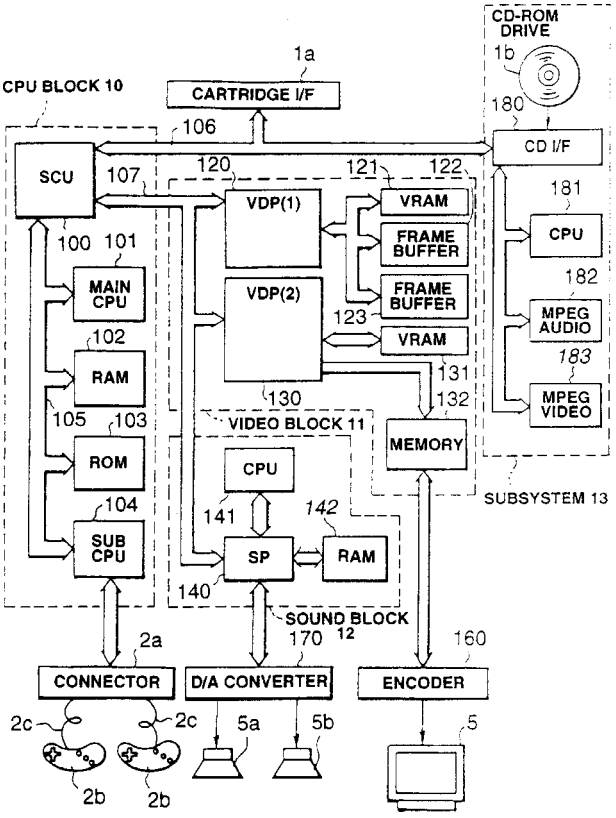
This talking compass uses a Hall effect sensor to detect the orientation of the device case with respect to the Earth's magnetic field and a speech output circuit to announce the direction the case is pointing.—DLR

6,572,475

43.72.Lc DEVICE FOR SYNCHRONIZING AUDIO AND VIDEO OUTPUTS IN COMPUTERIZED GAMES

Teruo Okabe and Hiroshi Yagi, assignors to Kabushiki Kaisha Sega Enterprises
3 June 2003 (Class 463/30); filed in Japan 28 January 1997

A game device is described for synchronizing video output with



sounds using audio parameters on the basis of a display parameter of an object.—HHN

6,466,654

43.72.Ne PERSONAL VIRTUAL ASSISTANT WITH SEMANTIC TAGGING

Robert S. Cooper *et al.*, assignors to Avaya Technology Corporation
15 October 2002 (Class 379/88.01); filed 6 March 2000

This voice command recognizer uses a grammar arrangement long known as a slot grammar, which provides a narrow limit to the branching factor, or complexity, of the set of recognizable utterances. The grammar is basically ''command verb''+''object''+''tag.'' The commands all concern computer file management operations. The object is typically a stored file, a message, a fax, etc. Most of the details reside in the final ''tag'' slot, which provides a goal, action, or result to be accomplished. This may include a designated person, a date, a status change, or any of several other types of destinations or results. The ''tag'' includes a link to another item in the database.—DLR

6,468,083

43.72.Ne GLOBAL COMMUNICATION MEANS

Joseph Mathias, Ridgeville, Ontario, Canada
22 October 2002 (Class 434/167); filed 29 September 2000

Rather than solving the admittedly difficult problem of computer speech recognition, the author of this patent would create a new artificial language with a clearly prescribed phonetic system, presumably easier for both computers and people to handle. The ''global'' specification is just that—all natural languages the world around would be dropped in favor of this new scheme. Goodbye English, French, etc. Of course, the patent completely ignores the fact that artificial languages, including this one, have invariably been found to be decidedly unnatural, not readily processed by the human brain. Regularity is not the blessing many suppose it to be.—DLR

6,471,420

43.72.Ne VOICE SELECTION APPARATUS VOICE RESPONSE APPARATUS, AND GAME APPARATUS USING WORD TABLES FROM WHICH SELECTED WORDS ARE OUTPUT AS VOICE SELECTIONS

Hidetsugu Maekawa *et al.*, assignors to Matsushita Electric Industrial Company, Limited
29 October 2002 (Class 395/2.79); filed in Japan 13 May 1994

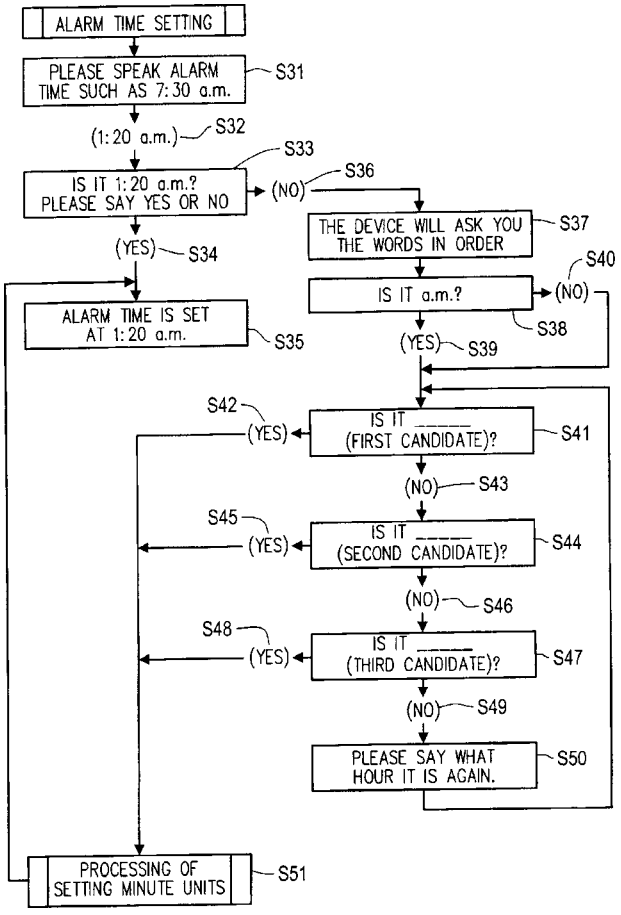
This human/computer interface uses a simplified voice analyzer and combines the resulting feature set with features extracted from a photocell lip position detector. The combined speech and lip movement feature set is used to control a computer-generated object, such as a game playing figure. Joystick inputs may also be added to the combination feature set.—DLR

6,564,185

43.72.Ne CONTINUOUS SPEECH RECOGNITION METHOD AND PROGRAM MEDIUM WITH ALTERNATIVE CHOICE SELECTION TO CONFIRM INDIVIDUAL WORDS

Yasunaga Miyazawa *et al.*, assignors to Seiko Epson Corporation
13 May 2003 (Class 704/251); filed in Japan 8 September 1998

A continuous speech recognition system analyzes the user's speech



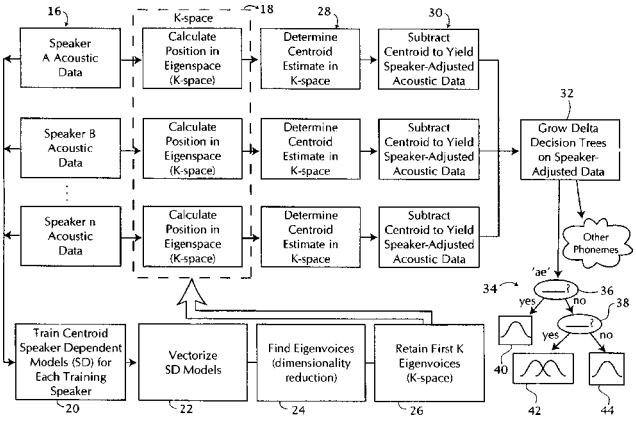
input and produces an affirmative/negative recognition result in order to accurately confirm the overall recognition performance.—HHN

6,571,208

43.72.Ne CONTEXT-DEPENDENT ACOUSTIC MODELS FOR MEDIUM AND LARGE VOCABULARY SPEECH RECOGNITION WITH EIGENVOICE TRAINING

Roland Kuhn *et al.*, assignors to Matsushita Electric Industrial Company, Limited
27 May 2003 (Class 704/250); filed 29 November 1999

The eigenvoice technique makes up the core of this patent to train and develop context-dependent acoustic models for phonemes. Maximum



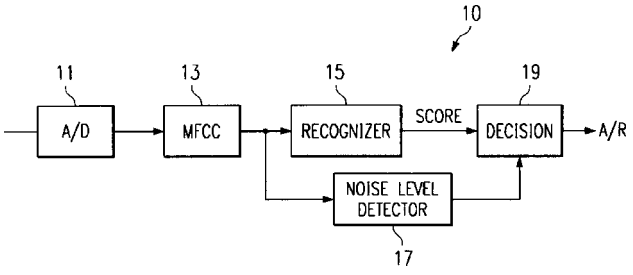
likelihood estimation methods are used to develop common decision tree frameworks.—HHN

6,577,997

43.72.Ne SYSTEM AND METHOD OF NOISE-DEPENDENT CLASSIFICATION

Yifan Gong, assignor to Texas Instruments Incorporated
10 June 2003 (Class 704/252); filed 27 April 2000

Speech recognition accuracy may degrade significantly in noisy environments. A classification scheme makes a decision to accept or reject the speech recognition output based on the average noise level in the input



signal and the probability of speech recognition success relative to a known vocabulary list.—DAP

6,577,999

43.72.Ne METHOD AND APPARATUS FOR INTELLIGENTLY MANAGING MULTIPLE PRONUNCIATIONS FOR A SPEECH RECOGNITION VOCABULARY

James R. Lewis and Barbara Ballard, assignors to International Business Machines Corporation
10 June 2003 (Class 704/270); filed 8 March 1999

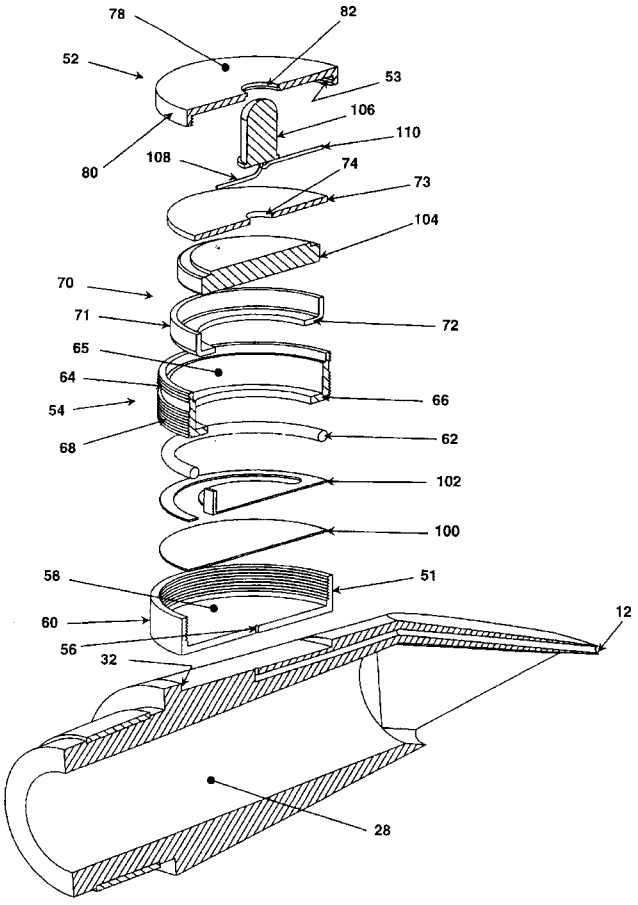
To manage multiple pronunciations in a speech recognition system, this method assigns a base quality metric for each of several acoustic models. The occurrence of certain events statistically determines the accuracy of each acoustic model.—HHN

6,570,077

43.75.Ef TRAINING DEVICE FOR MUSICAL INSTRUMENTS

Stacy P. Goss, Albertville, Alabama
27 May 2003 (Class 84/477 R); filed 6 March 2002

The inventor believes that developing sufficient wind pressure to support the reed oscillation is an important issue with beginning players. Accordingly, imagine a separate air channel **12** that leads to a diaphragm **100** that pushes on a spring switch **102** lighting LED **106** (note battery **104**).



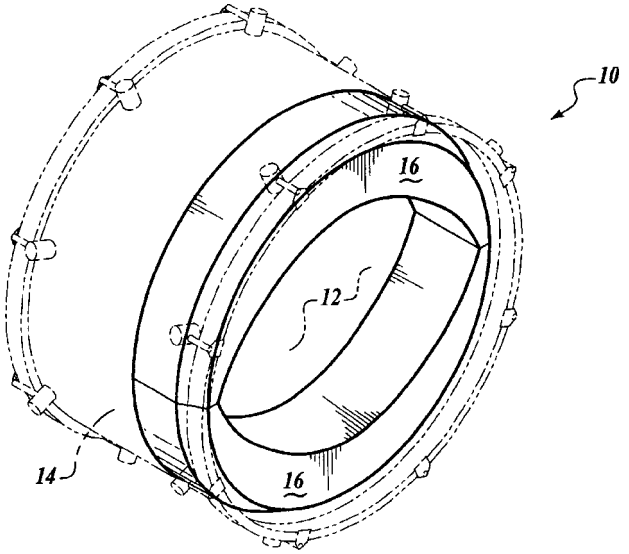
While this reviewer acknowledges the importance of breath support, particularly with double reeds, looking down at the binary output of a LED is not going to help with the myriad other issues of playing reed instruments. And how do you clean the saliva from the air channel?—MK

6,573,441

43.75.Hi DRUM PILLOW AND METHOD FOR USING SAME

Trent M. Norris, Jr., Port Angeles, Washington
3 June 2003 (Class 84/411 M); filed 5 June 2001

It is common practice to add a pillow in front of the bass drum membrane to deaden the sound, particularly for garage or studio environments. The inventor proposes an annular pillow so that the ''higher harmonics



generated at the edge'' are attenuated. Drum head vibrations are not that simple, but the invention will perform more or less as the inventor intended.—MK

6,576,826

43.75.Tv TONE GENERATION APPARATUS AND METHOD FOR SIMULATING TONE EFFECT IMPARTED BY DAMPER PEDAL

Masao Kondo and Satoshi Miyata, assignors to Yamaha Corporation
10 June 2003 (Class 84/615); filed in Japan 22 February 2000

Use of the damper pedal in an acoustic piano permits the string vibrations to continue until the pedal is released. The question is: how is an electronic piano supposed to accomplish the same action? In addition to the normal tone, the inventors propose storing the additive damper pedal tone, the tone after the pedal is released, and the additional ''key off'' tone. By analyzing where the pedal is depressed with respect to the key on event, these additional tones can be blended in to create a more realistic pedal tone.—MK

6,576,827

43.75.Wx MUSIC SOUND SYNTHESIS WITH WAVEFORM CACHING BY PREDICTION

Motoichi Tamura, assignor to Yamaha Corporation
10 June 2003 (Class 84/622); filed in Japan 23 March 2001

This patent combines two ideas: First, that musical instrument parameters can be analyzed (but are not disclosed in the patent) and then converted to vectors (also not disclosed). Then, during playback, these vectors must be interpreted. Naturally, if you cache the vectors, then there is less bus traffic, less disk traffic, and the overall system is faster. However, the encoding scheme is reminiscent of vector quantization and the caching of waveform data is well known (also see United States Patent 6,365,816 [reviewed in J. Acoust. Soc. Am. **112**(4), 1243 (2002)]).—MK

6,568,123

43.80.Nd BLOOD-SUCKING INSECT CONTROL STATION

J. Roy Nelson *et al.*, assignors to Bugjammer, Incorporated
27 May 2003 (Class 43/107); filed 20 June 2001

The authors of this patent fervently believe that adding a heartbeat sound makes a biting insect trap all the more attractive. Accordingly, imagine a combination speaker, CO_2 dispenser, heater, etc.—MK

6,561,979

43.80.Qf MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD

Andrew J. Wood and Kevin S. Randall, assignors to Acuson Corporation
13 May 2003 (Class 600/437); filed 14 September 1999

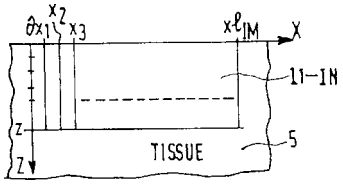
The principal feature of this portable ultrasound system is the electromagnetic shielding afforded by encapsulation of individual components of the system. The compartmentalization of the components isolates EMI between the components themselves as well as the system and the ambient world. The EMI shielding also contributes to overall structural rigidity and, with the proper design, can provide heat dissipation for the internal components. A transducer connector is also described as providing high-density interconnects between the transducer and the ultrasound system and facilitating easy handling.—DRR

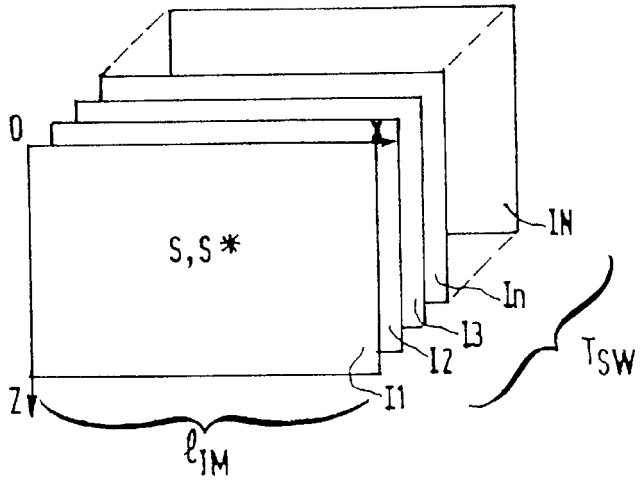
6,561,981

43.80.Qf ULTRASONIC METHOD AND SYSTEM FOR SHEAR WAVE PARAMETER ESTIMATION

Odile Bonnefous, assignor to Koninklijke Philips Electronics N.V.
13 May 2003 (Class 600/443); filed in the European Patent Office 26 April 2000

This patent relates to an ultrasonic system for determining the local propagation velocity of transient shear waves in tissue, for displaying a sequence of velocity images of the transient shear waves, and for deriving tissue elasticity information. Shear waves are generated in the tissue by an external mechanical vibration source, producing displacement of tissue





particles. A standard ultrasonic diagnostic system measures the velocity of the particles and the velocity of the front of the shear waves. A computer program is applied to estimate parameters such as elasticity from the velocity of the shear wave front.—DRR

6,565,512

43.80.Qf SYSTEM FOR ESTIMATING BLADDER VOLUME

Dipankar Ganguly *et al.*, assignors to SRS Medical Systems, Incorporated
20 May 2003 (Class 600/449); filed 17 November 2000

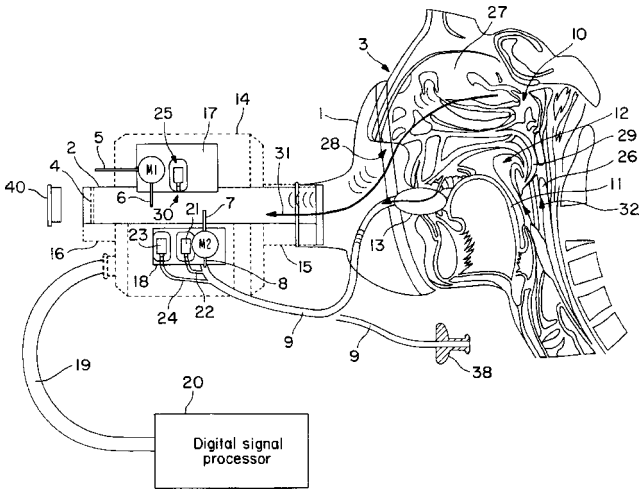
Information about the volume of urine in the bladder is important clinically for several types of patients suffering from bladder dysfunction. The noninvasive device covered by this patent is a monitoring system that sequentially scans a bladder with ultrasonic beams, sectioning the bladder into transverse planes. The system includes a transducer with a number of piezoelectric elements in a relatively thin elastomeric pad and/or substrate that is applied to the patient's lower abdomen. The elements are spaced and grouped in such a manner that the acoustic beams they produce lie on a number of planes, with the first and last planes separated by approximately 70°. The system establishes from the scan lines associated with a given plane a number of points on each of the front and back walls of the bladder. It then fits a curve to each of the sets of points. On the basis of these two joined curves, the system calculates the cross-sectional area of the bladder in the plane. After determining the area of each of the planes, the system can determine the volume of the bladder essentially by summing up the weighted elements formed by the planar areas.—DRR

6,565,517

43.80.Qf APPARATUS AND METHODS FOR RHINOMANOMETRY

Steen Brabrand Rasmussen, assignor to Rhinometrics A/S
20 May 2003 (Class 600/529); filed in Denmark 29 July 1998

Rhinomanometry is an examination method for determining flow resistances in the nose. The apparatus here consists of a measuring tube having a proximal end for connecting to a patient's nose and a distal end leading to the ambient atmosphere, a measuring gate, a means for determining air flow in the measuring tube, a pressure sensor for measuring the pressure in



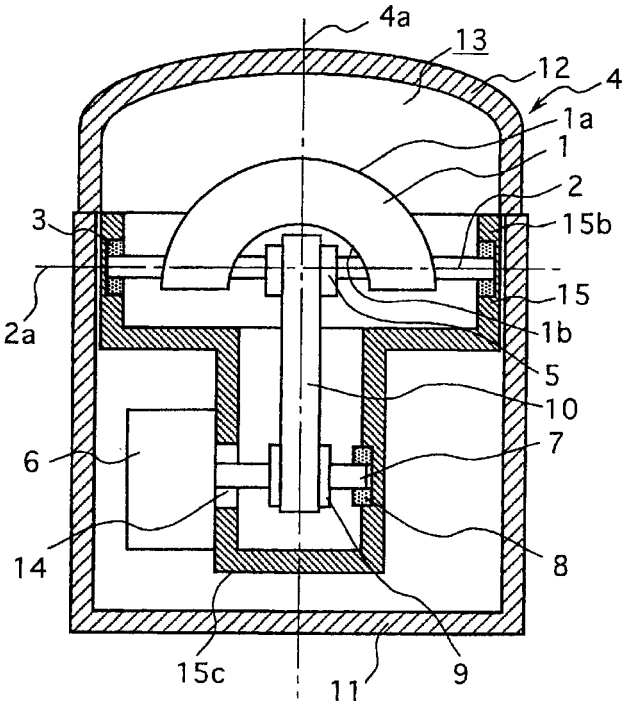
the tube's proximal end, and another pressure sensor for measuring the pressure at the measuring gate. The apparatus contains means for generating a sound signal at the measuring gate and means to receive sound signals in the tube proximal end.—DRR

6,569,100

43.80.Qf ULTRASONIC PROBE AND METHOD OF PRODUCING SAME

Eiichi Okawa *et al.*, assignors to Matsushita Electric Industrial Company, Limited
27 May 2003 (Class 600/445); filed in Japan 17 November 2000

This probe, designed to be more compact and said to be more accurate, includes a driven rotating shaft mounted inside of, and supported by, a housing. An ultrasonic transducer, mounted on the shaft so as to be swingable, converts an electrical signal to and from an ultrasound signal. A motor shaft is connected by a belt to the driven shaft. Constructing the probe in this



manner makes it possible to reduce the rotation radius of the transducer independently of the size of the electric motor by spacing the rotation axes of the drive motor and the transducer. Transferring the drive motor rotation

into transducer rotation without causing a backlash is said to prevent deterioration of the probe's accuracy.—DRR

6,569,101

43.80.Qf MEDICAL DIAGNOSTIC ULTRASOUND INSTRUMENT WITH ECG MODULE, AUTHORIZATION MECHANISM AND METHODS OF USE

Jen U. Quistgaard *et al.*, assignors to Sonosite, Incorporated
27 May 2003 (Class 600/459); filed 19 April 2001

This is a handheld ultrasound instrument that incorporates enhanced diagnostic modes including pulse wave Doppler, time-motion analysis, and tissue harmonic imaging. An external electrocardiograph (ECG) recording unit is also described. The ECG unit adapts for use with the handheld instrument to enable ECG monitoring simultaneously with an ultrasonic scan in *b*-mode, Doppler, color Doppler, *m*-mode, and PW mode. The handheld instrument also includes a secure mechanism that allows any combination of diagnostic modes to be enabled by the manufacturer (and later to enable or disable any one or group of diagnostic modes). The patent also covers a method for a manufacturer to maintain a database of handheld instrument capabilities after it enters medical distribution and use.—DRR

6,569,102

43.80.Qf MINIATURIZED ULTRASOUND APPARATUS AND METHOD

Mir A. Imran *et al.*, assignors to Zonare Medical Systems, Incorporated
27 May 2003 (Class 600/459); filed 18 May 2001

This ultrasound apparatus for examining tissue in a region of interest has a housing with a viewing aperture. An ultrasonic transducer consisting of a number of elements is deployed in the viewing aperture. Electrical pulses are supplied to the transducer to excite ultrasound signals into the body and the same transducer detects the ultrasonic reflections from the body, yielding electrical signals that are gain-corrected in accordance with time. In-phase and out-of-phase components of the electrical signals are then digitized and collected to form one image for a single frame of tissue in the region of interest in the body.—DRR

6,572,554

43.80.Qf METHOD AND APPARATUS FOR INTRAVASCULAR TWO-DIMENSIONAL ULTRASONOGRAPHY

Paul G. Yock, assignor to SciMed Life Systems, Incorporated
3 June 2003 (Class 600/463); filed 20 June 2002

This catheter is inserted into the blood vessel of a patient to provide ultrasound imaging of the vessel wall. The catheter incorporates a tubular element and an internally housed drive cable for effective circumferential scan about the catheter that contains means for generating ultrasound signals. Both the tubular element and the drive are of size and flexibility sufficient to permit their introduction into the vessel and subsequent advancement through the vessel to the location where imaging is to be executed.—DRR

6,575,908

43.80.Qf BALANCE BODY ULTRASOUND SYSTEM

Stephanie A. Barnes *et al.*, assignors to Sonosite, Incorporated
10 June 2003 (Class 600/443); filed 21 August 2002

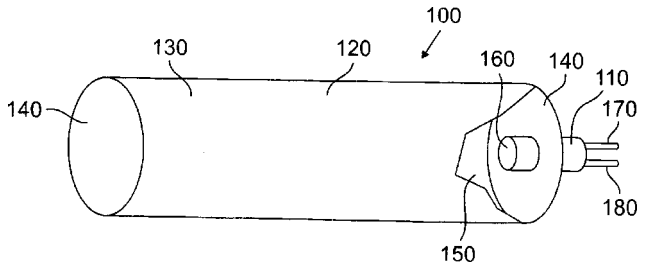
The subject portable ultrasound system consists of a transducer assembly connected to a balance body and a number of control elements arranged in an ergonomic fashion on the body so that a user may hold the system and operate at least one of the control elements with one hand. The balance body also incorporates system electronics, a power supply, and a user interface. The interface includes a *D*-controller and a touch screen.—DRR

6,575,916

43.80.Qf APPARATUS AND METHOD FOR DETECTING VERY LOW FREQUENCY ACOUSTIC SIGNALS

Michael E. Halleck *et al.*, assignors to iLife Solutions, Incorporated
10 June 2003 (Class 600/528); filed 24 March 2000

For detecting extremely low-frequency acoustic signals, this apparatus contains a sensor capable of detecting signals in the frequency range of 0.1 to 30 Hz. The sensor consists of a chamber featuring portions that form a cavity. A low-frequency microphone is placed within the cavity. Another embodiment consists of a chamber having portions that form a resonant



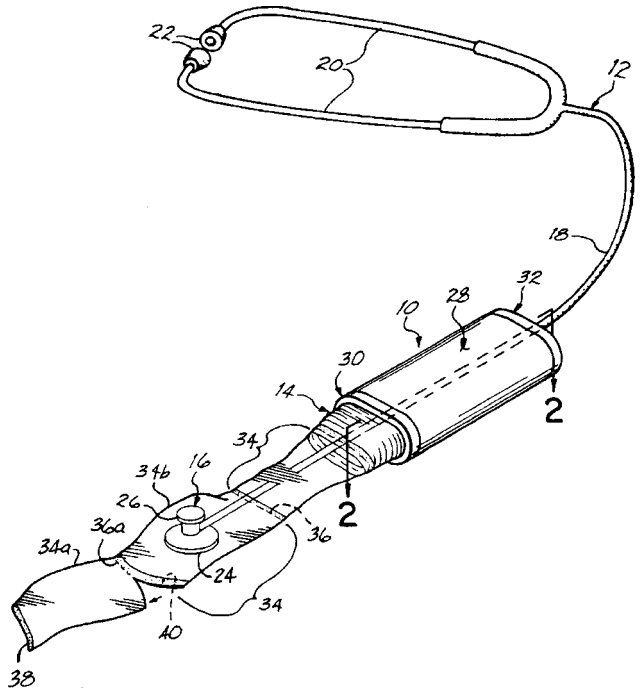
cavity, a lower frequency microphone within the resonant cavity, and a membrane that covers the resonant cavity. Low-frequency acoustic signals that are incident on the membrane cause the membrane to vibrate and amplify the acoustic signals within the resonant cavity. The sensor can provide information concerning physiological conditions, such as respiration and cardiac activity.—DRR

6,575,917

43.80.Qf PROTECTIVE-SLEEVE CARTRIDGE AND STETHOSCOPE INCORPORATING SAME

Jennifer S. Giroux *et al.*, assignors to St. Joseph Solutions LLC
10 June 2003 (Class 600/528); filed 12 March 2002

What we have here is a protective sleeve cartridge that is directly mountable onto a stethoscope. Thus, a clinician does not have to move to a box of covers, remove a single cover from the dispenser, pull the cover up, and cover the head of the stethoscope—all of which would need to be done



before the next patient. Instead, because the protective sleeve cartridge is located on the stethoscope itself, it remains at hand at all times, thereby making it quick and easy to have an unused sleeve portion covering the stethoscope head each time the clinician examines a different patient.—DRR

6,575,922

43.80.Qf ULTRASOUND SIGNAL AND TEMPERATURE MONITORING DURING SONO-THROMBOLYSIS THERAPY

Jim Fearnside and Al Kyle, assignors to Walnut Technologies
10 June 2003 (Class 601/2); filed 17 October 2000

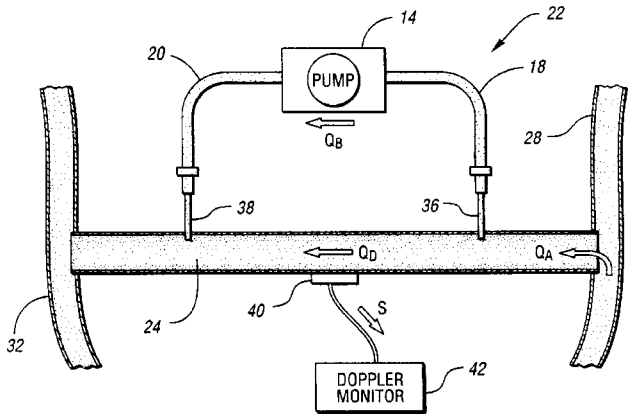
This system for sono-thrombosis provides real-time, continuous intensity and temperature monitoring. An ultrasound delivery column is established for the patient being treated for a thrombotic or emolic occlusion. The column includes an ultrasound transducer, a coupling medium, and a hydrophone. The transducer is operated in a conventional manner to deliver ultrasound to the occlusion. During the operation of the transducer, characteristics of the ultrasound are monitored on a real-time basis with the hydrophone. The method also includes monitoring the temperature near the occlusion on a real-time basis. The intensity of the delivered ultrasound is maintained within a fixed intensity range.—DRR

6,575,927

43.80.Qf SYSTEM AND METHOD FOR DETERMINING BLOOD FLOW RATE IN A VESSEL

William F. Weitzel *et al.*, assignors to The Regents of the University of Michigan
10 June 2003 (Class 604/8); filed 12 May 1999

A system and method are provided for measuring the performance of a vessel, such as a hemodialysis access (which communicates blood between two locations within the patient). A conduit, e.g., an external dialysis circuit or an intravascular catheter, is supplied in fluid communication with the vessel and features a means of diverting blood from the vessel at a certain



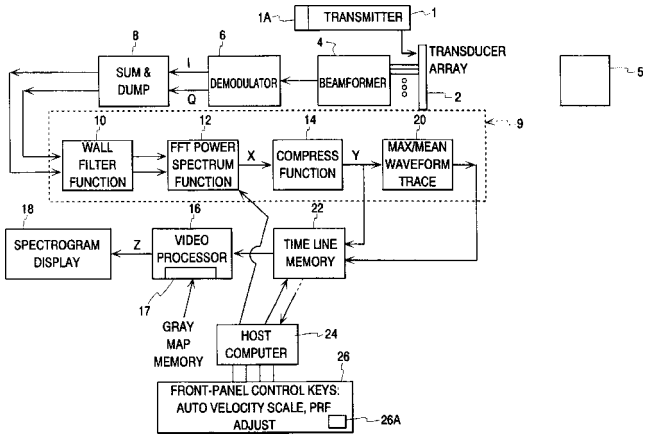
point into the conduit. An ultrasonic sensor in communication with the vessel generates at least one signal that is a function of the blood flow rate in the vessel downstream from the diversion point. The downstream flow depends on the determined conduit flow rate and the performance of the vessel can thus be established based on the signal. Moreover, one or more processors can be provided in communication with the sensor to determine a flow rate in the vessel upstream from the diversion point.—DRR

6,577,967

43.80.Qf AUTOMATIC ADJUSTMENT OF VELOCITY SCALE AND PULSE REPETITION FREQUENCY FOR DOPPLER ULTRASOUND SPECTROGRAMS

Larry Y. L. Mo *et al.*, assignors to General Electric Company
10 June 2003 (Class 702/76); filed 26 November 2001

This device, to be used in ultrasound diagnostic systems that measure the velocity of fluid flow using special Doppler techniques, provides automatic adjustment of velocity scale and pulse repetition frequency for such systems. A Doppler unit generates Doppler signals from ultrasound back-



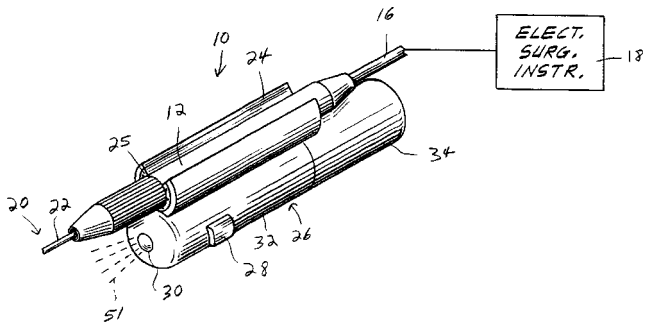
scatter and stores them in a memory. A logic unit executes a signal analysis algorithm that automatically adjusts the polarity and position of the velocity scale of a display and automatically adjusts the pulse repetition rate of the transmitter.—DRR

6,562,032

43.80.Sh ELECTROSURGICAL INSTRUMENT WITH VIBRATION

Alan G. Ellman and Jon C. Garito, both of Hewlett, New York
13 May 2003 (Class 606/41); filed 26 March 2001

Both rf energy and ultrasound energy are combined in this handpiece attachment to enable a surgeon to modulate all types of tissues with different degrees of effectiveness. In one embodiment, a small unbalanced motor causes the handpiece containing the surgical electrode (which also receives



rf energy) to vibrate as a whole, thereby imparting ultrasonic energy to the electrode. The patent proposes that the device can be applied to skin resurfacing but may also be used for tissue modulation, coagulation, hemostasis, and other electrosurgical procedures.—DRR

6,565,520

43.80.Sh APPARATUS FOR ULTRASONIC THERAPEUTIC TREATMENT

Michael John Radley Young, assignor to Orthosonics Limited
20 May 2003 (Class 601/2); filed 23 February 1999

This device for noninvasively treating skeleton/muscular injuries or for diagnosing bone fractures involves applications of two components of ultrasonic energy, at different frequencies in the range between 10 kHz and 4 MHz, to an external surface of living tissue. The two transducers generating these separate frequencies are placed apart but fairly close to each other, and the axial distribution of the echoes provides the diagnosis.—DRR

6,565,521

43.80.Sh SYSTEM AND METHOD OF VESSEL REMOVAL

Barry N. Silberg, Santa Rosa, California
20 May 2003 (Class 601/2); filed 30 November 2001

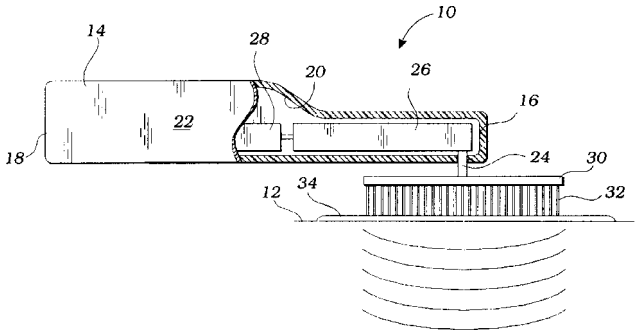
This is a system for harvesting a blood vessel for subsequent use in a grafting procedure, such as saphenous vein graft harvesting for a coronary bypass surgical operation. A quantity of a solution is infused into the tissue surrounding the portion of the vessel to be removed. An external ultrasonic transducer supplies an energy field to the tissue to loosen the intercellular connections between the tissue and the vessel to be removed. Once the energy field has been applied, the portion of the vessel to be removed is separated from surrounding tissue and tributary vessels are ligated. The portion of the vessel is then transected and removed from the body.—DRR

6,569,170

43.80.Sh METHOD OF CLEANING SKIN

David L. Kellogg, Irvine, California
27 May 2003 (Class 606/131); filed 20 September 2001

This is a skin-cleaning version of the popular ultrasonic toothbrush. The ultrasonic skin cleaner is essentially a handle with a gripping surface, a brush mounting point at one end of the handle, and an ultrasonic vibrator positioned within the hollow interior of the handle and operably attached to



the brush mounting point. A battery also occupies a portion of the inside of the hollow cylinder to provide power to the ultrasonic vibrator. The brush may be used in conjunction with a liquid cleaning fluid to cleanse the skin.—DRR

6,575,173

43.80.Sh ULTRASONIC DEVICE FOR THE TREATMENT OF HAIR AND OTHER FIBERS

Ke Ming Quan *et al.*, assignors to The Procter & Gamble Company
10 June 2003 (Class 132/119.1); filed 31 August 2001

This device uses ultrasound for treating hair (or other fibers). A comb device, excited by a piezoelectric ultrasonic generator, has a number of protuberances with a natural bending frequency extending outward from the comb device. Another embodiment features a reflector for reflecting the ultrasonic waves deployed at the distal end of the comb device.—DRR

6,575,956

43.80.Sh METHODS AND APPARATUS FOR UNIFORM TRANSCUTANEOUS THERAPEUTIC ULTRASOUND

Axel F. Brisken *et al.*, assignors to Pharmasonics, Incorporated
10 June 2003 (Class 604/500); filed 5 November 1999

This device is intended to deliver a uniform field of ultrasonic energy over a wide target range in the patient's body. The advantage of a wide beam delivering uniform energy over a large tissue volume is that cellular uptake of injected substances such as therapeutic DNA can be substantially enhanced over the entire region without inflicting tissue damage or degrading an injectate. In the preferred embodiment, the wide beam delivery system consists of a housing having an opening at its distal end with an ultrasound

(generally planar) transducer suspended within the housing. The ultrasound transducer is positioned in contact with an acoustic compliant material, such as water or water with additives, filling the housing. A flexible skin contact window, fabricated from silicon rubber, polyethylene, or other plastic material, is stretched across the opening at the distal end of the housing. The skin contact window is pressed against the patient's flesh during use.—DRR

6,576,220

43.80.Sh NON-INVASIVE METHODS FOR SURGERY IN THE VASCULATURE

Evan C. Unger, assignor to ImaRx Therapeutics, Incorporated
10 June 2003 (Class 424/9.32); filed 10 May 2001

This device relates to the field of magnetic resonance imaging (MRI) and, more specifically, to the use of stabilized gas-filled vesicles as contrast medium for MRI-directed ultrasound surgery. Basically, the MRI is used in conjunction with an administered contrast medium to identify the region of the patient requiring surgery and ultrasound is applied to the region to carry out the surgery. A secondary MRI scanning may follow the application of the ultrasound or the ultrasound operation may occur simultaneously with the MRI scanning. The gas-filled vesicles may contain a therapeutic agent to be released to a localized region of a patient through application of the ultrasound.—DRR

6,561,982

43.80.Vj METHOD AND APPARATUS FOR ULTRASOUND IMAGING IN THE PRESENCE OF CONTRAST AGENTS, PARTICULARLY IN THE FIELD OF CARDIOLOGY

Barbara Greppi and Marino Cerofolini, assignors to Esaote, S.p.A.
13 May 2003 (Class 600/458); filed in Italy 6 July 2000

Ultrasonic imaging is synchronized to occur during the systolic phase of the heart cycle using an electrocardiographic signal. During synchronized imaging, the intensity of the ultrasonic beams is determined using an intermediate mechanical index value for controlled destruction of contrast agent microbubbles to improve imaging in the presence of contrast agents.—RCW

6,569,404

43.80.Vj PHASE SHIFT COLLOIDS AS ULTRASOUND CONTRAST AGENTS

Steven C. Quay, assignor to Amersham Health A/S
27 May 2003 (Class 424/9.52); filed 6 June 1995

These contrast agents consist of a colloidal liquid-in-liquid dispersion in which the dispersed liquid phase has a boiling point below the temperature of the animal to be imaged. Following administration, the dispersed liquid changes phase to an echogenic gaseous foam. Emulsions with particle sizes below 1000 nm can be made and the resulting microbubbles typically have a diameter below 10 000 nm.—RCW

6,572,547

43.80.Vj TRANSESOPHAGEAL AND TRANSNASAL, TRANSESOPHAGEAL ULTRASOUND IMAGING SYSTEMS

David G. Miller *et al.*, assignors to Koninklijke Philips Electronics N.V.
3 June 2003 (Class 600/437); filed 31 July 2001

These imaging systems employ a two-dimensional tranducer array at the distal end of an elongated probe. Beams are transmitted over a selected spatial pattern defined by azimuthal and elevation orientations. Echoes are received from a tissue volume defined by the orientations and by a selected range of depth. Orthographic projection views over the tissue volume are displayed.—RCW

6,572,548

43.80.Vj METHOD AND APPARATUS FOR ULTRASOUND IMAGING, PARTICULARLY FOR THREE-DIMENSIONAL IMAGING

Marino Cerofolini, assignor to Esaote, S.p.A.
3 June 2003 (Class 600/443); filed in Italy 22 June 2000

User-defined parameters in an imaging volume are employed to scan only in a region that coincides with the scan plane or image projection plane specified by the parameters.—RCW

6,572,549

43.80.Vj HIGH FRAME RATE EXTENDED FIELD OF VIEW ULTRASOUND IMAGING SYSTEM AND METHOD

Jing-Ming Jong and Paul Detmer, assignors to Koninklijke Philips Electronics NV
3 June 2003 (Class 600/443); filed 18 December 2001

Speckle in images is processed using cross-correlation to track the movement of an ultrasound probe. By tracking the movement of the probe, images are registered with each other and combined with other images to create a three-dimensional extended field of view. Images used for cross-correlation are acquired more quickly than images of other regions to allow relatively fast scanning of the probe.—RCW

6,572,551

43.80.Vj IMAGING CATHETERS FOR VOLUMETRIC INTRALUMINAL ULTRASOUND IMAGING

Stephen W. Smith *et al.*, assignors to Duke University
3 June 2003 (Class 600/459); filed 11 April 2000

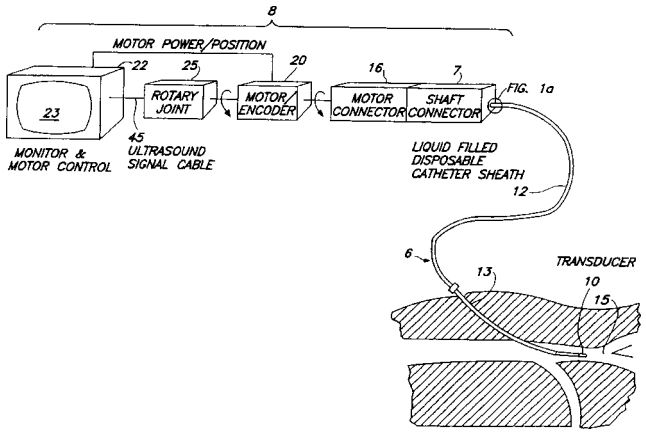
These catheters have an elongated body with an ultrasonic phased array transducer at the distal end. A volume forward of the distal end is scanned. A tool can be positioned at an interstitial site where one or more transducer elements are not present.—RCW

6,572,553

43.80.Vj MEDICAL IMAGING DEVICE

Robert J. Crowley, assignor to SciMed Life Systems, Incorporated
3 June 2003 (Class 600/463); filed 4 January 2002

This device includes a disposable liquid-confining sheath that supports an elongated flexible drive shaft rotating an imaging probe such as an ultrasound transducer. A trocar permits the imaging probe to be inserted where



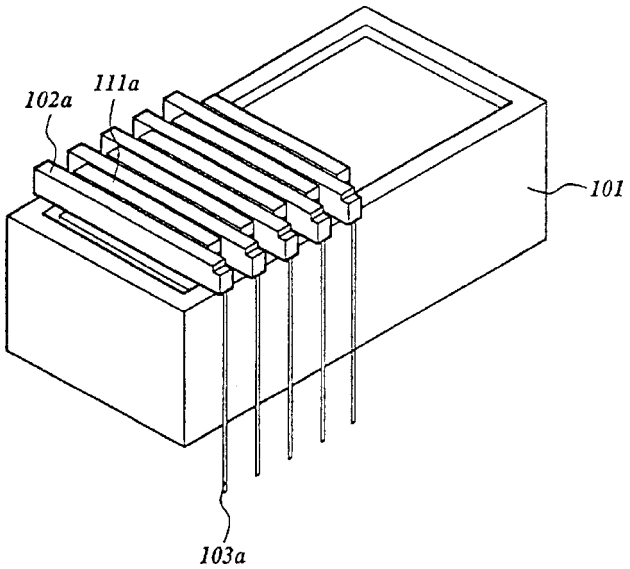
no natural passageway exists. The trocar is adapted to maintain registration with the imaging probe while the probe rotates. An inflatable balloon can be included at the distal end.—RCW

6,575,909

43.80.Vj ULTRASOUND PROBE HAVING TRANSDUCER ELEMENTS WITH DIFFERENT FREQUENCY CENTERS

Nicolaas De Jong *et al.*, assignors to Oldelft B.V.
10 June 2003 (Class 600/458); filed in The Netherlands 25 January 2000

This probe, intended for ultrasound imaging using contrast agents, contains two interleaved arrays of transducer elements. One interleaved array has elements which are longer and have a lower center frequency than



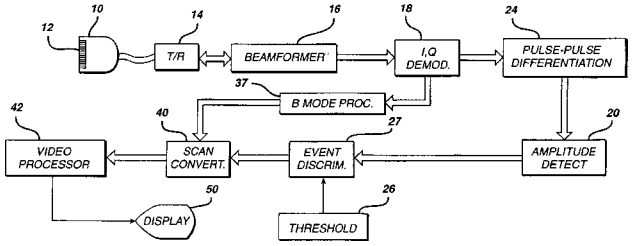
the elements in the other array. The two arrays are supported in a frame that facilitates transmitting with a higher intensity using the longer array elements. The shorter elements are used for both transmitting and receiving.—RCW

6,575,910

43.80.Vj ULTRASONIC IMAGE PERSISTENCE USING CONTRAST AGENTS

Michalakis Averkiou *et al.*, assignors to Koninklijke Philips Electronics N.V.
10 June 2003 (Class 600/458); filed 8 November 2001

Coherent imaging of ultrasonic contrast agents is performed and harmonic signals from contrast agents are detected. A dual display permits viewing a real-time image that shows the location of a contrast agent simultaneously with a triggered contrast image. Contrast agent detection and



imaging include methods to measure perfusion rate characteristics, perform multizone and multifrequency contrast imaging, display perfusion of tissue, and suppress contrast image artifacts.—RCW

U.S. Navy sources and receivers for studying acoustic propagation and climate change in the ocean (L)

John L. Spiesberger[a)]
Department of Earth and Environmental Science, 240 South 33rd Street, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6316

(Received 13 February 2003; revised 24 July 2003; accepted 25 August 2003)

Sounds from a U.S. Navy SSQ-110A source are received at high signal-to-noise ratios at ocean-basin scales at two Sound Surveillance Systems in the Pacific. The sounds have sufficient pulse resolution to study climatic variations of temperature. The acoustic data can be understood using ray and parabolic approximations to the wave equation. Modeled internal waves decrease pulse resolution from 0.01 to 0.1 s, consistent with observations. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1619983]

Pages: 2557–2560

PACS numbers: 43.30.Qd, 43.30.Re, 43.30.Pc [RAS]

I. INTRODUCTION

Sounds from explosive sources in the ocean have been identified over long distances using rays since the 1940s.[1] Such sources have been used to study acoustic propagation and to aid in locating underwater volcanoes[2] and missiles that fall into the sea.[3] Electronically controlled sources and the U.S. Navy's Sound Surveillance System (SOSUS) have been used to study climatic changes of temperature in the ocean.[4–6] However, scientists may not be able to afford to deploy enough sources to resolve many important climatic variations by means of tomography with the available number of SOSUS stations.[7] This paper demonstrates that explosive sources from routine U.S. Navy operations generate sufficiently loud sounds at basin scales so that temporally resolved echoes can be identified using established models for sound speed fields and acoustic propagation. These kinds of data could be used to see if climatic temperature variations could be detected by means of tomography.

II. DATA

A source (SSQ-110A) was detonated at 23.589 °N 208.717 °E on 13 November 1997 (Fig. 1). Signals were received on two of the nine SOSUS stations shown in Fig. 1 of Ref. 4 at distances between 1000 and 3000 km. The location and time of the detonation were known within a few kilometers and a few seconds, respectively. Many pulses arrived at the receivers with high signal-to-noise ratios in a band from about 50 to 200 Hz (Fig. 2).

III. MODELS

Vertical profiles of sound speed are computed along geodesics using Del Grosso's algorithm[8] and Levitus' climatological averages[9] of temperature and salinity for Fall. The depth of minimum speed varies from 760 m at the source to 500 and 600 m, respectively, at receivers one and two. Sound speeds are transformed to Cartesian coordinates via the Earth flattening transformation.

Some of the models incorporate sound speed fluctuations obeying a Garrett–Munk spectrum of internal waves.[10] These fluctuations are added to the climatological field described above. Internal wave modes are computed at range intervals of 80 km to account for changes in water depth, buoyancy frequency, and sound speed. The vertical displacements of these modes are set to zero at the surface and bottom. For each 80-km interval, a three-dimensional field of internal waves is computed in a box of 80 km by 80 km by D m where D is the average depth of the ocean in that interval. A vertical slice through the box gives the vertical displacements along the geodesic. Internal wave energy is given the standard value[10] which worked before[11] though controversy exists.[12–14] Remaining details for constructing internal waves follow Ref. 11 exactly.

The c_0 insensitive parabolic approximation[15] is used to compute a two-dimensional field of sound along a geodesic from 0- to 8000-m depth. Travel times of pulses are computed with an accuracy of a few milliseconds.[15] The field is modeled at each of 1024 acoustic frequencies from 50 to 150 Hz. The impulse response is synthesized with an inverse Fourier transform resulting in a time series with a period of 10.24 s. Near regions where the sound interacts with the bottom, the computational grid has an interval of 25 m in range and 3.9 m in depth. In other regions, the intervals are 100 m in range in 7.8 m in depth. These values are sufficient to obtain convergence within a few decibels.

Fans of rays are traced using a program, zray, that is a modification of ray.[16] Eigenrays are found using another program. These programs have been used to identify acoustic paths before.[17] Rays reflect specularly from the bottom. Both geometric and nongeometric arrivals are found. Nongeometric arrivals are those that provide energy at the receiver on the shadow sides of caustics. The contribution from a caustic at any angle is included if it is within 1000 m of the receiver's depth at the receiver's range. For lack of a more reliable value, rays that reflect from the bottom suffer an attenuation of 3 dB per bounce.

IV. IDENTIFYING ACOUSTIC PATHS

As shown below, the most reliable models use the parabolic approximation and Levitus' climatology[9] with a spec-

a)Electronic mail: johnsr@sas.upenn.edu

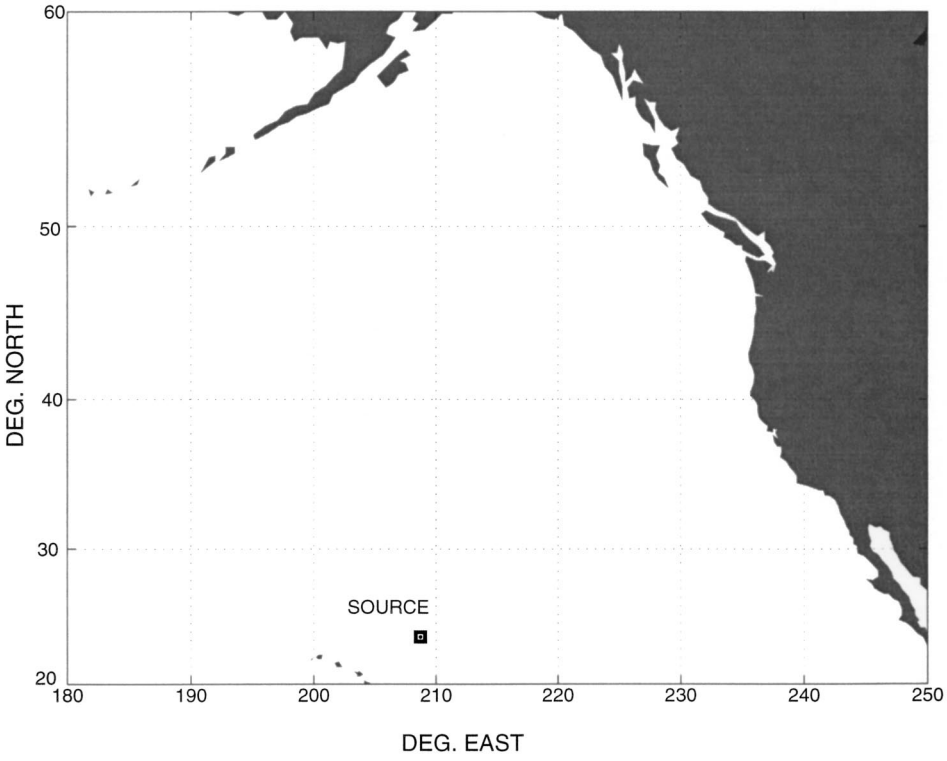


FIG. 1. The explosive source used in this experiment is an SSQ-110A. It is used in normal U.S. Navy operations. Signals are received at two of the nine Sound Surveillance Systems (SOSUS) shown in Fig. 1 of Ref. 4.

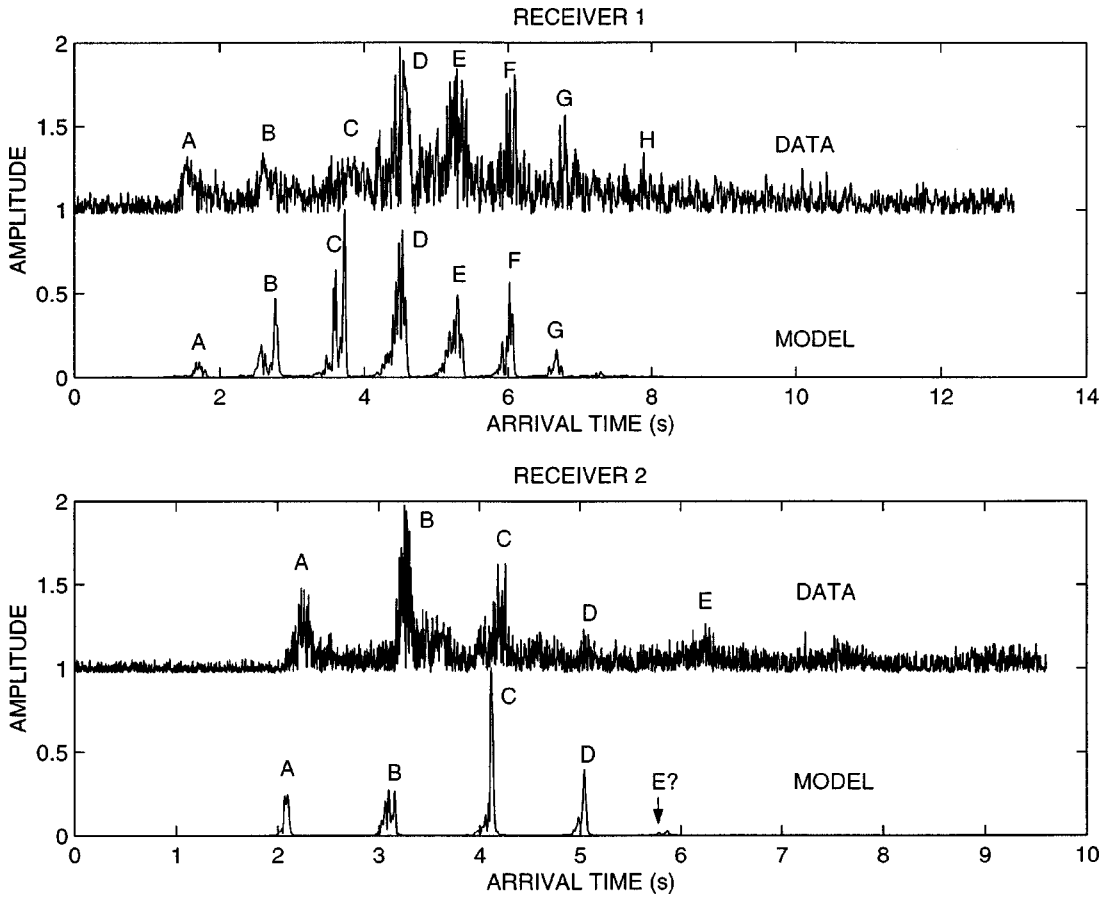


FIG. 2. Comparison of data with models at both receivers. A constant arrival time is added to the model to align with the data and a value of about one is added to the amplitudes of the data so they do not overlap the model. Models are computed using the c_0 insensitive parabolic approximation[15] and a Garrett–Munk[10] spectrum of internal waves superimposed on Levitus' climatological average conditions for Fall.[9] Top: Arrivals except for H are predicted by the model. Bottom: Arrivals except for perhaps E are predicted by the model. Note the time scale differs from the top panel.

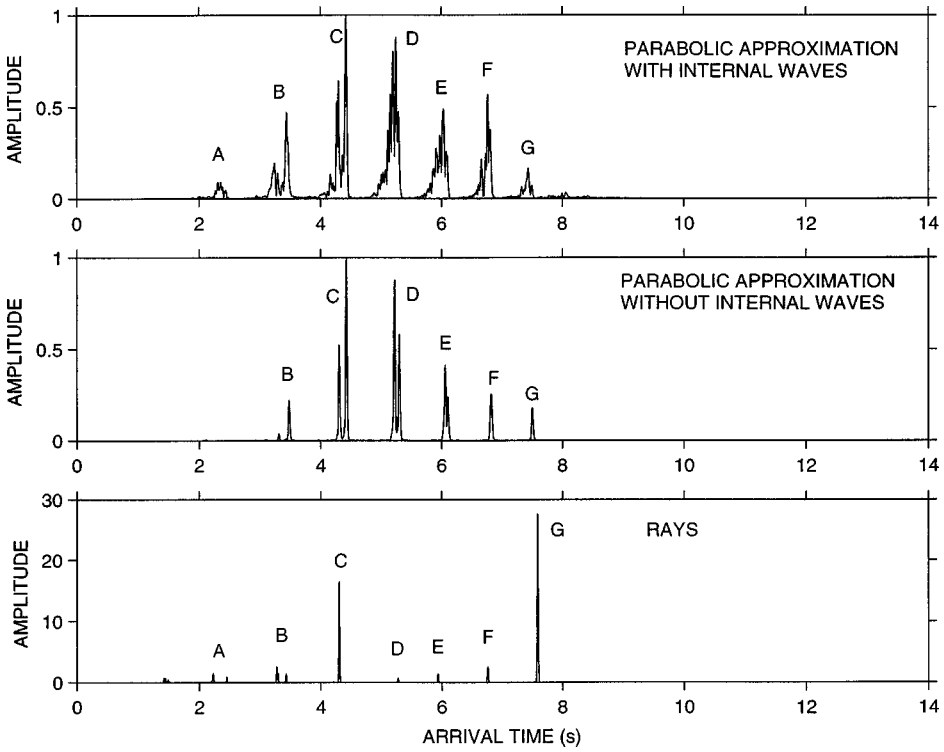


FIG. 3. Three models for the acoustic data at receiver 1. The top two models use the c_0 insensitive parabolic approximation[15] with and without a spectrum of internal waves.[10] The ray model uses the same sound speed field as the parabolic approximation without internal waves on the computational grid used by the parabolic approximation. Arrivals $A-G$ are the same as Fig. 2 (top). Travel times of all three models are shifted by identical amounts so that travel times between models can be compared. Amplitudes from the parabolic approximation are normalized to unity. Amplitudes from the ray model are unnormalized.

trum of internal waves[10] (Fig. 2). At receiver one, arrivals $A-G$ appear in the data and model. The last arrival at H is not in the model. At receiver two, arrivals $A-D$ appear in the data and model. The last arrival at E is probably not in the model. Except for the last arrivals, the relative travel times in the models resemble the data. None of the rays reflect from the bottom or surface.

The received pulses have widths from 0.1 to 0.2 s, similar to those modeled. Since the bandwidth of the model is 100 Hz, the emitted width is $1/(100\ \text{Hz}) = 0.01$ s. This is ten times too small. Internal waves are responsible for the broadening to 0.1 s at both receivers (Figs. 3 and 4).

The correspondence between the arrival times of the data and ray models is good (Figs. 3 and 4). However, the relative amplitudes from rays are much less reliable than those from the parabolic approximation.

Arrival A at receiver one appears in the parabolic ap-

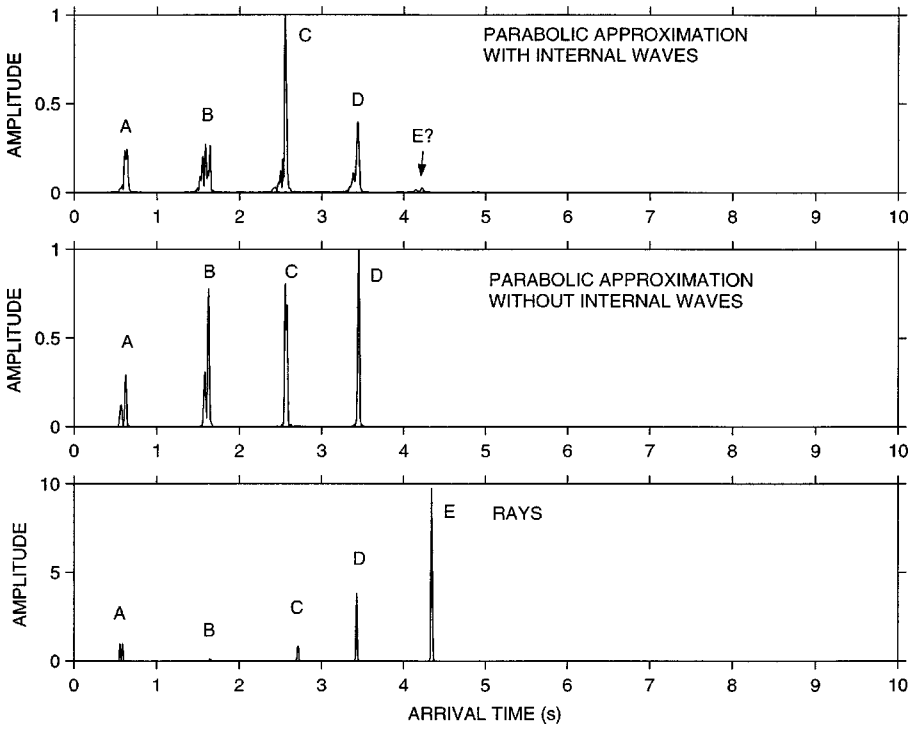


FIG. 4. Three models for the acoustic data at receiver 2. The top two models use the c_0 insensitive parabolic approximation[15] with and without a spectrum of internal waves,[10] respectively. The ray model uses the same sound speed field as the parabolic approximation without internal waves on the computational grid used by the parabolic approximation. Arrivals $A-E$ are the same as the bottom panel of Fig. 2. Travel times of all three models are shifted by identical amounts so that travel times between models can be compared. Amplitudes from the parabolic approximation are normalized to unity. Amplitudes from the ray model are unnormalized.

proximation with internal waves (Fig. 3). These waves scatter the sound down to the depth of the receiver. Without scattering, the caustic from arrival *A* is too far above the receiver to detect an arrival in the parabolic approximation. The ray model yields arrival *A* without scattering because amplitudes from the ray model are not as reliable.

At receiver one, the rays have upper turning depths of about 100 to 400 m near the source. Some rays have upper turning depths near the surface at the receiver. At receiver two, upper turning depths are about 50 to 400 m near the source and receiver. As has been noted in other experiments, the rays composing each arrival have slightly different turning depths and quite different turning ranges.[14,17,18]

V. CONCLUSION

Sounds from U.S. Navy operations can be identified with ray and parabolic approximations of the wave equation over basin-scales at Sound Surveillance stations. The fidelity is adequate for detecting climatic changes of about 1 s from El Niño and the Southern Oscillation.[7] The next step involves a tomographic inversion to estimate the accuracy with which temperature variations are estimated at climatic scales given the navigational $[O(1)\ \text{km}]$ and timing uncertainties $[O(1)\ \text{s}]$ associated with the source. Models[19–21] indicate that climatic signals can be detected in the presence of these errors. An experimental demonstration is needed.

Although success has been had using rays to identify paths over basin-scales in the ocean,[4,22–24] it appears that the relative amplitudes of the arrivals are estimated with better accuracy from a parabolic approximation[11] and a field of internal waves.[10] This field spreads the pulse resolution from ideal values of about 0.01 s to 0.1 to 0.2 s as observed.

Neither the ray nor parabolic approximations of the wave equation yield the last arrival at either receiver (Figs. 2–4), even with a realistic spectrum of internal waves. A similar problem was solved in another experiment by including a realistic mesoscale.[11] There, the mesoscale created a bias of about $+0.6$ s for sounds traveling near the depths of minimum speed in the waveguide.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research contract N00014-00-C-0317. I thank CDR Rick Shema, (Retired) for providing the data, John Tralies (Naval Air Warfare Center, Patuxent River, MD) for answers to many questions, Bruce Einfalt and Carter Ackerman (Penn State) for computer software, and the reviewers.

[1] M. Ewing and J. L. Worzel, ''Long-range sound transmission,'' in *Propagation of Sound in the Ocean* (Geological Society of America, New York, 1948), Memoir 27, Chap. 3, pp. 1–35.
[2] R. H. Johnson, ''Synthesis of point data and path data in estimating sofar speed,'' J. Geophys. Res. **74**, 4559–4570 (1969).
[3] G. R. Hamilton, ''Time variations of sound speed over long paths in the ocean,'' in *International Workshop on Low-frequency Propagation and Noise*, Woods Hole, MA (1977), Vol. 1, pp. 7–30.
[4] J. L. Spiesberger and K. Metzger, ''Basin scale ocean monitoring with acoustic thermometers,'' Oceanography **5**, 92–98 (1992).
[5] ATOC Consortium (A. B. Baggeroer, T. G. Birdsall, C. Clark, J. A. Colosi, B. D. Cornuelle, D. Costa, B. D. Dushaw, M. Dzieciuch, A. M. G. Forbes, C. Hill, B. M. Howe, J. Marshall, D. Menemenlis, J. A. Mercer, K. Metzger, W. Munk, R. C. Spindel, D. Stammer, P. F. Worcester, and C. Wunsch), ''Ocean climate change: Comparison of acoustic tomography, satellite altimetry, and modeling,'' Science **281**, 1327–1332 (1998).
[6] P. N. Mikhalevsky, A. Gavrilov, A. B. Baggeroer, and M. Slavinsky, ''Experiment tests use of acoustics to monitor temperature and ice in Arctic ocean,'' EOS Trans. Am. Geophys. Union **76**, No. 27 (1995).
[7] J. L. Spiesberger, H. E. Hurlburt, M. Johnson, M. Keller, S. Meyers, and J. J. O'Brien, ''Acoustic thermometry data compared with two ocean models: The importance of Rossby waves and ENSO in modifying the ocean interior,'' Dyn. Atmos. Oceans **26**, 209–240 (1998).
[8] V. A. Del Grosso, ''New equation for the speed of sound in natural waters with comparisons to other equations,'' J. Acoust. Soc. Am. **56**, 1084–1091 (1974).
[9] S. Levitus, ''Climatological atlas of the world ocean,'' in NOAA Prof. Pap. 13. U.S. Government Printing Office, Washington, DC, 1982.
[10] C. Garrett and W. Munk, ''Space-time scales of internal waves,'' Geophys. Fluid Dyn. **2**, 225–264 (1972).
[11] M. A. Wolfson and J. L. Spiesberger, ''Full wave simulation of the forward scattering of sound in a structured ocean: A comparison with observations,'' J. Acoust. Soc. Am. **106**, 1293–1306 (1999).
[12] J. A. Colosi, E. K. Scheer, S. M. Flatte, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, P. F. Worcester, B. M. Howe, J. A. Mercer, R. C. Spindel, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, ''Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3202–3218 (1999).
[13] S. M. Flatte, J. A. Colosi, M. A. Dzieciuch, and P. F. Worcester, ''Acoustic observations of internal-wave strength in the Mid-Pacific in 1989 and 1996,'' J. Acoust. Soc. Am. **100**, 2582 (1996).
[14] 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).
[15] J. B. Bowlin, J. L. Spiesberger, T. F. Duda, and L. E. Freitag, Ocean acoustical ray-tracing software RAY, Woods Hole Oceanographic Technical Report, WHOI-93-10 (1993).
[16] J. L. Spiesberger and F. D. Tappert, ''Kaneohe acoustic thermometer further validated with rays over 3700 km and the demise of the idea of axially trapped energy,'' J. Acoust. Soc. Am. **99**, 173–184 (1996).
[17] J. Simmen, S. M. Flatte, and G. Wang, ''Wavefront folding, chaos, and diffractions for sound propagation through ocean internal waves,'' J. Acoust. Soc. Am. **102**, 239–255 (1997).
[18] J. L. Spiesberger, ''Ocean acoustic tomography: travel-time biases,'' J. Acoust. Soc. Am. **77**, 83–100 (1985).
[19] A. Fabrikant, J. L. Spiesberger, A. Silivra, and H. E. Hurlburt, ''Estimating Climatic Temperature Change in the Ocean with Synthetic Acoustic Apertures,'' IEEE J. Ocean. Eng. **23**(1), 20–25 (1998).
[20] J. L. Spiesberger, A. Fabrikant, A. Silivra, and H. E. Hurlburt, ''Mapping Climatic Temperature Changes in the Ocean with Acoustic Tomography: Navigational Requirements,'' IEEE J. Ocean. Eng. **22**(1), 128–142 (1997).
[21] A. Silivra, J. L. Spiesberger, A. Fabrikant, and H. E. Hurlburt, ''Acoustic tomography at basin-scales and clock errors,'' IEEE J. Ocean. Eng. **22**(1), 143–150 (1997).
[22] J. L. Spiesberger, T. G. Birdsall, K. Metzger, R. A. Knox, C. W. Spofford, and R. C. Spindel, ''Measurements of Gulf Stream meandering and evidence of seasonal thermocline development using long range acoustic transmissions,'' J. Phys. Oceanogr. **13**, 1836–1846 (1983).
[23] D. E. Norris, J. L. Spiesberger, and D. W. Merdes, ''Comparison of basin-scale acoustic transmissions with rays and further evidence for a structured thermal field in the northeast Pacific,'' J. Acoust. Soc. Am. **103**, 182–194 (1998).
[24] B. D. Dushaw, B. M. Howe, J. A. Mercer, R. C. Spindel, and the ATOC Group (A. B. Baggeroer, T. G. Birdsall, C. Clark, J. A. Colosi, B. D. Cornuelle, D. Costa, B. D. Dushaw, M. Dzieciuch, A. M. G. Forbes, B. M. Howe, D. Menemenlis, J. A. Mercer, K. Metzger, W. H. Munk, R. C. Spindel, P. F. Worcester, and C. Wunsch), ''Multimegameter-range acoustic data obtained by bottom-mounted hydrophone arrays for measurement of ocean temperature,'' IEEE J. Ocean. Eng. **24**, 203–214 (1999).

Determination of transport parameters in air-saturated porous materials via reflected ultrasonic waves

Z. E. A. Fellah[a)]
National Institute of Health and Medical Research (INSERM U556), 151 cours Albert Thomas, 69424 Lyon Cedex 03, France

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

S. Berger and W. Lauriks
Laboratorium voor Akoestiek en Thermische Fysica, Katholieke Universiteit Leuven, Celestijnenlaan 200 D, B-3001 Heverlee, Belgium

P. Trompette and J.-Y. Chapelon
National Institute of Health and Medical Research (INSERM U556), 151 cours Albert Thomas, 69424 Lyon Cedex 03, France

(Received 22 May 2003; revised 9 August 2003; accepted 2 September 2003)

An ultrasonic reflectivity method of evaluating the acoustic parameters of porous materials saturated by air (or any other gas) is discussed. The method is based on experimental detection of waves reflected at normal incidence by the first and second interface of the material. This method is based on a temporal model of direct and inverse scattering problems for the propagation of transient ultrasonic waves in a homogeneous isotropic slab of porous material with a rigid frame [Fellah *et al.*, J. Acoust. Soc. Am. **113**, 61–73 (2003)]. Generally, the conventional ultrasonic approach can be used to determine tortuosity, and viscous and thermal characteristic lengths via transmitted waves. Porosity cannot be estimated in transmitted mode because of its very weak sensitivity. First interface use of the reflected wave at oblique incidence leads to the determination of porosity and tortuosity [Fellah *et al.*, J. Acoust. Soc. Am. **113**, 2424–2433 (2003)] but this is not possible at normal incidence. Using experimental data of reflected waves by the first and second interface at normal incidence simultaneously leads to the determination of porosity, tortuosity, viscous and thermal characteristic lengths. As with the classic ultrasonic approach for characterizing porous material saturated with one gas, both characteristic lengths are estimated individually by assuming a given ratio between them. Tests are performed using weakly resistive industrial plastic foams. Experimental and numerical results, and prospects are discussed. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1621393]

PACS numbers: 43.20.Bi, 43.20.Hq [ANN] Pages: 2561–2569

I. INTRODUCTION

The propagation of sound in fluid-saturated porous media with rigid solid frames is of great interest for a wide range of industrial applications. With air as the pore fluid,[1–3] applications can be found in noise control, nondestructive material characterization, thermoacoustically controlled heat transfer, etc.

Many applications, such as medical imaging or inverse scattering, require a study of the behavior of pulses traveling into porous media. When a broadband ultrasound pulse passes through a layer of a medium, the pulse wave form changes as a result of attenuation and dispersion of the medium. The classic method for predicting a change in the wave form of a signal passing through a medium relies on the system's impulse response. According to the theory of linear systems, the output signal is a convolution of the input signal and the system impulse response. Many media, including porous materials and soft tissues, have been observed to have an attenuation function that increases with frequency. As a result, higher frequency components of the pulse are attenuated more than lower frequency components. After passing through the layer, the transmitted pulse is not just a scaled down version of the incident pulse, but has a different shape. Dispersion refers to the phenomenon observed when the phase velocity of a propagating wave changes with frequency. Dispersion causes the propagating pulse wave form to change because wave components with different frequencies travel at different speeds. An understanding of the interaction of ultrasound with a porous medium in both time and frequency domains, and the ability to determine the change of wave form when propagating ultrasound pulses, should be useful in designing array transducers and in quantitative ultrasound tissue characterization.

To cope efficiently with the specific problems appearing in transient acoustic field propagation, new approaches are required.[4,5] At present, most analyses of signal propagation are carried out in the frequency domain using the Fourier

a)Electronic mail: fellah@lyon.inserm.fr

transform, and the results are translated to the time domain, and vice-versa. However this has several limitations. The first is that the transformation is difficult to compute accurately enough numerically for nonanalytical functions. For example, using the Fourier transform to obtain time domain results for a lossy material is a more complicated approach than using true time domain analysis, and the numerical results are less accurate. The second disadvantage is that by working in the frequency domain, some numerical information is lost or hard to recover. For example, in the case of noisy data it may be difficult to reconstruct the chronological events of a signal by phase unwrapping. Consequently, it is difficult to obtain a deep understanding of transient signal propagation using the frequency domain method.

A time domain approach differs from a frequency analysis in that the susceptibility functions describing viscous and thermal effects are convolution operators acting on velocity and pressure, and therefore a different algebraic formalism must be applied to solve the wave equation. The time domain response of the material is described by an instantaneous response and a ''susceptibility'' kernel responsible for memory effects. The attraction of a time domain based approach is that analysis is naturally bounded by the finite duration of ultrasonic pressures and is consequently the most appropriate approach for the transient signal.

In this work, we present a reflectivity method for measuring the physical parameters describing the propagation of ultrasonic pulses in air-saturated porous materials. This method is based on a temporal model of direct and inverse scattering problems for the propagation of transient ultrasonic waves in a homogeneous isotropic slab of porous material with a rigid frame, initially introduced by the authors.[6] The principle of this method is to measure the wave reflected by the first and second interface of a slab of porous material, and to solve the inverse problem in order to estimate the physical parameters (porosity, tortuosity, viscous and thermal characteristic lengths). The advantage of this method over classic ultrasonic methods is that all these parameters can be determined simultaneously using experimental reflected data (the ratio between the viscous and thermal characteristic lengths is fixed as in classical acoustical methods[7,8]).

The outline of this paper is as follows. Section II recalls a time domain model and the basic equations of wave propagation in porous material. Section III is devoted to the direct problem and the expression of reflection and transmission kernels in the time domain at normal incidence. In Sec. IV, the sensitivity of porosity and tortuosity, as well as viscous and thermal characteristic lengths are discussed, showing the effect of each parameter on the waves reflected by the first and second interface. Section V contains the inverse problem and the appropriate procedure, based on the least square method, which is used to estimate the acoustic parameters. Finally in Sec. VI, experimental validation using ultrasonic measurement is discussed for air-saturated industrial plastic foams.

II. MODEL

In porous material acoustics, a distinction can be made between two situations depending on whether the frame is moving or not. In the first case, the wave dynamics due to coupling between the solid frame and the fluid are clearly described by the Biot theory.[9,10] In air-saturated porous media, the structure is generally motionless and the waves propagate only in the fluid. This case is described by the equivalent fluid model which is a particular case in the Biot model, in which fluid-structure interactions are taken into account in two frequency response factors: dynamic tortuosity of the medium $\alpha(\omega)$ given by Johnson *et al.*[11] and dynamic compressibility of the air in the porous material $\beta(\omega)$ given by Allard.[1] In the frequency domain, these factors multiply the fluid density and compressibility, respectively, and show the deviation from fluid behavior in free space as frequency increases. In the time domain, they act as operators and in the asymptotic domain (high frequency approximation) their expressions are given[4,6] by

$$\tilde{\alpha}(t)=\alpha_\infty\left(\delta(t)+\frac{2}{\Lambda}\left(\frac{\eta}{\pi\rho_f}\right)^{1/2}t^{-1/2}\right), \tag{1}$$

$$\tilde{\beta}(t)=\left(\delta(t)+\frac{2(\gamma-1)}{\Lambda'}\left(\frac{\eta}{\pi \mathrm{Pr}\rho_f}\right)^{1/2}t^{-1/2}\right). \tag{2}$$

In these equations, $\delta(t)$ is the Dirac function, $\mathrm{Pr}=0.71$ is the Prandtl number, $\eta=1.84.10^{-5}\,\mathrm{kg\,m\,s^{-1}}$ and $\rho_f=1.23\,\mathrm{kg\,m^{-3}}$ are, respectively, fluid viscosity and fluid density, and $\gamma=1.4$ is the adiabatic constant. The relevant physical parameters of the model are the medium tortuosity α_∞, and viscous and thermal characteristic lengths Λ and Λ'. In this model the time convolution of $t^{-1/2}$ with a function is interpreted as a semi-derivative operator according to the definition of the fractional derivative of order ν given in Samko *et al.*[12]

$$D^\nu[x(t)]=\frac{1}{\Gamma(-\nu)}\int_0^t(t-u)^{-\nu-1}x(u)du, \tag{3}$$

where $\Gamma(x)$ is the gamma function.

In this framework, the basic equations[4,6] for our model can be written as

$$\rho_f\tilde{\alpha}(t)*\frac{\partial v_i}{\partial t}=-\nabla_i p, \quad \frac{\tilde{\beta}(t)}{K_a}*\frac{\partial p}{\partial t}=-\nabla\cdot\mathbf{v}, \tag{4}$$

where $*$ denotes the time convolution operation, p is acoustic pressure, $\mathbf{v}$ is particle velocity, and $K_a=141\,820\,\mathrm{Pa}$ is the bulk modulus of air. The first equation is the Euler equation, the second one is the constitutive equation.

For a wave propagating along the x axis, these equations become

$$\rho_f\alpha_\infty\frac{\partial v(x,t)}{\partial t}+2\frac{\rho_f\alpha_\infty}{\Lambda}\left(\frac{\eta}{\pi\rho_f}\right)^{1/2}\int_0^t\frac{\partial v(x,t')/\partial t'}{\sqrt{t-t'}}dt' =-\frac{\partial p(x,t)}{\partial x}, \tag{5}$$

$$\frac{1}{K_a}\frac{\partial p(x,t)}{\partial t}+2\frac{\gamma-1}{K_a\Lambda'}\left(\frac{\eta}{\pi \mathrm{Pr}\rho_f}\right)^{1/2}\int_0^t\frac{\partial p(x,t')/\partial t'}{\sqrt{t-t'}}dt' =-\frac{\partial v(x,t)}{\partial x}. \tag{6}$$

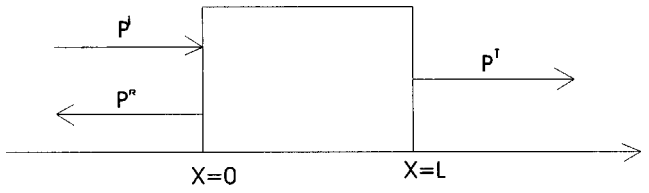


FIG. 1. Problem geometry.

In these equations, the convolutions express the dispersive nature of the porous material. They take into account the memory effects due to the fact that the medium's response to wave excitation is not instantaneous but needs more time to take effect.

The generalized lossy wave equation in the time domain is derived from the basic equations (5) and (6) by elementary calculation as follows:

$$\frac{\partial^2 p(x,t)}{\partial x^2} - A\frac{\partial^2 p(x,t)}{\partial t^2} - B\int_0^t \frac{\partial^2 p(x,t)/\partial t'^2}{\sqrt{t-t'}}dt' - C\frac{\partial p(x,t)}{\partial t} = 0, \tag{7}$$

where coefficients A, B, and C are constants given by

$$A=\frac{\rho_f \alpha_\infty}{K_a}, \quad B=\frac{2\alpha_\infty}{K_a}\sqrt{\frac{\rho_f \eta}{\pi}}\left(\frac{1}{\Lambda}+\frac{\gamma-1}{\sqrt{\mathrm{Pr}}\,\Lambda'}\right), \quad C=\frac{4\alpha_\infty(\gamma-1)\eta}{K_a \Lambda\Lambda'\sqrt{\mathrm{Pr}}}, \tag{8}$$

respectively. The first one concerns the velocity $c=1/\sqrt{\rho_f \alpha_\infty/K_a}$ of the wave in the air in the porous material. α_∞ gives the refractive index of the medium which changes the wave velocity from $c_0=\sqrt{K_a/\rho_f}$ in free space to $c=c_0/\sqrt{\alpha_\infty}$ in the porous medium. The other coefficients essentially depend on the characteristic lengths Λ and Λ' and express fluid-structure viscous and thermal interactions. The constant B governs signal spreading while C is responsible for wave attenuation.

III. DIRECT PROBLEM

The direct scattering problem involves determining the scattered field as well as the internal field, arising when a known incident field impinges on a porous material with known physical properties. Green's function[13] for the modified wave equation in the porous medium must be known in order to compute a solution for the direct problem. In that case, the internal field is given by the time convolution for Green's function with the incident wave, and the reflected and transmitted fields are deduced from the internal field and boundary conditions.

In this section some notation is introduced. The problem geometry is given in Fig. 1. A homogeneous porous material occupies region $0\leqslant x\leqslant L$. This medium is assumed to be isotropic and to have a rigid frame. A short sound pulse impinges normally on the medium from the left. It generates an acoustic pressure field $p(x,t)$ and an acoustic velocity field $v(x,t)$ within the material, which satisfy Eqs. (5) and (6).

To derive the reflection and transmission scattering operators, it is assumed that the pressure field and flow velocity are continuous at the material boundary,

$$p(0^+,t)=p(0^-,t), \quad p(L^-,t)=p(L^+,t), \quad v(0^-,t)=\phi v(0^+,t), \quad v(L^+,t)=\phi v(L^-,t), \tag{9}$$

where ϕ is the porosity of the medium and the positive and negative superscript denotes the limit from left and right, respectively. Assumed initial conditions are

$$p(x,t)|_{t=0}=0, \quad \left.\frac{\partial p}{\partial t}\right|_{t=0}=0, \tag{10}$$

which means that the medium is idle for $t=0$.

If the incident sound wave is generated in region $x\leqslant 0$, then the expression of the acoustic field in the region to the left of the material is the sum of the incident and reflected fields,

$$p_1(x,t)=p^i\left(t-\frac{x}{c_0}\right)+p^r\left(t+\frac{x}{c_0}\right), \quad x<0. \tag{11}$$

Here, $p_1(x,t)$ is the field in the region $x<0$, p^i and p^r denote the incident and reflected fields, respectively. In addition, a transmitted field is produced in the region to the right of the material. This has the form

$$p_3(x,t)=p^t\left(t-\frac{L}{c}-\frac{(x-L)}{c_0}\right), \quad x>L. \tag{12}$$

[$p_3(x,t)$ is the field in region $x>L$ and p^t is the transmitted field.]

The incident and scattered fields are related by scattering operators (i.e., reflection and transmission operators) for the material. These are integral operators represented by

$$p^r(x,t)=\int_0^t \tilde{R}(\tau)p^i\left(t-\tau+\frac{x}{c_0}\right)d\tau = \tilde{R}(t)*p^i(t)*\delta\left(t+\frac{x}{c_0}\right), \tag{13}$$

$$p^t(x,t)=\int_0^t \tilde{T}(\tau)p^i\left(t-\tau-\frac{L}{c}-\frac{(x-L)}{c_0}\right)d\tau = \tilde{T}(t)*p^i(t)*\delta\left(t-\frac{L}{c}-\frac{(x-L)}{c_0}\right). \tag{14}$$

In Eqs. (13) and (14) functions $\tilde{R}$ and $\tilde{T}$ are the reflection and transmission kernels, respectively, for incidence from the left. Note that the lower limit of integration in Eqs. (13) and (14) is given as 0, which is equivalent to assuming that the incident wave front first impinges on the material at $t=0$. The operators $\tilde{R}$ and $\tilde{T}$ are independent of the incident field used in the scattering experiment and depend only on the properties of the materials.

Using boundary and initial conditions (9) and (10), reflection and transmission scattering operators can be derived,[6] taking into account the n-multiple reflections in the material:

$$\tilde{R}(t)=\left(\frac{1-E}{1+E}\right)\sum_{n\geqslant 0}\left(\frac{1-E}{1+E}\right)^{2n}\left[F\left(t,2n\frac{L}{c}\right)-F\left(t,(2n+2)\frac{L}{c}\right)\right], \tag{15}$$

$$\tilde{T}(t)=\frac{4E}{(1+E)^2}\sum_{n\geqslant 0}\left(\frac{1-E}{1+E}\right)^{2n}F\left(t+\frac{L}{c_0},(2n+1)\frac{L}{c}\right). \tag{16}$$

In these equations, $E=\phi/\sqrt{\alpha_\infty}$ and F are the medium's Green function[13] given by

$$F(t,k)=\begin{cases}0 & \text{if } 0\leqslant t\leqslant k\\ \Xi(t)+\Delta\displaystyle\int_0^{t-k}h(t,\xi)d\xi & \text{if } t\geqslant k\end{cases} \tag{17}$$

with

$$\Xi(t)=\frac{b'}{4\sqrt{\pi}}\frac{k}{(t-k)^{3/2}}\exp\left(-\frac{b'^2k^2}{16(t-k)}\right), \tag{18}$$

where $h(\tau,\xi)$ has the following form:

$$h(\tau,\xi)=-\frac{1}{4\pi^{3/2}}\frac{k}{\sqrt{(\tau-\xi)^2-k^2}}\frac{1}{\xi^{3/2}}\times\int_{-1}^{1}\exp\left(-\frac{\chi(\mu,\tau,\xi)}{2}\right)(\chi(\mu,\tau,\xi)-1)\frac{\mu d\mu}{\sqrt{1-\mu^2}} \tag{19}$$

and where $\chi(\mu,\tau,\xi)=(\Delta\mu\sqrt{(\tau-\xi)^2-k^2}+b'(\tau-\xi))^2/8\xi$, $b'=Bc_0^2\sqrt{\pi}$, $c'=C\cdot c_0^2$ and $\Delta^2=b'^2-4c'$.

In most cases of air-saturated porous materials, multiple reflection effects may be negligible because of the high attenuation of sound waves in these media. This depends on the thickness L and physical properties of the porous material. For porous material with high flow resistivity, only the wave reflected by the first interface is seen experimentally and all other reflection contributions are neglected. So, by taking into account only the first reflections at interfaces $x=0$ and $x=L$, the expression of the reflection operator will be given by

$$\tilde{R}(t)=r(t)+\Re(t), \tag{20}$$

with

$$r(t)=\left(\frac{1-E}{1+E}\right)\delta(t),\quad \Re(t)=-\frac{4E(1-E)}{(1+E)^3}F\left(t,\frac{2L}{c}\right). \tag{21}$$

$r(t)$ is the instantaneous response of the porous material corresponding to the reflection contribution at the first interface ($x=0$). $\Re(t)$ is equivalent to reflection at the second interface $x=L$. The part of the wave corresponding to $r(t)$ is not subjected to dispersion but simply multiplied by the factor $(1-E)/(1+E)$. The experimental detection of the reflection contribution by the second interface depends strongly on the sample's thickness and acoustic properties (porosity, tortuosity, viscous and thermal characteristic lengths). It may be interesting to be able to measure the second reflection contribution, because it could offer an alternative method for determining acoustic parameters which are now measured acoustically only in transmission mode.[6] In fact the wave reflected by the second interface depends on the medium's Green function F which is expressed in terms of the ultrasonic parameters. This wave is diffusive compared to the wave reflected by the first interface.

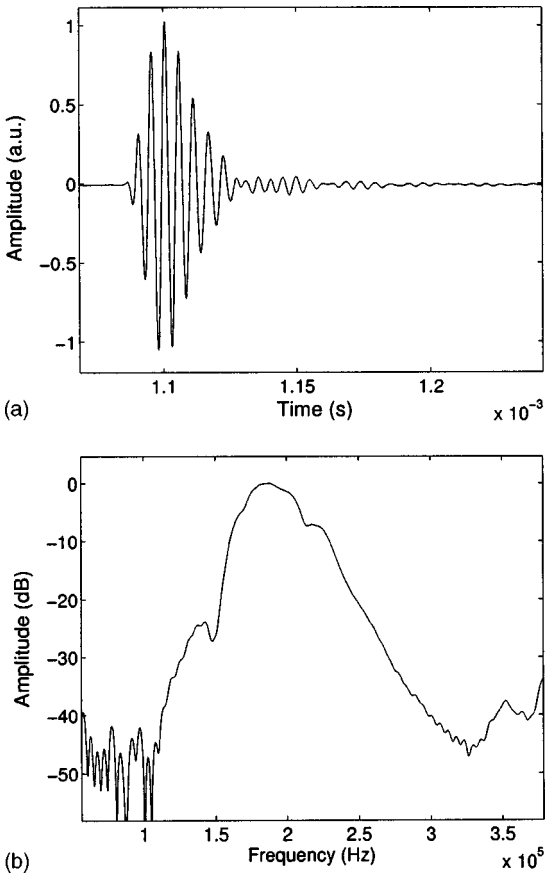


FIG. 2. (a) Incident signal. (b) Spectrum of the incident signal.

IV. ACOUSTIC PARAMETER SENSITIVITY

In this section, numerical simulations of waves reflected by a slab of porous material are run by varying the independent geometrical parameters of a porous medium described acoustically using the theory developed in Sec. III. 50% variation is applied to the governing parameters. A first numerical simulation is produced. The incident signal used in the simulation is given in Fig. 2(a). Its spectrum is given in Fig. 2(b). The numerical values chosen for the physical parameters correspond to quite common acoustic materials, as follows: thickness $L=1$ cm, tortuosity $\alpha_\infty=1.05$, viscous characteristic length $\Lambda=300$ μm, thermal characteristic length $\Lambda'=900$ μm, and porosity $\phi=0.97$. The result of the simulation is a signal as shown in Fig. 3(a). Amplitude is given by an arbitrary unit and the point number given in the

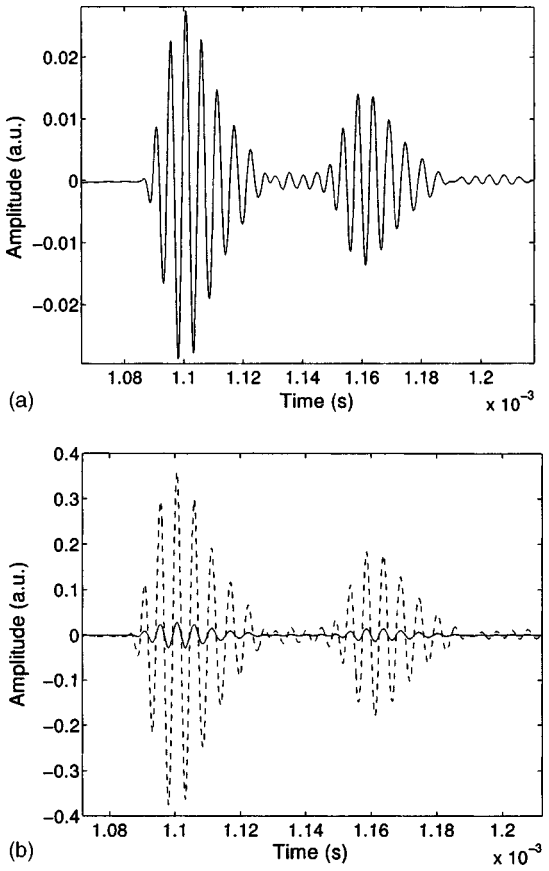


FIG. 3. (a) Simulated reflected signal. (b) Comparison between simulated reflected signals corresponding to $\phi=0.97$ (solid line) and $\phi=0.48$ (dashed line).

abscissa is proportional to time. In Fig. 3, the two signals correspond to the waves reflected by the first and second interface, respectively. The speed of the first wave is the speed of free air 340 $\mathrm{m\,s^{-1}}$ and the speed of the second wave, which propagates inside the porous material, is 331 $\mathrm{m\,s^{-1}}$ ($c=c_0/\sqrt{\alpha_\infty}$).

Figure 3(b) shows the results obtained after reducing porosity by 50% of its initial value. The first signal (solid line) corresponds to the simulated reflected signal for $\phi=0.97$ and the second one (dashed line) to $\phi=0.48$. The values of the other parameters have been kept constant ($L=1$ cm, $\alpha_\infty=1.05$, $\Lambda=300$ μm, and $\Lambda'=900$ μm). The sensitivity of porosity in reflected mode can be seen for a 50% change. There is an important change of amplitude in the two waves (first and second interface). By reducing porosity, the amplitudes of the wave reflected by the first and second interface increase by 775% and 800% of their initial values, respectively. This result can be predicted by the fact that when porosity decreases, the porous material becomes very resistive and the incident wave is then more reflected by the porous material.

When the thickness decreases by 50% of its initial value, the reflected signal also changes. The values of the ultrasonic parameters have been kept constant ($\alpha_\infty=1.05$, $\phi=0.97$, $\Lambda=300$ μm, $\Lambda'=900$ μm). Figure 4 shows a comparison of two reflected signals with two different thicknesses. The first one (solid line) corresponds to a thickness of 1 cm and the second one (dashed line) to 0.5 cm. By reducing the thickness of the material, the distance propagated by the second reflected wave in the porous material falls and is less attenuated. It leads to an increase of amplitude of 40% of its initial value in the second reflected wave.

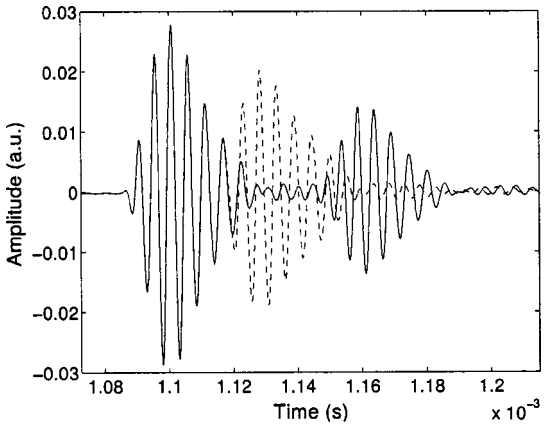


FIG. 4. Comparison between simulated reflected signals corresponding to $L=1$ cm (solid line) and $L=0.5$ cm (dashed line).

Now let us look at the effect of tortuosity. Figure 5 shows a comparison of two simulated reflected signals, the first (solid line) corresponding to $\alpha_\infty=1.05$ and the second (dashed line) to $\alpha_\infty=1.57$. In this simulation, the values of the thickness, porosity, viscous and thermal characteristic lengths have been kept constant ($L=1$ cm, $\Lambda=300$ μm, $\Lambda'=900$ μm, and $\phi=0.97$). From the two signals of Fig. 5, it can be seen that tortuosity plays an important part in reflected waves. By increasing the value of tortuosity, the speed of the wave reflected by the second interface has fallen from 331 to 270 $\mathrm{m\,s^{-1}}$. However, the amplitude of the two waves increases, the wave reflected by the first interface increases by 420% of its initial amplitude and the wave reflected by the second interface increases by 250% of its initial amplitude. On increasing tortuosity, the inertial couplings between fluid and structure also increase, the porous material becomes more resistive, and then the reflected operators [Eq. (21)] of the first and second interface increase. At the same time, the wave reflected by the second interface is much delayed and its amplitude is less amplified compared with the wave reflected by the first interface. This can be ex-

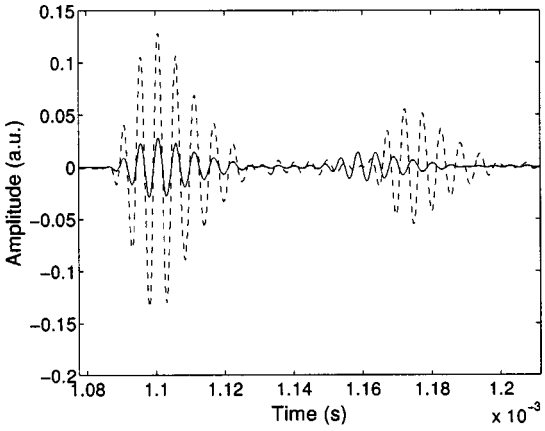


FIG. 5. Comparison between simulated reflected signals corresponding to $\alpha_\infty=1.05$ (solid line) and $\alpha_\infty=1.57$ (dashed line).

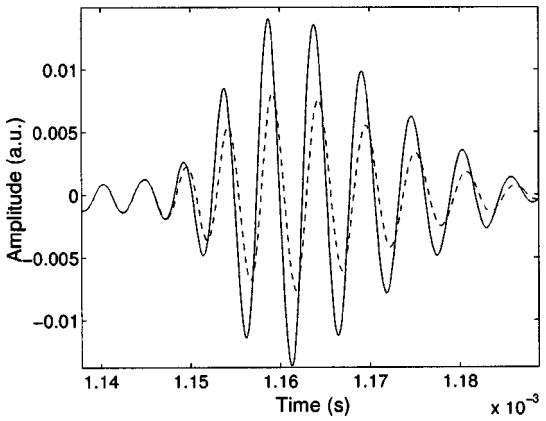


FIG. 6. Comparison between simulated reflected signals at the second interface corresponding to $\Lambda=300$ μm (solid line) and $\Lambda=150$ μm (dashed line).

plained by additional inertial losses in the second reflected wave due to the propagation phenomenon.

Figure 6 shows the sensitivity of viscous characteristic length Λ on the reflected wave by the second interface (the reflected wave by the first interface is not affected by Λ). The signal plotted as a solid line corresponds to $\Lambda=300$ μm and the dashed line transmitted signal to $\Lambda=150$ μm. The values of the other parameters have been kept constant ($L=1$ cm, $\alpha_\infty=1.05$, $\Lambda'=900$ μm, and $\phi=0.97$). It can be seen that the amplitude of the reflected wave by the second interface has fallen by 85% of its initial value. A very small change occurs at the speed of the wave reflected by the second interface due to the dispersion phenomenon governed by Λ. This length appears in the second term in the asymptotic expression of dynamic tortuosity at high frequencies [Eq. (1)], and as shown by this numerical simulation, Λ plays a less important role in reflected waves than tortuosity, which appears in the first order of expression of the tortuosity operator $\tilde{\alpha}(t)$ [Eq. (1)]. In transmission mode, viscous characteristic length plays an important role in dispersion, especially for resistive porous materials.

Figure 7 shows a comparison of two simulated reflected signals corresponding to two different thermal characteristic lengths Λ' (only the reflected wave by the second interface is presented in Fig. 7, the reflected wave by the first interface is not affected by Λ'). The first one (solid line) corresponds to $\Lambda'=900$ μm and the second one (dashed line) to $\Lambda'=450$ μm. A small change is only observed in the wave reflected by the second interface, in which its amplitude has fallen by 8% of its initial value.

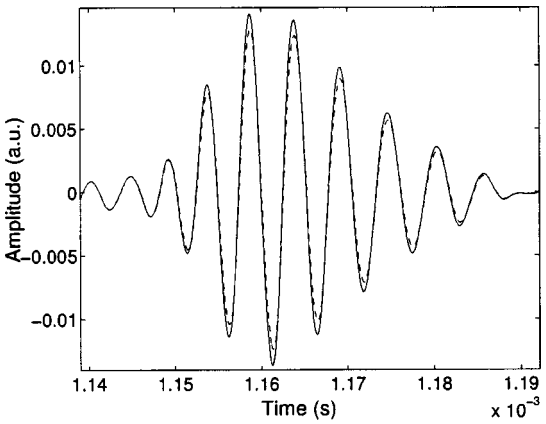


FIG. 7. Comparison between simulated reflected signals at the second interface corresponding to $\Lambda'=900$ μm (solid line) and $\Lambda'=450$ μm (dashed line).

From this study, we can gain an insight into the sensitivity of each physical parameter used in this theory. The parameters which are involved in propagation and dispersion phenomena such as viscous and thermal characteristic lengths Λ and Λ' are only involved in the wave reflected by the second interface wave form. This wave propagates inside the porous material and is thus subjected to viscous and thermal interactions between fluid and structure, responsible for damping the acoustic wave in porous material. However, the wave reflected by the first interface is immediately reflected by the material and does not propagate inside the porous medium. This wave is not sensitive to parameters describing dispersion phenomena such as Λ and Λ'.

We also note from these simulations that some parameters such as porosity ϕ, tortuosity α_∞, and thickness play an important role simultaneously in both waves reflected by the first and second interface wave forms compared to other parameters which act on the wave form of just one of the two waves.

V. INVERSE PROBLEM

The propagation of acoustic waves in a slab of porous material in the asymptotic domain is characterized by four parameters, namely, porosity ϕ, tortuosity α_∞, viscous characteristic length Λ, and thermal characteristic length Λ', the values of which are crucial for the behavior of sound waves in such materials. So, it is of some importance to develop new experimental methods and efficient tools for their estimation. Therefore, a basic inverse problem associated with the slab may be stated as follows: by measuring signals transmitted and/or reflected outside the slab, find the values of the medium's parameters. The inverse problem has been solved in transmission mode in Ref. 6 and an estimate of α_∞, Λ, and Λ' has been made, giving a good correlation between experiment and theory. Porosity has not been estimated in transmitted mode because of its very weak sensitivity. The inverse problem for reflected waves has been solved only for the wave reflected by the first interface at oblique incidence, giving good results for tortuosity and porosity in plastic foams.[14] The determination of porosity and tortuosity is not possible at normal incidence from the reflected wave at the first interface. Viscous and thermal characteristic lengths have not been estimated, because they do not depend on the wave reflected by the first interface. It may be interesting to develop a method to measure all the parameters in one mode (reflection or transmission), although it seems to be impossible in transmitted mode.[6] However for weakly resistive porous materials, such as some plastic foams and fibrous materials, it may be possible to detect the wave reflected by the second interface experimentally, and this gives the same physical information as the transmitted wave, so we can then hope to use these experimental data (waves reflected by the first and second interface) to estimate

all the physical parameters describing propagation. A study of the sensitivity of the physical parameters involved in the reflected signal shows that it may be possible to obtain the effect of each parameter in one or two reflected waves.

As shown in Sec. V, solution of the direct problem involves a system of two operators expressed as functions of ϕ, α_∞, Λ, and Λ'. The inversion algorithm for finding the values of the slab parameters in reflected mode is based on an appropriate procedure: find the values of the parameters ϕ, α_∞, Λ, and Λ' such that the reflected signal describes the scattering problem in the best possible way (e.g., in the least-squares sense). Because of the weak sensitivity of Λ' in the porous material's reflected response, the Λ'/Λ ratio is set to 3 (Λ'/Λ is classically set from 2 to 3 for plastic foams[7,8]).

The inverse problem is to find values for parameters ϕ, α_∞, and Λ which minimize the function

$$U(\phi,\alpha_\infty,\Lambda)=\int_0^t (p^r_{\text{expt}}(x,t)-p^r(x,t))^2 dt,$$

where $p^r_{\text{expt}}(x,t)$ is the experimentally determined reflected signal, $p^r(x,t)$ is the reflected wave predicted from Eq. (13). However, because the equations are nonlinear, the analytical method of solving the inverse problem using the conventional least-squares method is tedious. In our case, a numerical solution of the least-squares method can be found which minimizes $U(\phi,\alpha_\infty,\Lambda)$ defined by

$$U(\phi,\alpha_\infty,\Lambda)=\sum_{i=1}^{i=N} (p^r_{\text{expt}}(x,t_i)-p^r(x,t_i))^2, \tag{22}$$

where $p^r_{\text{expt}}(x,t_i)_{i=1,2,\ldots,N}$ is the discrete set of values of the experimental reflected signal and $p^r(x,t_i)_{i=1,2,\ldots,N}$ is the discrete set of values of the simulated reflected signal. Section VI discusses solution of the inverse problem based on experimental reflected data.

VI. ULTRASONIC MEASUREMENT

In application of this model, some numerical simulations are compared with experimental results. Experiments are performed in air using a broadband Ultran NCT202 transducer with a central frequency of 190 kHz in air and a bandwith of 6 dB extending from 150 to 230 kHz. This transducer is used simultaneously as a transmitter and receiver to detect the reflected wave. Pulses of 400 V are provided by a 5052PR Panametrics pulser/receiver. The signals received are amplified to 90 dB and filtered above 1 MHz to avoid high frequency noise (energy is totally filtered by the sample in this upper frequency domain by the wave reflected by the second interface). Electronic interference is eliminated by 1000 acquisition averages. The experimental setup is shown in Fig. 8. Signal duration is important in solving the inverse problem; its spectrum must verify the condition of high frequency approximation[4,6] referred to in Sec. II.

Consider a sample of plastic foam, M1, of two different thicknesses, the first (M1-a) with a thickness of 1.00 ± 0.01 cm and the second (M1-b) 0.62 ± 0.01 cm. Sample M1 was characterized using classic methods[6,14] given the following physical parameters: $\phi=0.94\pm0.005$, $\alpha_\infty=1.07\pm0.005$, $\Lambda=200\pm5$ μm, $\Lambda'=600\pm15$ μm.

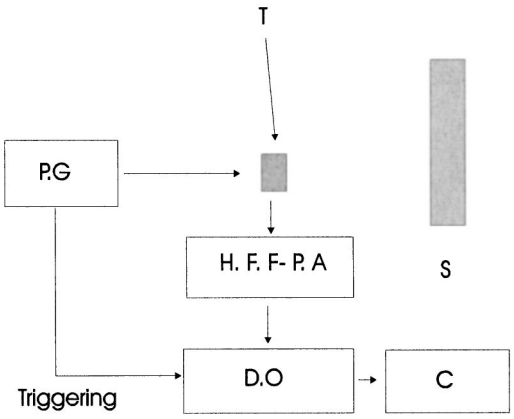


FIG. 8. Experimental setup of the ultrasonic measurements in reflected mode. PG: pulse generator, HF F-PA: high frequency filtering -pre-amplifier, DO: digital oscilloscope, C: computer, T: transducer, S: sample.

Consider the sample M1-a: Figure 9 shows the experimental incident signal generated by the transducer. After solving the inverse problem numerically, we find the following optimized values $\phi=0.945\pm0.005$, $\alpha_\infty=1.065\pm0.005$, $\Lambda=210\pm5$ μm, and $\Lambda'=630\pm15$ μm. Using these values, we present in Figs. 10(a), (b), and (c) the variations in the minimization function U given in Eq. (22) with porosity, tortuosity, and viscous characteristic length, respectively. In Fig. 11, we show a comparison between an experimental reflected signal and simulated reflected signals for the optimized values of porosity, tortuosity, viscous and thermal characteristic lengths. The difference between the two curves is slight, which leads us to conclude that the optimized values of the physical parameters are correct.

Let us now solve the inverse problem for sample M1-b. The experimental incident signal generated by the transducer is shown in Fig. 12. The optimized values of these parameters are $\phi=0.938\pm0.005$, $\alpha_\infty=1.07\pm0.005$, $\Lambda=170\pm5$ μm, and $\Lambda'=510\pm15$ μm. Figure 13 shows a comparison of an experimental reflected signal and a simulated signal obtained by optimization from the inverse problem. Here, again, the correlation of theoretical prediction and experimental data is good.

Using another plastic foam M2 (thickness 1 cm), the incident signal generated by the transducer is given in Fig.

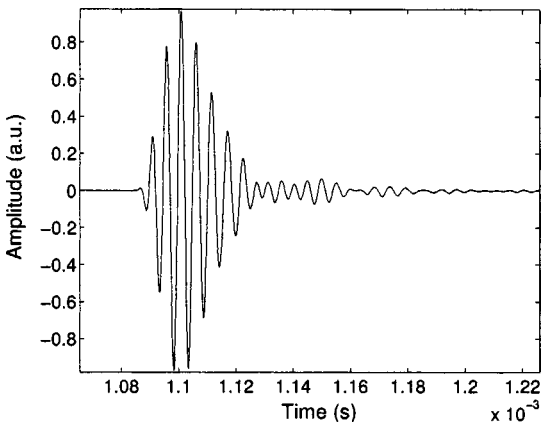


FIG. 9. Experimental incident signal generated by the transducer for the plastic foam sample M1-a.

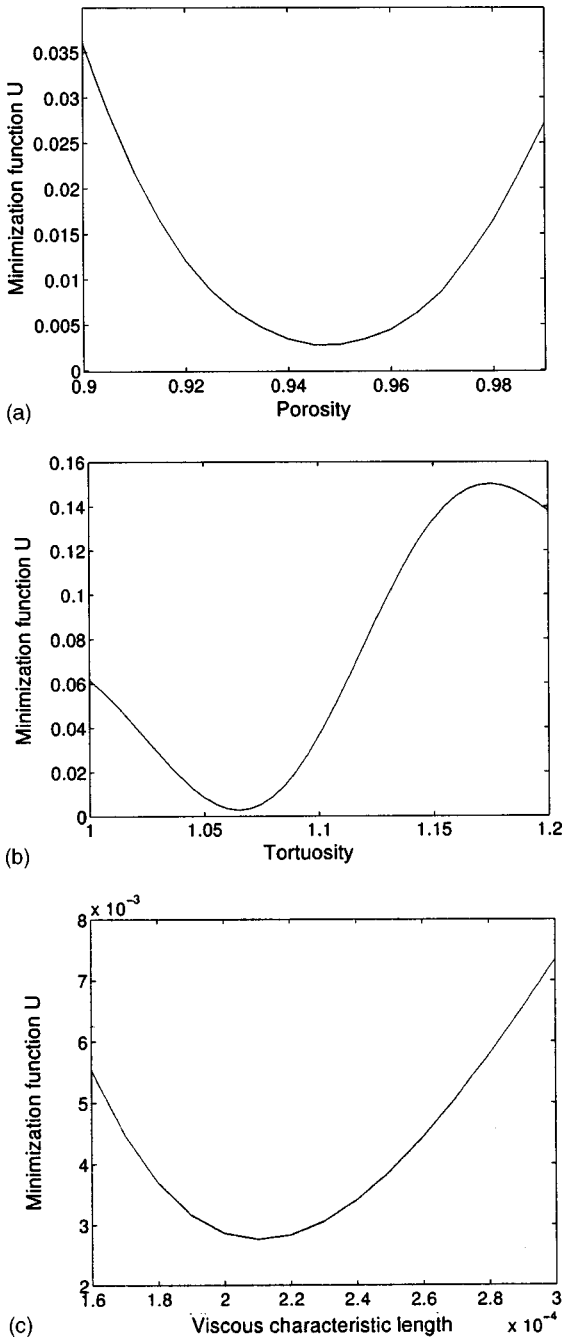


FIG. 10. (a) Variation of the minimization function U with porosity for $\alpha_\infty=1.065$ and $\Lambda=210$ μm (sample M1-a). (b) Variation of the minimization function U with tortuosity for $\phi=0.945$ and $\Lambda=210$ μm (sample M1-a). (c) Variation of the minimization function U with viscous characteristic length for $\phi=0.945$ and $\alpha_\infty=1.065$ (sample M1-a).

14. The results, after solving the inverse problem for variation of the minimization function with porosity, tortuosity and viscous characteristic lengths, are: $\phi=0.97\pm0.005$, $\alpha_\infty=1.075\pm0.005$, $\Lambda=380\pm5$ μm, and $\Lambda'=1140\pm15$ μm. In Fig. 15, we show a comparison of an experimental reflected signal and a reflected simulated signal using optimized values of physical parameters. Classic methods[6,14] give the following characteristics; $\phi=0.97\pm0.05$, $\alpha_\infty=1.07\pm0.005$, $\Lambda=360\pm5$ μm, and $\Lambda'=1080\pm15$ μm.

It can be seen that for the three samples, M1-a, M1-b, and M2, the optimized values obtained using this method are

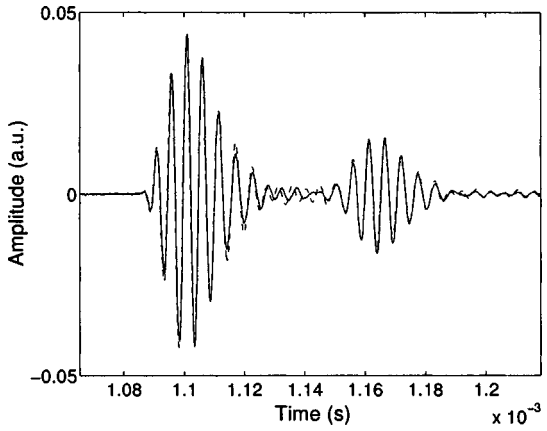


FIG. 11. Comparison between experimental reflected signal (solid line) and simulated reflected signal (dashed line) for the plastic foam sample M1-a ($\alpha_\infty=1.065$, $\phi=0.945$, and $\Lambda=210$ μm).

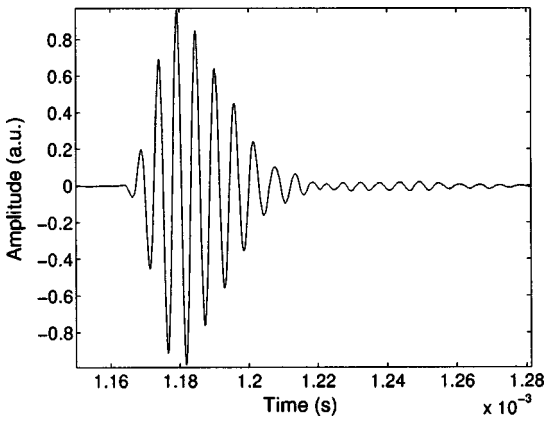


FIG. 12. Experimental incident signal generated by the transducer for the plastic foam sample M1-b.

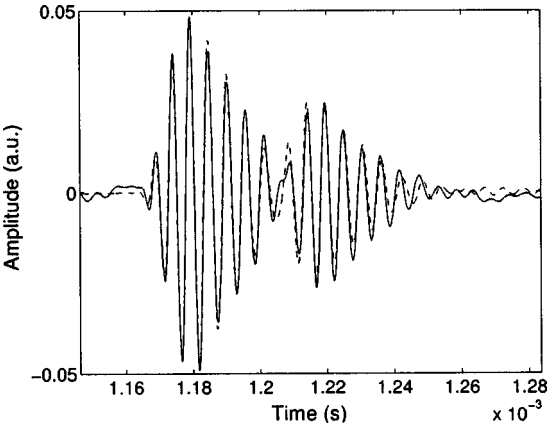


FIG. 13. Comparison between experimental reflected signal (solid line) and simulated reflected signal (dashed line) obtained by optimization from the inverse problem, for the plastic foam sample M1-b ($\alpha_\infty=1.070$, $\phi=0.938$, and $\Lambda=170$ μm).

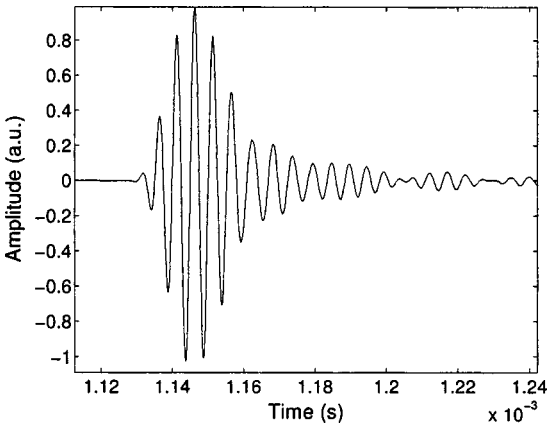


FIG. 14. Experimental incident signal generated by the transducer for the plastic foam sample M2.

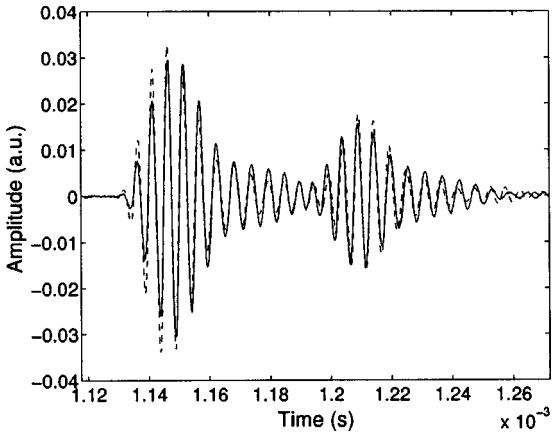


FIG. 15. Comparison between experimental reflected signal (solid line) and simulated reflected signal (dashed line) obtained by optimization from the inverse problem, for the sample M2 ($\alpha_\infty = 1.075$, $\phi = 0.97$, and $\Lambda = 380$ μm).

close to those produced using classic methods.[6,14]

This method seems to be efficient for estimating the physical parameters needed to describe sound propagation in air-saturated porous media such as plastic foams. However, when the reflected wave by the second interface cannot be detected experimentally because of strong damping of the acoustic wave inside the porous material, it will be not possible to estimate the transport parameters since the reflected wave at the first interface gives just a relation between α_∞ and ϕ at normal incidence.

VII. CONCLUSION

A method for measuring transport parameters in porous materials simultaneously, using measurements of waves reflected at the first and second interface, has been proposed. This method is based on a temporal model of direct and inverse scattering problems affecting the propagation of transient ultrasonic waves in a homogeneous isotropic slab of porous material with a rigid frame.

Generally, porosity and tortuosity can be evaluated by the wave reflected at the first interface at oblique incidence[14] but this is not possible at normal incidence. Viscous and thermal characteristic lengths can be estimated only by transmitted waves. Porosity cannot be determined from transmit ted waves because of its weak sensitivity in this mode. The advantage of the proposed method is that all the parameters can be determined at normal incidence (the ratio between viscous and thermal lengths is fixed as in classical methods[7,8] based on transmission measurement).

The main principle of this method is the experimental detection of reflected contributions from the first and the second interface of the medium. The properties of these two contributions are used to estimate the four acoustical parameters needed for ultrasonic propagation in porous material with a rigid frame, by solving the inverse problem.

Studying the sensitivity of each reflected wave parameter demonstrates the importance of each contribution (first and second interface) for the inversion. Numerical and experimental validation for weak resistive air-saturated industrial plastic foams is given to validate this method.

[1] J. F. Allard, *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Chapman and Hall, London, 1993).
[2] K. Attenborough, "On the acoustic slow wave in air-filled granular media," J. Acoust. Soc. Am. **81**, 93–102 (1987).
[3] K. Attenborough, "Models for the acoustical properties of air-saturated granular media," Acta Acustica **1**, 213 (1993).
[4] Z. E. A. Fellah and C. Depollier, "Transient acoustic wave propagation in rigid porous media: A time-domain approach," J. Acoust. Soc. Am. **107**, 683–688 (2000).
[5] T. L. Szabo, "Time domain wave equations for lossy media obeying a frequency power law," J. Acoust. Soc. Am. **96**, 491–500 (1994).
[6] Z. E. A. Fellah, M. Fellah, W. Lauriks, and C. Depollier, "Direct and inverse scattering of transient acoustic waves by a slab of rigid porous material," J. Acoust. Soc. Am. **113**, 61–73 (2003).
[7] C. Ayrault, A. Moussatov, B. Castagnède, and D. Lafarge, "Ultrasonic characterization of plastic foams via measurements with static pressure variations," Appl. Phys. Lett. **74**, 3224–3226 (1999).
[8] A. Moussatov, C. Ayrault, and B. Castagnède, "Porous material characterization-ultrasonic method for estimation of tortuosity and characteristic length using a barometric chamber," Ultrasonics **39**, 195–202 (2001).
[9] M. A. Biot, "The theory of propagation of elastic waves in fluid-saturated porous solid. I. Low frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[10] M. A. Biot, "The theory of propagation of elastic waves in fluid-saturated porous solid. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[11] 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).
[12] S. G. Samko, A. A. Kilbas, and O. I. Marichev, *Fractional Integrals and Derivatives: Theory and Applications* (Gordon and Breach Science Publishers, Amsterdam, 1993).
[13] Z. E. A. Fellah, M. Fellah, W. Lauriks, C. Depollier, J. Y. Chapelon, and Y. C. Angel, "Solution in time domain of ultrasonic propagation equation in a porous material," Wave Motion **38**, 151–163 (2003).
[14] Z. E. A. Fellah, S. Berger, W. Lauriks, C. Depollier, C. Aristégui, and J.-Y. Chapelon, "Measuring the porosity and the tortuosity of porous materials via reflected waves at oblique incidence," J. Acoust. Soc. Am. **113**, 2424–2433 (2003).

Modified Szabo's wave equation models for lossy media obeying frequency power law

W. Chen[a] and S. Holm[b]
Simula Research Laboratory, P.O. Box 134, 1325 Lysaker, Oslo, Norway

(Received 31 December 2002; revised 27 August 2003; accepted 2 September 2003)

Szabo's models of acoustic attenuation [Szabo, J. Acoust. Soc. Am. **96**(1), 491–500 (1994)] comply well with the empirical frequency power law involving noninteger and odd-integer exponent coefficients while guaranteeing causality, but nevertheless encounter the troublesome issues of hypersingular improper integral and obscurity in implementing initial conditions. The purpose of this paper is to ease or remove these drawbacks of the Szabo's models via the Caputo fractional derivative concept. The positive time-fractional derivative is also introduced to include the positivity of the attenuation processes. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1621392]

PACS numbers: 43.20.Bi, 43.20.Hq, 43.35.Bf [MO] Pages: 2570–2574

I. INTRODUCTION

The effect of attenuation plays a prominent role in many acoustic and ultrasound applications, for instance, the ultrasound second harmonic imaging and high-intensity focused ultrasound beam for therapeutic surgery. Due to the fractal microstructures of media, such acoustic attenuation typically exhibits a frequency dependency characterized by a power law (He, 1998).

$$S(\mathbf{x}+\Delta\mathbf{x}) = S(\mathbf{x}) e^{-\alpha(\omega)\Delta\mathbf{x}}, \tag{1}$$

$$\alpha(\omega) = \alpha_0 |\omega|^y, \quad y \in [0,2], \tag{2}$$

where S represents the amplitude of an acoustic field variable such as velocity and pressure; ω denotes the angular frequency; $\Delta\mathbf{x}$ is wave propagation distance; and the tissue-specific coefficients α_0 and y are empirically obtained by fitting measured data. The power law (2) is applicable for a broad range of frequency of practical interest. It has long been known that the time-space mathematical model of the attenuation formulas (1) and (2) with a power $y \neq 0,2$ is not easily constituted with the standard integer-order partial differential equation (PDE) methodology. Thus, the attenuations involving $y \neq 0,2$ are also called the anomalous attenuations.

Recent decades have witnessed great effort devoted to developing a variety of anomalous attenuation models. These modeling approaches, however, have their respective advantages and disadvantages. For example, the frequency-domain model via the Laplace transform is simple but limited to linear cases, and also computationally expensive (Wismer and Ludwig, 1995; Rossikhin and Shitikova, 1997; Ginter, 2000). The adaptive Rayleigh model (Wojcik *et al.*, 1995) is as easy as the common PDE models to be implemented but may not be feasible to important broadband pulse propagation (Chen and Holm, 2002). The multiple relaxation model (Nachman *et al.*, 1990; Mast *et al.*, 2001) works well for relaxation-dominated attenuations but requires large amounts of computational effort (Yuan *et al.*, 1999), and demands strenuous effort to estimate many obscure relaxation parameters, and may not be applicable to anomalous attenuations involving singular memory processes (Hanyga, 1999). The fractional calculus models have better accuracy for general anomalous attenuations with fewer necessary parameters (Baglegy and Torvik, 1983; Makris and Constantinou, 1991; Ochmann and Makarov, 1993; Enelund, 1996; Gaul, 1999), but are mathematically more complicated and inflict nontrivial cost of the numerical solution. Recently, Szabo (1993, 1994a, b) presented the convolution integral wave models, which comply well with arbitrary power-law frequency dependences while guaranteeing causality and involving only the parameters α_0 and y. However, it is not straightforward to implement the initial conditions in the Szabo's models. In addition, the severely hypersingular improper integral hinders the numerical solution of the models.

The paper aims to remedy the above-mentioned drawbacks of the Szabo's models. It is found that the Szabo's models can be recast with the Riemann–Liouville fractional derivative for noninteger power y. Then, the basic strategy of this study is that the Riemann–Liouville fractional derivative is replaced by the Caputo fractional derivative to derive a modified Szabo's model, where the initial conditions can be easily prescribed and the hypersingular integral is regularized inherently. To simplify the expression and formalize for potential applications in a standard setting, a positive time-fractional derivative is also introduced to include the positivity in modeling of the frequency power-law attenuation with arbitrary exponent y.

II. SZABO'S CONVOLUTIONAL INTEGRAL MODEL

By a thorough examining of the dispersion equations of the frequency-independent damped wave equation and frequency-squared dependent thermoviscous wave equation, Szabo (1994a) presented a general dispersion equation for arbitrary power y of frequency-dependent attenuation

$$k^2 = (\omega/c_0)^2 + i2(\omega/c_0)a_0|\omega|^y, \tag{3}$$

where the wave number $k = \beta + i\alpha$ corresponds to the Fourier transform of the space variable. β and α are, respectively, related to the dispersion and the attenuation. For more details on the induction of the dispersion equation (3) see

[a]Electronic mail: wenc@simula.no
[b]Electronic mail: sverre.holm@simula.no

Szabo (1994a). Under the so-called conservative smallness condition [Eq. (22) of Szabo, 1994a]

$$\alpha/\beta \approx \alpha_0 |\omega|_{\lim}^{y}/\beta_0 = \alpha_0 |\omega|_{\lim}^{y-1} c_0 \leqslant 0.1, \tag{4}$$

the dispersion equation (3) approximates the power law (2) quite well and leads to the Szabo's time convolutional integral model

$$\Delta p = \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} + \frac{2\alpha_0}{c_0} L_{\alpha,y,t} * p, \tag{5}$$

where* denotes the convolution operation, and

$$L_{\alpha,y,t} = \begin{cases} \delta(t) & y = 0, \\ \dfrac{2}{\pi t^3}, & y = 1, \\ -\dfrac{\Gamma(y+2)\cos[(y+1)\pi/2]}{\pi |t|^{y+2}}, & y \neq 1,\ 0 < y < 2, \\ -\delta^3(t), & y = 2, \end{cases} \tag{6}$$

in which Γ denotes the gamma function, and δ represents the Dirac delta function. Equation (5) uses the relationship

$$\delta^n(t) * p(t) = \frac{\partial^n p}{\partial t^n}. \tag{7}$$

Note that the convolution operation in (5) implies infinite limits on the time integral, and (5) is therefore not causal. To add causality in terms of the time-causal relation (Szabo, 1994a), $L_{a,y,t}$ is replaced by

$$S_y(p) = 2H(t) L_{\alpha,y,t}, \quad 0 \leqslant y \leqslant 2, \tag{8}$$

where $H(t)$ is the Heaviside operator. Assuming that $t=0$ is the initial quiescence instant, $S_y(p)$ is expressed as

$$S_y(p) = \begin{cases} \partial p/\partial t, & y = 0, \\ -\dfrac{2\Gamma(y+2)\cos[(y+1)\pi/2]}{\pi} \displaystyle\int_0^t \frac{p(\tau)}{(t-\tau)^{y+2}}\, d\tau, & \\ \qquad 0 < y < 2, & \\ -\partial^3 p/\partial t^3, & y = 2. \end{cases} \tag{9}$$

Note that $y=1$ term in (9) is included in the noninteger expression. $S_y(p)$ is called the Szabo operator. Equation (5) is revised as

$$\Delta p = \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} + \frac{2\alpha_0}{c_0} S_y(p). \tag{10}$$

It is observed that the Szabo operator encounters the hypersingular improper integral when $y \neq 0, 2$ and does not explicitly show how to implement the initial conditions.

Szabo (1994a) and Szabo and Wu (2000) also briefly mentioned that the Szabo's model shared some similarity with the fractional derivative equation models of the frequency-dependent attenuation such as the Bagley–Torvik attenuation model (Bagley and Torvik, 1983). One essential difference between both types of anomalous attenuation models, however, is that the Szabo's model guarantees the positive definition operation of noninteger y by using (Lighthill, 1962)

$$\mathrm{FT}^-(|\omega|^y) = \Gamma(y+1)\cos[(y+1)\pi/2]/(\pi |t|^{y+1}) = s(t), \tag{11}$$

where FT^- is the inverse Fourier transform operation. To deal with the odd integer y cases, Szabo (1994a) employed the inverse Fourier transform of the sign function (Lighthill, 1962)

$$\mathrm{FT}^-(\omega^y \operatorname{sgn}(\omega)) = \Gamma(y+1)(-1)^{(y+1)/2}/(\pi t^{y+1}). \tag{12}$$

It is found that (12) can be actually covered within (11). The inverse Fourier transforms (11) and (12) play an essential part in Szabo's creating the causal lossy wave equation model (5).

Compared with the Szabo's models, the fractional derivative attenuation models in the sense of Riemann–Liouville or Caputo fails to recover dispersion equation (3) due to

$$\mathrm{FT}^-[(-i\omega)^y P] = \partial^y p/\partial t^y, \tag{13}$$

when y is a fraction or an odd integer.

On the other hand, the fractional derivative equation models in the Caputo sense, however, have an obvious advantage over the Szabo's model in that they inherently regularize the hypersingular improper integral and naturally require and implement the integer-order initial conditions (Seredynska and Hanyga, 2000). Section III presents a model equation combining the merits of the Szabo's model and the Caputo fractional derivative models via a newly defined positive time-fractional derivative.

III. MODIFIED SZABO'S WAVE EQUATION MODEL WITH THE CAPUTO FRACTIONAL DERIVATIVE

In terms of the Riemann–Liouville fractional derivative $D*$, defined in the Appendix, the Szabo's operator (9) is rewritten as

$$S_y(p) = \begin{cases} \partial p/\partial t, & y = 0, \\ \dfrac{4}{\pi}\displaystyle\int_0^t \frac{p(\tau)}{(t-\tau)^3}\, d\tau, & y = 1, \\ -\dfrac{2\Gamma(y+2)\Gamma[-(y+1)]\cos[(y+1)\pi/2]}{\pi} & \\ \qquad \times D_*^{y+1} p, & y \neq 1,\ 0 < y < 2, \\ -\partial^3 p/\partial t^3, & y = 2. \end{cases} \tag{14}$$

Note that the initial quiescent instant is set to zero in (14). It is found from (14) that the Szabo's model implicitly uses the Riemann–Liouville derivative in the cases of the noninteger exponent y. As discussed in the Appendix, the Caputo fractional derivative is more appropriate for modeling of engineering initial value problems, while the Riemann–Liouville fractional derivative model is not well posed (Seredynska and Hanyga, 2000). Thus, the Szabo's loss operator is modified via the Caputo fractional derivative [see (A5) and (A11) in the Appendix] as

$$Q_y(p)=\begin{cases}\partial p/\partial t, & y=0,\\ \dfrac{2}{\pi}\displaystyle\int_0^t \frac{D^2p(\tau)}{(t-\tau)}\,d\tau, & y=1,\\ -\dfrac{2\Gamma[-(y+1)]\Gamma(y+2)\cos[(y+1)\pi/2]}{\pi} & \\ \quad\times D^{y+1}p, & y\neq 1,\ 0<y<2,\\ -\partial^3 p/\partial t^3, & y=2.\end{cases} \tag{15}$$

Note that when $0<y<2$, the Riemann–Liouville derivative in (14) is replaced by the Caputo derivative in (15), and when $y=1$, the definition $Q_y(p)$ still has the singularity at the origin. Substituting the loss operator $Q_y(p)$ instead of $S_y(p)$ into (10) produces the linear wave equation model of the power-law frequency-dependent attenuation

$$\Delta p=\frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2}+\frac{2\alpha_0}{c_0}Q_y(p). \tag{16}$$

The model equation (16) eases the hypersingular improper integral in the Szabo's model (10) through a regularization process of the Caputo derivative [see (A5) in the Appendix] by invoking the integer-order initial conditions naturally. It is worth mentioning that (16) has the same dispersion equation as the Szabo's model equation (10) since Eq. (A11) in the Appendix shows that the Caputo fractional derivative differs from the Riemann–Liouville fractional derivative only in that the former augments a polynomial series with the initial conditions as coefficients, which has no effect on the dispersion equation. Thus, the modified Szabo's model equation (16) holds the causality of the original Szabo's model (10).

IV. A POSITIVE TIME FRACTIONAL DERIVATIVE

The positivity of the attenuation (damping) operation [see power-law formula (2) and dispersion equation (3)] required by the decay of the global energy (Matignon *et al.*, 1998) can be well preserved via a positive fractional calculus operator. The space fractional Laplacian (Hanyga, 2001; Chen and Holm, 2002) is the positive definition operator and thus very suitable to describe the anomalous attenuation behaviors, while the traditional fractional time derivative lacks this crucial positive property. In this section, a positive time fractional derivative in terms of (11) is introduced, whose Fourier transform characterizes the positive operation as follows:

$$\mathrm{FT}^+(D^{|\eta|}p)=|\omega|^{\eta}P(\omega), \quad 0<\eta<2, \tag{17}$$

where FT^+ is the Fourier transform operation; $P(\omega)$ is the Fourier transform of the pressure signal $p(t)$; and

$$D_*^{|\eta|}p=s(t)*p(t)=\frac{1}{q(\eta)}\int_0^t\frac{p(\tau)}{(t-\tau)^{\eta+1}}\,d\tau, \tag{18}$$

$$D^{|\eta|}p=\begin{cases}\dfrac{-1}{\eta q(\eta)}\displaystyle\int_0^t\frac{D^1p(\tau)}{(t-\tau)^{\eta}}\,d\tau, & 0<\eta\leqslant 1,\\ \dfrac{1}{\eta(\eta-1)q(\eta)}\displaystyle\int_0^t\frac{D^2p(\tau)}{(t-\tau)^{\eta-1}}\,d\tau, & 1<\eta<2,\end{cases} \tag{19}$$

where $s(t)$ is defined in (11), and

$$q(\eta)=\frac{\pi}{2\Gamma(\eta+1)\cos[(\eta+1)\pi/2]}. \tag{20}$$

The definitions (18) and (19) of the positive fractional derivative are to combine the Fourier transform relationship (11), respectively, with the Riemann–Liouville fractional derivative (A7) and with the Caputo fractional derivative (A8) in the Appendix to hold the Fourier transform relationship (17). Equation (19) can be expanded by the integration by parts into (18) augmented a regularization series. Note that for $\eta=1$ the positive fractional derivative still confronts singularity at the origin. The positive fractional derivative (19) can also be further generalized by

$$D^{|\eta|}u=\begin{cases}\dfrac{-1}{(\eta-2k)q(\eta)}\displaystyle\int_0^t\frac{D^{2k+1}u(\tau)}{(t-\tau)^{\eta-2k}}\,d\tau, & \\ \quad 2k<\eta\leqslant 2k+1,\\ \dfrac{1}{(\eta-2k)[\eta-(2k+1)]q(\eta)}\displaystyle\int_0^t\frac{D^{2k+2}u(\tau)}{(t-\tau)^{\eta-(2k+1)}}\,d\tau, & \\ \quad 2k+1<\eta<2k+2,\end{cases} \tag{21}$$

where k is a non-negative integer. The following formulas also hold:

$$D^{|\eta|+q}p=D^{|\eta|}D^qp\neq D^qD^{|\eta|}p=D^{q+|\eta|}p, \tag{22}$$

where q is a positive integer number. Note that $D^{|1|}p\neq D_*^{|1|}p\neq D^1p$. It is straightforward to have

$$\mathrm{FT}^+(D^{|\eta|+q}p)=(-i\omega)^q|\omega|^{\eta}P(\omega), \quad 0<y<2. \tag{23}$$

In terms of the new fractional derivative definition, the loss operator $Q_y(p)$ defined in (15) is rewritten as

$$Q_y(p)=\begin{cases}D^1p, & y=0,\\ D^{|y|+1}p, & 0<y<2,\\ -D^3p, & y=2,\end{cases} \tag{24}$$

where $y=0$ corresponds to the damped wave equation, also known as the electromagnetic equation (Szabo, 1994a) or the Telegrapher's equation; and $y=2$ leads to the thermoviscous equation (Blackstock, 1967; Pierce, 1989), also called the augmented wave equation (Johnson and Dudgeon, 1993). In most cases of practical interest $0<y<2$ occurs. It is noted that the fractional time derivative Burgers equation for anomalous attenuations presented by Ochmann and Makarov (1993) requires a sign change of viscous coefficient between $0\leqslant y<1$ and $1<y\leqslant 2$, whereas the definition (19) of the positive time fractional derivative avoids the sign change across $y=1$ in (24).

If defining

$$D^{|\eta|}p=\begin{cases}p, & \eta=0,\\ -D^2p, & \eta=2,\end{cases} \tag{25}$$

then the modified Szabo equation (16) can be rewritten as

$$\Delta p = \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} + \frac{2\alpha_0}{c_0} D^{|y|+1} p. \tag{26}$$

Note that the definition (19) implies that unlike the even-order derivative, the positive fractional time derivative of the odd order is not equivalent to the corresponding standard derivative.

V. CONCLUSIONS

The modified Szabo's model equation (16) includes the Caputo fractional derivative, the existence, uniqueness, and well-posedness (i.e., continuous dependence on data) of whose type of equations have been well proved (Seredynska and Hanyga, 2000). The corresponding numerical solution can also be carried out with the well-developed standard numerical methods for the Caputo fractional derivative equations (Diethelm, 1997; Podlubny, 1999; Diethelm and Ford, 2002). It is worth stressing that the modified Szabo's model is closely related to the singular Volterra equation of the second kind, the numerical solutions of which are discussed in Press *et al.* (1992) and many other publications.

In this study, only the linear model of wave equation was involved. It is straightforward to extend the present strategy to a variety of the corresponding parabolic equations and nonlinear equation models developed by Szabo (1993, 1994a). It is also noted that the mathematical models for anomalous dissipation behaviors are mostly phenomenological (Adhikari, 2000). In other words, these models are to describe attenuation phenomena but do not necessarily reflect various physical and chemical mechanisms behind the scenes. Broadly speaking, there are two types of attenuation (damping) model methodology, space and time operations. The present time-domain model underlies the memory effect, also often called heredity, relaxation, or hysteresis in some publications, which depends on the past history of motion. On the other hand, the space-fractional Laplacian models reflect the fractal microstructures of media and can describe the frequency power-law attenuation quite well (Chen and Holm, 2002). When $y \geqslant 1$, the modified Szabo's time-domain models require the second-order initial condition, which is not readily available in many practical cases, and it is thus reasonable to use the space-domain attenuation model instead. When $y < 1$, the present modified Szabo's model is the model of choice since its numerical solution is relatively easier than the fractional Laplacian space-domain model.

ACKNOWLEDGMENTS

The work reported here is sponsored by Simula Research Laboratory with the project ''Mathematical and numerical modeling of medical ultrasound wave propagation.''

APPENDIX: FRACTIONAL TIME DERIVATIVES

The Riemann–Liouville fractional integral is an essential concept to understand the fractional derivative and is given by (Samko *et al.*, 1987)

$$J^q\{\psi(t)\} = \frac{1}{\Gamma(q)}\int_a^t \frac{\psi(\tau)}{(t-\tau)^{1-q}}\,d\tau, \quad 0 < q, \tag{A1}$$

where q and a are real valued. The corresponding Riemann–Liouville fractional derivative is interpreted as

$$D_*^\lambda\{\psi(t)\} = \frac{d}{dt}[J^{1-\lambda}\{\psi(t)\}], \quad 0 < \lambda < 1, \tag{A2}$$

which can be further elaborated into

$$D_*^\lambda\{\psi(t)\} = \frac{1}{\Gamma(1-\lambda)}\frac{d}{dt}\int_a^t \frac{\psi(\tau)}{(t-\tau)^\lambda}\,d\tau$$

$$= \frac{1}{\Gamma(-\lambda)}\int_a^t \frac{\psi(\tau)}{(t-\tau)^{1+\lambda}}\,d\tau = J^{-\lambda}\{\psi(t)\}. \tag{A3}$$

The Riemann–Liouville fractional derivative, however, has some notable disadvantages in engineering applications such as the hypersingular improper integral, where the order of singularity is higher than the dimension, and nonzero of the fractional derivative of constants, e.g., $D^\lambda 1 \neq 0$, which would entail that dissipation does not vanish for a system in equilibrium (Samko *et al.*, 1987; Seredynska and Hanyga, 2000) and invalidates the causality. The Caputo fractional derivative has instead been developed to overcome these drawbacks (Caputo, 1967; Caputo and Mainardi, 1971) as defined below:

$$D^\lambda\{\psi(x)\} = J^{1-\lambda}\left[\frac{d}{dx}\psi(x)\right]. \tag{A4}$$

Integration by parts of (A4) yields

$$D^\lambda\{\psi(t)\} = \frac{1}{\Gamma(1-\lambda)}\int_a^t \frac{1}{(t-\tau)^\lambda}\frac{d\psi(\tau)}{d\tau}\,d\tau$$

$$= D_*^\lambda\{\psi(t)\} - \frac{\psi(a)}{\Gamma(1-\lambda)}\frac{1}{(t-a)^\lambda}. \tag{A5}$$

It is observed that the right-hand second term in (A5) regularizes the fractional derivative to avoid the potential divergence from singular integration at $t=a$. In addition, the Caputo fractional differentiation of a constant results in zero. For instance, Seredynska and Hanyga (2000) pointed out that the solution of the following time fractional derivative model:

$$D^2 u + \gamma D^{1+\eta} u + F(u) = 0, \quad 0 < \eta \leqslant 2, \tag{A6}$$

is not C^2 smooth if the Caputo fractional derivative is replaced by the Riemann–Liouville fractional derivative. The Caputo fractional derivative also implicitly includes the initial function value at $t=a$ as shown in (A5), which is convenient in handling initial value problem. Therefore, the fractional derivative defined in the Caputo sense is essential in the modeling of various anomalous attenuation behaviors (Makris and Constantinou, 1991; Seredynska and Hanyga, 2000). The numerical solution of the Caputo derivative equation like (A6) is well developed, as described in the monograph by Podlubny (1999).

The fractional derivatives in the Riemann–Liouville and Caputo senses can, respectively, in general be expressed as

$$D_*^{\mu}\{\psi(x)\} = D^m[J^{m-\mu}\psi(x)], \tag{A7}$$

$$D^{\mu}\{\psi(x)\} = J^{m-\mu}[D^m\psi(x)]. \quad m-1 < \mu \leqslant m, \tag{A8}$$

where m is an integer, and $D^m\psi = d^m\psi/dt^m$. The Riemann–Liouville fractional derivative (A7) can be recast as (Seredynska and Hanyga, 2000)

$$D_*^{\mu}\psi = \theta_{-\mu-1} * \psi, \tag{A9}$$

where

$$\theta_{\mu}(t) = |t|^u/\Gamma(\mu+1). \tag{A10}$$

By integration by parts, the Caputo fractional derivative (A8) can be reduced to

$$D^{\mu}\psi = \theta_{-\mu-1} * \psi - \sum_{k=0}^{m-1} D^k\psi(0)\,\theta_{k-\mu}$$

$$= D_*^{\mu}\psi - \sum_{k=0}^{m-1} D^k\psi(0)\,\theta_{k-\mu}, \tag{A11}$$

where the first term of the right-hand side is in fact the Riemann–Liouville fractional derivative.

Adhikari, S. (**2000**). "Damping Models for Structural Vibration," Ph.D. thesis, Cambridge University.

Bagley, R. L., and Torvik, P. J. (**1983**). "A theoretical basis for the application of fractional calculus to viscoelasticity," J. Rheol. **27**, 201–210.

Blackstock, D. T. (**1967**). "Transient solution for sound radiated into a viscous fluid," J. Acoust. Soc. Am. **41**, 1312–1319.

Caputo, M. (**1967**). "Linear models of dissipation whose *Q* is almost frequency independent. II," Geophys. J. R. Astron. Soc. **13**, 529–539.

Caputo, M., and Mainardi, F. (**1971**). "A new dissipation model based on memory mechanism," Pure Appl. Geophys. **91**, 134–147.

Chen, W., and Holm, S. (**2002**). "Fractional Laplacian time-space models and Lévy stable distribution for linear and nonlinear frequency-dependent lossy media," Research Report of Simula Research Laboratory, No. 2002-09.

Diethelm, K. (**1997**). "An algorithm for the numerical solution of differential equations of fractional order," Electronic Trans. Numer. Anal. **5**, 1–6.

Diethelm, K. (**2000**).

Diethelm, K., and Ford, N. J. (**2002**). "Analysis of fractional differential equations," J. Math. Anal. Appl. **265**, 229–248.

Enelund, M. (**1996**). "Fractional Calculus and Linear Viscoelasticity in Structural Dynamics," Ph.D thesis, Chalmers University of Technology, Sweden.

Gaul, L. (**1999**). "The influence of damping on waves and vibrations," Mech. Syst. Signal Process. **13**(1), 1–30.

Ginter, S. (**2000**). "Numerical simulation of ultrasound-thermotherapy combining nonlinear wave propagation with broadband soft-tissue absorption," Ultrasonics **27**, 693–696.

Hanyga, A. (**1999**). "Simple memory models of attenuation in complex viscoporous media," in Proceedings of the 1st Canadian Conference on Nonlinear Solid Mechanics, Victoria, BC, 16–20 June.

Hanyga, A. (**2001**). "Multi-dimensional solutions of space-fractional diffusion equations," Proc. R. Soc. London, Ser. A **457**, 2993–3005.

He, P. (**1998**). "Simulation of ultrasound pulse propagation in lossy media obeying a frequency power law," IEEE Trans. Ultra. Ferro. Freq. Contr. **45**(1), 114–125.

Johnson, D., and Dudgeon, D. (**1993**). *Array Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Lighthill, M. J. (**1962**). *Introduction to Fourier Analysis and Generalized Functions* (Cambridge University Press, Cambridge).

Makris, N., and Constantinou, M. C. (**1991**). "Fractional-derivative Maxwell model for viscous dampers," J. Struct. Eng. **117**(9), 2708–2724.

Mast, T. D., Souriau, L. P., Liu, D. L., Tabei, M., Nachman, A. I., and Wagg, R. C. (**2001**). "A *k*-space method for large-scale models of wave propagation in tissue," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**(2), 341–354.

Matignon, D., Audounet, J., and Montseny, G. (**1998**). "Energy decay for wave equations with damping of fractional order," LAAS Report 98031.

Nachman, A. I., Smith, J., and Waag, R. C. (**1990**). "An equation for acoustic propagation in inhomogeneous media with relaxation losses," J. Acoust. Soc. Am. **88**(3), 1584–1595.

Ochmann, M., and Makarov, S. (**1993**). "Representation of the absorption of nonlinear waves by fractional derivative," J. Acoust. Soc. Am. **94**(6), 3392–3399.

Pierce, A. D. (**1989**). *Acoustics, an Introduction to its Physical Principles and Applications* (Acoustical Society of America, Woodbury, NY).

Podlubny, I. (**1999**). *Fractional Differential Equations* (Academic, New York).

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (**1992**). *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge), Chaps. 18.2 and 18.3, pp. 786–794, Volterra Equations.

Rossikhin, Y. A., and Shitikova, M. V. (**1997**). "Applications of fractional calculus to dynamic problems of linear and nonlinear hereditary mechanics of solids," ASME Appl. Mech. Rev. **50**, 15–67.

Samko, S. G., Kilbas, A. A., and Marichev, O. I. (**1987**). *Fractional Integrals and Derivatives: Theory and Applications* (Gordon and Breach Science, New York).

Seredynska, M., and Hanyga, A. (**2000**). "Nonlinear Hamiltonian equations with fractional damping," J. Math. Phys. **41**, 2135–2156.

Szabo, T. L. (**1993**). "Time Domain Nonlinear Wave Equations for Lossy Media," in *Advances in Nonlinear Acoustics: Proceedings of the 13th ISNA*, edited by H. Hobaek (World Scientific, Singapore), pp. 89–94.

Szabo, T. L. (**1994a**). "Time domain wave equations for lossy media obeying a frequency power law," J. Acoust. Soc. Am. **96**(1), 491–500.

Szabo, T. L. (**1994b**). "Causal theories and data for acoustic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **97**(1), 14–24.

Szabo, T. L., and Wu, J. (**2000**). "A model for longitudinal and shear wave propagation in viscoelastic media," J. Acoust. Soc. Am. **107**(5), 2437–2446.

Wismer, M., and Ludwig, R. (**1995**). "An explicit numerical time domain formulation to simulate pulsed pressure waves in viscous fluids exhibiting arbitrary frequency power law attenuation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**(6), 1040–1049.

Wojcik, G., Mould, J., Lizzi, Jr., F., Abboud, N., Ostromogilsky, M., and Vaughan, D. (**1995**). "Nonlinear modelling of therapeutic ultrasound," 1995 IEEE Ultrasonics Symposium Proceedings, pp. 1617–1622.

Yuan, X., Borup, D., and Wiskin, J. (**1999**). "Simulation of acoustic wave propagation in dispersive media with relaxation losses by using FDTD method with PML absorbing boundary condition," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**(1), 14–23.

Locally conformal method for acoustic finite-difference time-domain modeling of rigid surfaces[a)]

Julius G. Tolan[b)] and John B. Schneider[c)]
School of Electrical Engineering and Computer Science, Washington State University, Pullman, Washington 99164-2752

(Received 15 January 2001; accepted for publication 11 August 2003)

The finite-difference time-domain method is a simple but powerful numerical method for simulating full-wave acoustic propagation and scattering. However, the method can demand a large amount of computational resources. Traditionally, continuously curved boundaries are represented in a stair-step fashion and thus accurately modeling scattering from a boundary will require a finer discretization than would otherwise be necessary for modeling propagation in a homogeneous medium. However, a fine discretization might not be practical due to limited computational resources. A locally conformal technique is presented here for modeling acoustic scattering from continuously curved rigid boundaries. This technique is low cost, simple to implement, and gives better results for the same grid discretization than the traditional stair-step representation. These improvements can be traded for a coarser discretization which reduces the computational burden. The improved accuracy of this technique is demonstrated for a spherical scatterer. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1616576]

PACS numbers: 43.20.Fn, 43.30.Ft, 43.20.Px [VWS] Pages: 2575–2581

I. INTRODUCTION

The finite-difference finite-time (FDTD) method was introduced by Yee in 1966 for the purpose of modeling electromagnetic wave propagation.[1] The method is simple to implement and can readily be applied to other types of wave propagation phenomena, e.g., Refs. 2 and 3.

Modeling scattering from continuously curved boundaries with the traditional FDTD method requires a stair-step representation of those boundaries because the field update equations use material properties defined only at discrete spatial locations. When performing such modeling, using a finer discretization, i.e., reducing the spatial step size, will yield more accurate results; however, refining the grid increases the computational resource requirements. For example, when modeling a three-dimensional problem, if the grid step size is reduced by one-half, the required computer memory increases by a factor of 8 and the time required to solve the problem increases by a factor of 16. Due to limited computational resources, finer grid discretizations might not be practical and instead a coarser grid will have to be used and the resulting lower accuracy tolerated.

In this paper we introduce a locally conformal method for modeling rigid boundaries in acoustic FDTD simulations. Rigid boundaries have been used in various models including that of the human head.[4,5] An FDTD model of a rigid human head has been reported[6] where a stair-step representation was used.

The algorithm presented here uses special update equations for the pressure nodes adjacent to rigid boundaries while using regular update equations for all other nodes (including velocity nodes). The algorithm can be implemented with minimum effort, and it is found that the solution accuracy is improved significantly. Locally conformal methods for pressure-release surfaces have also been developed and can be found in the literature.[7] The locally conformal pressure-release schemes are only pertinent to the case of Dirichlet boundary conditions. Although they are similar in concept to the approach described here, they do not apply to the Neumann boundary conditions which govern rigid objects.

An alternative approach for better representing curved boundaries or boundaries not aligned with a Cartesian grid is a globally conformal technique such as presented by Botteldooren.[8] The construction of the grid in a globally conformal technique can be quite complicated. Furthermore, globally conformal techniques may not be able to use directly standard tools that have been developed for Cartesian grids (such as absorbing boundary conditions, near-field to far-field transformations, etc.). In the technique we propose this is not a concern since the technique does not distort the grid away from the surface of the scatterer. Finally, globally distorted grids have inherent numeric artifacts that are difficult to quantify *a priori* (e.g., a plane wave that propagates through a globally distorted grid representing a homogeneous medium typically will not emerge as a plane wave).

II. CONFORMAL TECHNIQUE

We begin our development with the governing differential equations in three dimensions,

$$\frac{\partial p}{\partial t} = -\rho c^2 \nabla \cdot \mathbf{v} = -\rho c^2 \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \right), \tag{1}$$

[a)]Presented in part at the 140th Meeting of the Acoustical Society of America.
[b)]Present address: Northrop Grumman Space Technology, 15051 Avenue of Sciences, San Diego, CA 92128.
[c)]Author to whom correspondence should be addressed; electronic mail: schneidj@eecs.wsu.edu

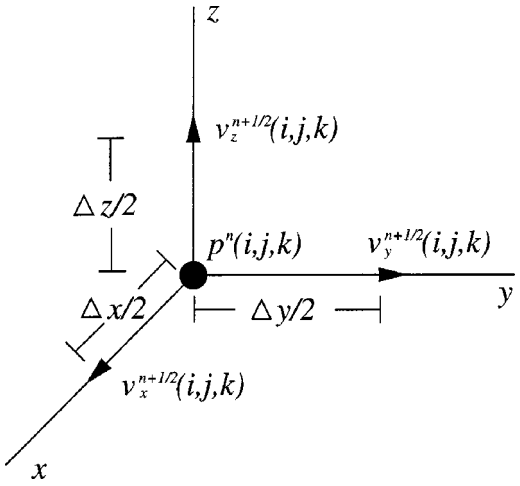


FIG. 1. Position of pressure and velocity nodes in a Cartesian grid. Different fields with the same spatial indices (i, j, k) are not collocated. For example, $v_x(i,j,k)$ and $p(i,j,k)$ are offset by a half spatial step in the x direction.

$$\frac{\partial \mathbf{v}}{\partial t} = -\frac{1}{\rho}\nabla p \rightarrow \begin{cases} \dfrac{\partial v_x}{\partial t} = -\dfrac{1}{\rho}\dfrac{\partial p}{\partial x}, \\ \dfrac{\partial v_y}{\partial t} = -\dfrac{1}{\rho}\dfrac{\partial p}{\partial y}, \\ \dfrac{\partial v_z}{\partial t} = -\dfrac{1}{\rho}\dfrac{\partial p}{\partial z}, \end{cases} \tag{2}$$

where p is the pressure, $\mathbf{v} = v_x\hat{\mathbf{a}}_x + v_y\hat{\mathbf{a}}_y + v_z\hat{\mathbf{a}}_z$ the velocity, ρ the density, and c the speed of sound. The density and speed may both be inhomogeneous.

To solve a problem using the FDTD method, we begin by replacing the derivatives with finite differences. Figure 1 shows a Cartesian grid arrangement of nodes that can be used for this replacement. The velocity and pressure nodes are known only at fixed locations in space and only at discrete times. The spatial location of the fields is indicated by the indices (i,j,k) such that $f(i,j,k)$ implies the physical location $(i\Delta x, j\Delta y, k\Delta z)$ where f is one of the fields and Δx, Δy, and Δz, are the step sizes in the x, y, and z directions, respectively. Furthermore, there is an additional offset of the fields as indicated in Fig. 1 so that each of the velocity components is offset a half spatial step from the pressure node with the same indices. (One could explicitly include these offsets in the indices, but they are dropped when implementing the algorithm in a computer. For notational convenience and to facilitate translation into code, we leave the offsets as being implied by the field component itself.) The temporal index is n. Velocity nodes are offset from pressure nodes by a half temporal step, but all velocity components are evaluated at the same instant of time, i.e., pressure nodes are defined to exist at integer multiples of the temporal step Δt while the velocity nodes v_x, v_y, and v_z, exist at times halfway between the times at which the pressures are defined. In contrast to the spatial indices, we will indicate explicitly the offset in temporal locations of the fields.

This arrangement of offset nodes allows the calculation of pressures from velocities at a previous time, and the calculation of velocities from these updated pressures. This alternating calculation is repeated for each successive time step

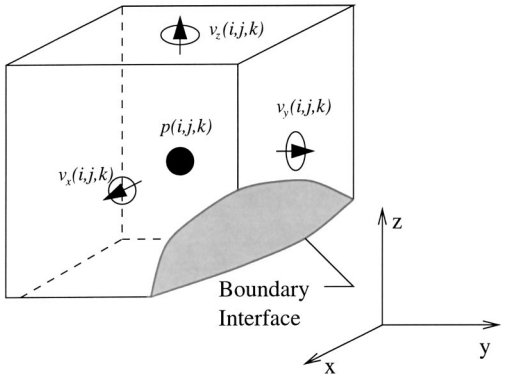


FIG. 2. Cell intersected by a rigid boundary with the portion inside the boundary removed.

as the simulation is run from an initial time to a final time. By replacing the derivatives with finite differences according to the node arrangement shown in Fig. 1, the following (traditional) update equations can be obtained from the governing differential equations (1) and (2):

$$\begin{aligned} p^{n+1}(i,j,k) = p^n(i,j,k) \\ - \Delta t \rho c^2 \Bigg(& \frac{[v_x^{n+1/2}(i,j,k) - v_x^{n+1/2}(i-1,j,k)]}{\Delta x} \\ & + \frac{[v_y^{n+1/2}(i,j,k) - v_y^{n+1/2}(i,j-1,k)]}{\Delta y} \\ & + \frac{[v_z^{n+1/2}(i,j,k) - v_z^{n+1/2}(i,j,k-1)]}{\Delta z} \Bigg), \end{aligned} \tag{3}$$

$$v_x^{n+1/2}(i,j,k) = v_x^{n-1/2}(i,j,k) - \frac{\Delta t}{\rho}\left(\frac{p^n(i+1,j,k) - p^n(i,j,k)}{\Delta x}\right), \tag{4}$$

$$v_y^{n+1/2}(i,j,k) = v_y^{n-1/2}(i,j,k) - \frac{\Delta t}{\rho}\left(\frac{p^n(i,j+1,k) - p^n(i,j,k)}{\Delta y}\right), \tag{5}$$

$$v_z^{n+1/2}(i,j,k) = v_z^{n-1/2}(i,j,k) - \frac{\Delta t}{\rho}\left(\frac{p^n(i,j,k+1) - p^n(i,j,k)}{\Delta z}\right), \tag{6}$$

where the superscripts represent the temporal index, and the arguments the spatial indices. These update equations require that material properties change only at discrete locations. This implies that continuous boundaries have to be represented in a stair-step fashion.

The conformal technique presented here is similar to the one presented by Dey and Mittra for modeling perfect conductors in three-dimensional electromagnetic simulations.[9] We begin by integrating Eq. (1) over a volume V, which is bounded by a surface S, and then use the divergence theorem to obtain:

$$\frac{d}{dt}\int_V p\, dv = -\rho c^2 \int_V \nabla \cdot \mathbf{v}\, dv = -\rho c \oint_S \mathbf{v} \cdot d\mathbf{s}. \tag{7}$$

The integration can now be applied to a unit cell in the FDTD grid, centered on the pressure node at (i,j,k), which is intersected by an arbitrary rigid boundary, as shown in Fig. 2. In applying the integration, note that the normal component of velocity is zero over the rigid boundary, so the surface integral over that area is zero. Velocities are assumed constant over the portion of each face that lies outside the rigid boundary and the pressure is assumed constant throughout the cell volume which is outside the boundary. Carrying out the integration with these assumptions yields

$$\begin{aligned}\frac{dp(i,j,k)}{dt}V(i,j,k) = -\rho c^2[&A_x(i,j,k)v_x(i,j,k)\\ &-A_x(i-1,j,k)v_x(i-1,j,k)\\ &+A_y(i,j,k)v_y(i,j,k)\\ &-A_y(i,j-1,k)v_y(i,j-1,k)\\ &+A_z(i,j,k)v_z(i,j,k)\\ &-A_z(i,j,k-1)v_z(i,j,k-1)], \quad (8)\end{aligned}$$

where $V(i,j,k)$ is the cell volume outside the rigid boundary and $A_m(i,j,k)$ is the area of a cell face outside the rigid boundary, where m represents one of the surface normal directions x, y, or z. Approximating the temporal derivative with a finite difference, the following pressure update equation is obtained

$$\begin{aligned}p^{n+1}(i,j,k) = p^n(i,j,k) - \frac{\rho c^2 \Delta t}{V(i,j,k)}[&A_x(i,j,k)v_x^{n+1/2}(i,j,k)\\ &-A_x(i-1,j,k)v_x^{n+1/2}(i-1,j,k)\\ &+A_y(i,j,k)v_y^{n+1/2}(i,j,k)\\ &-A_y(i,j-1,k)v_y^{n+1/2}(i,j-1,k)\\ &+A_z(i,j,k)v_z^{n+1/2}(i,j,k)\\ &-A_z(i,j,k-1)v_z^{n+1/2}(i,j,k-1)]. \quad (9)\end{aligned}$$

This equation is used to update only the pressure in unit cells that are intersected by a rigid boundary. The velocity nodes adjacent to these pressure nodes are updated with Eqs. (4)–(6) since no special treatment is required. Furthermore, all other fields outside the rigid boundary are updated with the usual update equations Eqs. (3)–(6). If Eq. (9) is applied to a unit cell that is totally outside the rigid boundary, it can easily be shown that Eq. (9) reduces to Eq. (3).

When implementing Eq. (9) the volume and areas should be calculated before time-stepping begins. To calculate these quantities, we first need to identify all cells intersected by rigid boundaries. This can be accomplished by checking all the corners of a particular cell to see if at least one corner lies on the opposite side of a rigid boundary from the others. To increase further the processing speed, cells that are far enough from boundaries that they do not run the risk of intersecting them can be screened out with a simpler criterion based only on the location of the cell center. The more precise corner checking test can then be applied to the remaining cells.

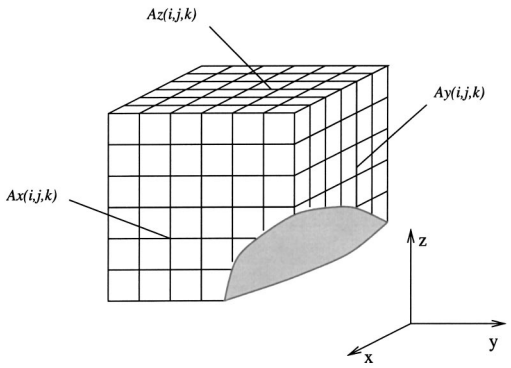


FIG. 3. A cell intersected by a rigid boundary which has been subdivided to facilitate volume and area calculations.

Once a cell is identified as being intersected by a boundary, its volume and areas can be calculated using a simple summation process. This process begins by subdividing a cell into smaller subcells, as illustrated in Fig. 3. The next step is to count all subcells that have their center outside the boundary. The total number is then used to determine the volume $V(i,j,k)$. Similarly, to find the face areas $A_m(i,j,k)$, the centers of all the squares on a particular cell face which result from the subdivision of the cell can be checked to see if they lie outside the rigid boundary. Their total number is then used to determine the areas $A_m(i,j,k)$. Note that the subdivision of cells is only done in the preprocessing phase to facilitate calculation of the geometric parameters needed in Eq. (9). The subdivision scheme described here merely requires a routine that can determine if a point is inside or outside the rigid boundary. If more detailed geometric information were needed, a different technique, such as analytic integration over the specified boundary, could be used to determine the geometric parameters. However, given the simplicity of the subdivision scheme and given a sufficiently fine level of subdivisions, it is anticipated that the subdivision approach will be preferable to the alternatives.

Stability of an FDTD simulation is another consideration that needs to be addressed. For propagation in a Cartesian grid, stability is ensured if

$$\Delta t \leqslant \frac{1}{c\sqrt{\frac{1}{\Delta x^2}+\frac{1}{\Delta y^2}+\frac{1}{\Delta z^2}}}. \quad (10)$$

For a uniform grid this reduces to, $c\Delta t/h \leqslant 1/\sqrt{3}$, where $\Delta x = \Delta y = \Delta z = h$. The dimensionless quantity $c\Delta t/h$ is the Courant number. However, when the conformal algorithm is implemented with a Courant number equal to the $1/\sqrt{3}$ limit, instability will occur for cells where the volume $V(i,j,k)$ is less then a certain minimum value $V_{\min}$ which was found empirically to be about 2% of the total volume of an undistorted cell. The instability can be removed if the cells with volume $V(i,j,k)$ less then $V_{\min}$ are treated as though they are completely inside the boundary, i.e., one resorts to stair stepping for these cells. For a Courant number of half the limit,

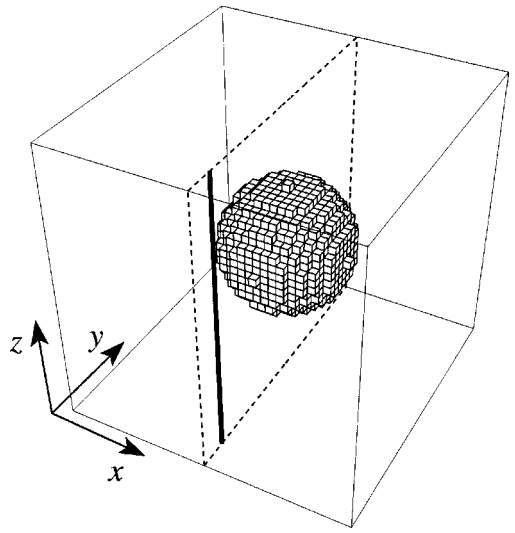


FIG. 4. Representation of the 3D computational domain used to obtain the scattered pressure from a rigid sphere. The sphere is depicted in accordance with its stair-step representation. The incident plane wave travels in the $+z$ direction and the scattered pressure is collected along the position of the heavy vertical line.

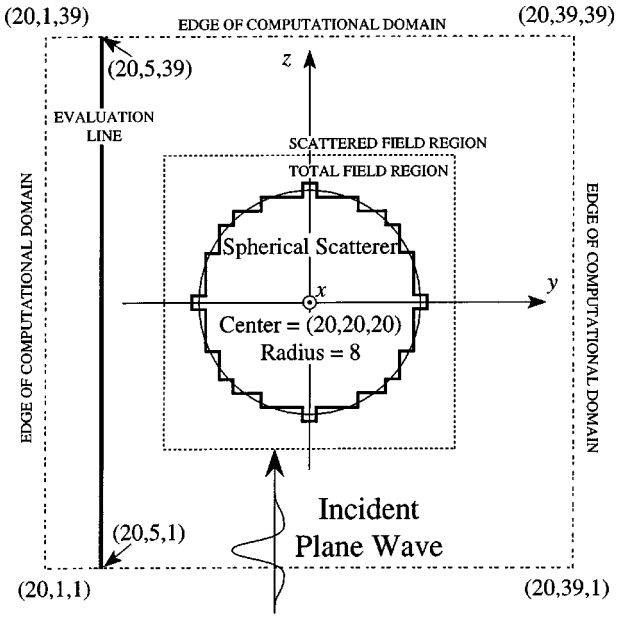


FIG. 5. Cross section of the computational domain at $x=20$. Scattered pressures are collected along the position of the heavy line.

$V_{\min}$ decreases to about 0.05% of the total volume of an undistorted cell. This yields a better representation of the boundary since one resorts to stair stepping only when 99.95% of a cell is already inside the rigid boundary; hence, only a slight error is introduced by assuming the remaining 0.05% of the cell is also within the rigid boundary.

The cost of lowering the Courant number is an increase in numeric dispersion and an increase in the number of time steps needed to propagate the field over a given distance. The increase in dispersion typically represents a small error and is more than offset with the improvement gained by the better representation of the rigid boundary. The increase in run time is inversely proportional to the reduction of the Courant number. Note that there is typically only a negligible increase in the amount of memory needed to implement the conformal technique. Hence if a computer is capable of running a simulation using a traditional staircase representation, that same simulation can be run at the same level of discretization on the same computer using the conformal technique. Reducing the Courant number by 50% and using the same level of discretization doubles the run time for the conformal technique. However, as shown in the next section, because of its increased accuracy, the conformal technique allows one to trade discretization (and accuracy) for run time. For example, if the spatial step size in the conformal technique were doubled, the simulation would run eight times faster than the traditional solution, and yet the conformal technique may still yield more accurate results.

III. RESULTS

The rigid-boundary conformal algorithm has been tested on a rigid sphere which is depicted as a staircased boundary in Fig. 4. This geometry was chosen since it permits comparison with an analytic solution. The sphere was illuminated with an incident plane wave traveling in the $+z$ direction and the scattered pressure collected over the heavy vertical line. The wave form of the incident field was chosen to be a Ricker wavelet with a peak spectral content corresponding to 20 cells per wavelength. The sphere has a radius of eight cells. The exact location of the sphere and the scattered pressure observation locations are shown in Fig. 5, which is a *yz* slice of the computational domain through the center of the sphere.

The computational domain had $39\times39\times39$ cells with an eight-cell wide perfectly matched layer (PML) surrounding it. The purpose of the PML, which is not shown in Fig. 5, is to prevent reflections from the termination of the FDTD grid.[3,10,11] A total-field/scattered-field formulation[12] was used to introduce the plane wave excitation. The simulation was run with a Courant number 50% of the limit and with a $V_{\min}$ value equal to 0.05% of the total volume of an undistorted cell. The simulation was run for 512 time steps and the scattered pressures at three different frequencies were obtained by use of a discrete Fourier transform (DFT).[12]

The FDTD results are compared with the exact analytic solution for pressures scattered from a rigid sphere of radius *a*, given by[13]

$$p(r,\theta)=\sum_{m=1}^{\infty} A_m P_m(\cos\theta) h_m(kr), \tag{11}$$

where P_m is the Legendre polynomial, h_m is the spherical Hankel function, r is the radial distance from the center of the sphere to the observation point, k is the wave number, and θ is the angle between the incident wave vector and the observation vector. The coefficients A_m are given by

$$A_m=-(2m+1)(i^m)\frac{j_m'(ka)}{h_m'(ka)}, \tag{12}$$

where the prime indicates differentiation with respect to the argument.

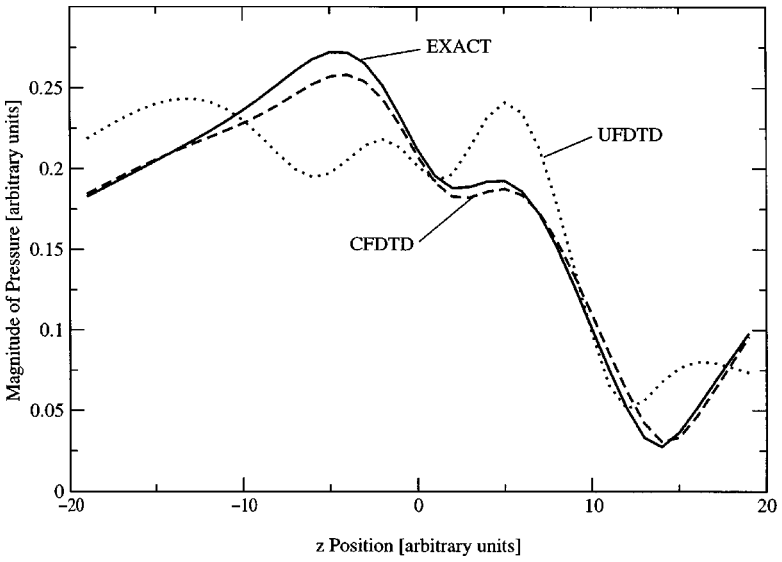


FIG. 6. Scattered pressure at 9.853 (nominally 10) points per wavelength. The independent variable is position in terms of cells where zero corresponds to the same z location as the center of the sphere.

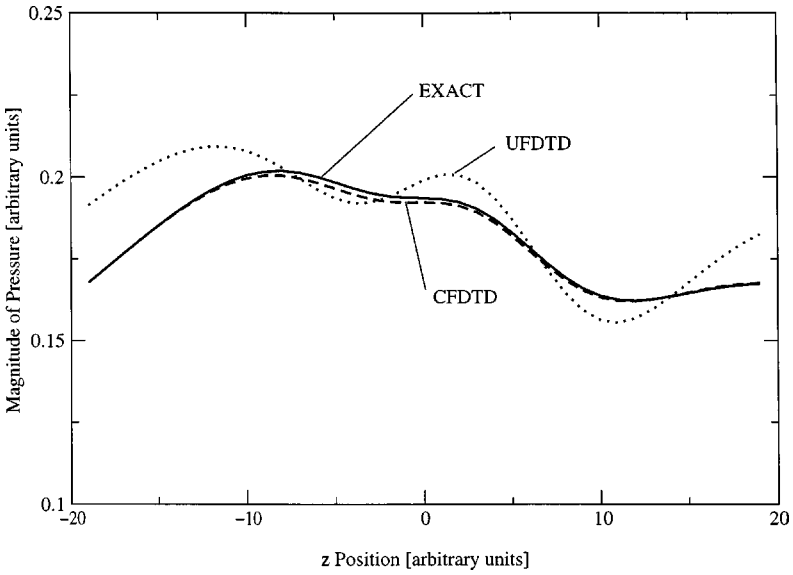


FIG. 7. Scattered pressure at 21.11 (nominally 20) points per wavelength.

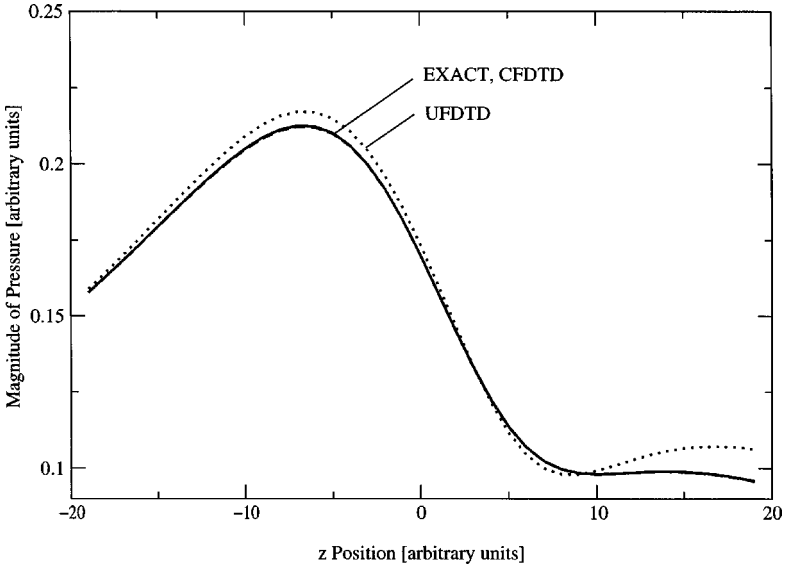


FIG. 8. Scattered pressure at 36.95 (nominally 40) points per wavelength.

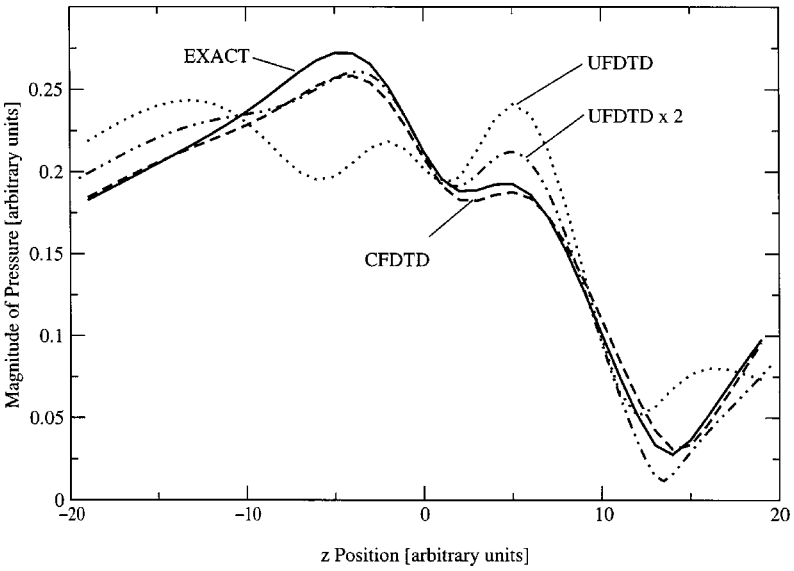


FIG. 9. Same frequency as in Fig. 6 corresponding to a *ka* of 5.026. The curve labeled "UFDTD×2" is obtained by using twice as many points per wavelength as were used to generate the previous results (i.e., by halving the spatial step size but maintaining the same geometry and *ka*).

Figures 6–8 show the scattered pressure for three different frequencies with wavelengths approximately equal to 10, 20, and 40 points per wavelength. (The actual frequencies that were obtained are dictated by the duration of the simulation and the nature of the discrete Fourier transform and correspond to 9.853, 21.11, and 36.95 points per wavelength. Nevertheless, for convenience we identify these discretizations by their approximate or "nominal" values.) Each plot shows the exact analytic result as well as the results obtained from the conformal FDTD (CFDTD) method and the traditional uniform FDTD (UFDTD) method (i.e., the stairstep approach). The pressures are normalized with respect to the incident wave. As shown in Fig. 6, at 10 points per wavelength the UFDTD method differs significantly from the exact result. The conformal method is much better but errors are still easily visible in the plot. As shown in Fig. 7, at 20 points per wavelength the UFDTD method still differs significantly from the exact solution while the CFDTD method is only slightly below the exact solution through the middle of the computational domain. As shown in Fig. 8, at 40 points per wavelength the exact and CFDTD results are nearly identical while the UFDTD result is still visibly different from the exact result. Note that these results pertain to any absolute frequency—the only important consideration is the size of the scatterer relative to a wavelength.

To illustrate further the superiority of the conformal technique, again consider the frequency and discretization shown in Fig. 6. At this frequency the scatterer has a *ka* of 5.026. We now construct a new uniform FDTD solution that halves the discretization used to obtain the previous results. Thus the sphere now has a radius of 16 cells and the computational domain is $78\times78\times78$ cells. Using this finer discretization we extract the results corresponding to a *ka* of 5.026. These results, labeled "UFDTD×2," are shown in Fig. 9 together with the previous results. By halving the discretization (and thus increasing the memory requirement by a factor of 8), the scattered field obtained using the UFDTD method has improved, but it is still not, on average, as good as the original CFDTD results. For this particular example the CFDTD method out-performs the UFDTD method even when the UFDTD method uses eight times the memory and runs eight times slower than the CFDTD method.

IV. CONCLUSION

The use of a stair-step representation of curved rigid boundaries in FDTD simulations can produce significant errors. The conformal method presented here requires modified pressure update equations only for cells adjacent to the boundary. The method is simple to implement and requires a negligible increase in memory relative to a traditional solution with the same discretization. The method does require a reduction in the Courant number to ensure stability and this can increase run-time. However, the increased accuracy of the conformal technique does offer the option of trading some accuracy (via use of a coarser discretization) for improved run time. This method was shown to produce improved results for scattered pressure from a rigid sphere and similar improvements can be expected for other geometries. Though results were not presented here, this method has been applied to other geometries which include both convex and concave surfaces.

ACKNOWLEDGMENTS

Funding for this work was provided by the Office of Naval Research, Code 321OA.

[1] K. S. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," IEEE Trans. Antennas Propag. **14**, 302–307 (1966).
[2] J. Virieux, "*SH*-wave propagation in heterogeneous media: Velocity-stress finite difference method," Geophysics **49**, 1933–1942 (1984).
[3] J. G. Maloney and K. E. Cummings. "Adaptation of FDTD techniques to acoustic modeling," 11th Annual Review of Progress in Applied Computational Electromagnetics, Monterey, CA, 1995, Vol. 2, pp. 724–731.
[4] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," J. Acoust. Soc. Am. **110**, 2440–2448 (2001).
[5] B. F. G. Katz, "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 (2001).
[6] T. Xiao and Q. H. Liu, "Finite difference computation of head-related

transfer function for human hearing,'' J. Acoust. Soc. Am. **113**, 2434–2445 (2003).
[7] J. B. Schneider, C. L. Wagner, and R. J. Kruhlak, ''Simple conformal methods for finite-difference time-domain modeling of pressure-release surfaces,'' J. Acoust. Soc. Am. **104**, 3219–3226 (1998).
[8] D. Botteldooren, ''Acoustical finite-difference time-domain simulation in a quasi-Cartesian grid,'' J. Acoust. Soc. Am. **95**, 2313–2319 (1994).
[9] S. Dey and R. Mittra, ''A locally conformal finite-difference time-domain (FDTD) algorithm for modeling three-dimensional perfectly conducting objects,'' IEEE Microw. Guid. Wave Lett. **7**, 273–275 (1997).
[10] X. Yuan, D. Borup, J. W. Wiskin, M. Berggren, R. Eidens, and S. A. Johnson, ''Formulation and validation of Berenger's PML absorbing boundary for the FDTD simulation of acoustic scattering,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 816–822 (1997).
[11] J.-P. Berenger, ''A perfectly matched layer for the absorption of electromagnetic waves,'' J. Comput. Phys. **114**, 185–200 (1994).
[12] A. Taflove and S. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 2nd ed. (Artech House, Boston, MA, 2000), pp. 356–359.
[13] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), pp. 1483–1486.

Ultrasonic wave interaction with multidirectional composites: Modeling and experiment

L. Wang and S. I. Rokhlin[a)]
The Ohio State University, Nondestructive Evaluation Program, Edison Joining Technology Center, 1248 Arthur E. Adams Drive, Columbus, Ohio 43221

(Received 10 March 2003; revised 4 August 2003; accepted 6 August 2003)

Phenomena of wave transmission through a multidirectional composite laminate immersed in a fluid have been investigated. Based on a recently-developed recursive stiffness matrix method time-domain beam models have been developed to simulate the problem. Experimental and theoretical results at frequency 2.25 MHz show that the transmission amplitude is highly dependent on lamina orientation and angle of incidence. Large transmission amplitude appears at small ($<10°$) and large incident angles ($45°-60°$). At intermediate incident angles ($16°-40°$) the transmission amplitude is almost zero. At high frequency, the residue epoxy layers between each lamina become important and corresponding resonances may be observed. These transmission phenomena have been interpreted in terms of Floquet waves. It shows that the pass and stop bands of the three Floquet waves obtained from the unit cell stiffness matrix determine the transmission amplitude distribution in frequency, incident angle and rotation angle domains. The effect of random deviation of the laminated structure periodicity has also been assessed. At normal incidence, the variation of the thickness of the epoxy residual layer between composite lamina has little effect on the overall stop and pass band structures but introduces random reverberation and scattering in the pass bands. It is shown that for oblique incidence the transmittivity spectrum of a composite with random lamina lay-up converges with increase of randomness to that of a $[0/-45/90/45]_{2s}$ quasi-isotropic composite. Randomization of lamina lay-up produces a small effect in the transmittivity spectrum for a quasi-isotropic composite. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1616920]

PACS numbers: 43.20.Gp, 43.20.Fn, 43.35.Cg [DEC] Pages: 2582–2595

I. INTRODUCTION

With the increasing use of advanced composites in a variety of modern applications, it has become a necessity to employ reliable and effective nondestructive evaluation (NDE) methods for their property characterization and damage assessment. Development of those methods depends on understanding wave propagation and scattering in these materials. Multidirectional composites are highly anisotropic multilayered structures, which significantly complicates ultrasonic wave propagation.[1–5] Selection of ultrasonic inspection parameters is very difficult for these structures without comprehensive modeling for optimization of the experimental conditions and data interpretation.

To model ultrasonic wave propagation in multidirectional composites, different solutions for wave propagation in layered anisotropic media have been utilized.[3–15] Methods based on the transfer matrix method have been extensively discussed for composites by Nayfeh[4,5] with experimental and theoretical examples for multidirectional composites given by Nayfeh and Chimenti[5] and Chimenti and Nayfeh.[8] A computationally stable global matrix method has been developed by Mal.[9] Different modifications of the global matrix method utilizing the stiffness matrix have been reported in Ref. 10 (this last method reduces by half the number of equations in the global system). Approximate solutions based on the finite element method have been discussed and reviewed by Datta[11] and Liu and Xi[12] (this method employs subdivisions of structure layers into a number of thin layers and leads to a large global system of linear equations). For a large number of layers the global system of equations becomes very large and requires special methods of solution. To avoid having to solve a large global matrix system Wang and Rokhlin[13,14] and Rokhlin and Wang[15] developed a stable reformulation of the transfer matrix method in the form of a recursive stiffness matrix method. A large number of experimental methods applicable to composite characterization have been reviewed by Chimenti.[16]

Advanced fiber-reinforced composites are usually assembled in different periodic stacking sequences of fiber directions (for example $0°/-45°/90°/45°$) to form multilayered structures as shown in Fig. 1. For the modeling of ultrasonic wave propagation, each layer is described as a monoclinic medium. In an anisotropic layered solid with periodicity, the wave propagation is significantly affected by the periodicity and depends on the behaviors of the Floquet waves propagating in an infinite periodic medium. Braga and Herrmann[17] addressed Floquet wave behavior in arbitrarily anisotropic infinite periodic media with a transfer matrix method using the Stroh formalism. They performed a very detailed analysis of the Floquet wave characteristic equation and studied stop and pass bands (Brillouin zones). Potel and Belleval and Potel *et al.*[18,19] applied the Floquet wave solution to finite periodic media and provided interesting numeri-

a)Author to whom correspondence should be addressed.

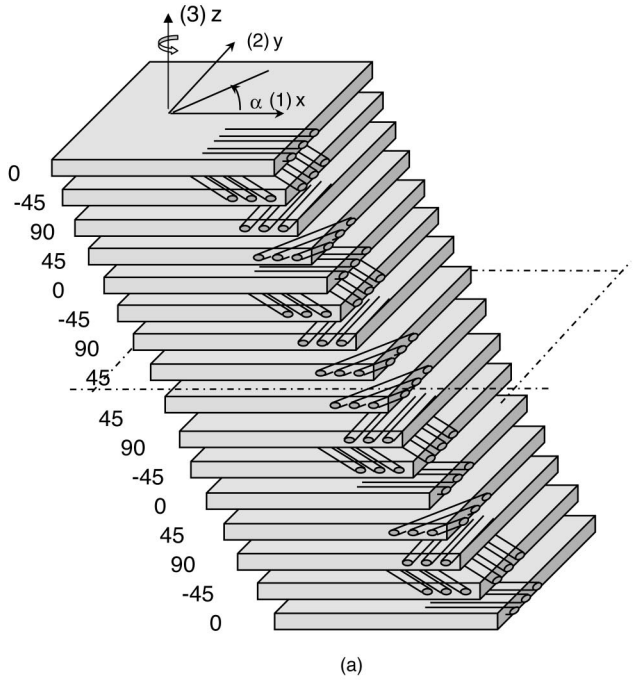


FIG. 1. (a) Lay-up of multidirectional $[0/-45/90/45]_{2s}$ composites, total thickness 3.1 mm and lamina thickness 0.194 mm. (b) Residual epoxy layer between each lamina, thickness 6μm.

cal illustrations. Shull *et al.* and Auld *et al.*[20,21] showed that the special Lamb wave dispersion behavior is related to the pass and stop bands of the Floquet wave spectrum in a plate formed by aluminum and aramid-epoxy composite layers. Based on the equivalency of the bulk waves in a homogeneous anisotropic medium and the Floquet waves in a periodic medium at low frequency, Wang and Rokhlin[22] described a Floquet wave-based dynamic homogenization method and used it to measure single lamina moduli in multi-directional composites.[23]

The reflection and transmission coefficients for a finite thickness composite structure are also related to the Lamb modes according to the "correspondence principle." [24] The validity of this principle depends on the density and the velocity in the fluid and the solid. The density of fiber-reinforced composites is very close to that of the immersion fluid (water), therefore the fluid loading effect becomes very important and reflection or transmission spectra are not necessarily closely associated with the Lamb wave dispersion spectra[24] in the unloaded plate.

Most practical ultrasonic inspections of multidirectional composites are performed in through-transmission mode using normally-incident ultrasonic waves. The use of oblique incident waves is complicated by strong lamina scattering and difficulties in the interpretation of the measurements. However application of obliquely-incident waves may in some cases be advantageous: (1) high sensitivity to inter-lamina fatigue damage; this damage appears in the laminas in the form of vertical cracks perpendicular to the lamina surface; (2) composite and lamina elastic property measurements;[23] (3) for bonded structure characterization.[25] Also, inspection of composite parts with curved shapes may require understanding and use of obliquely propagating waves.

In this paper, the reflection/transmission signatures in multidirectional composites are investigated comprehensively both theoretically and experimentally in the time and frequency domains. For modeling we use the recursive stiffness matrix method[13,14] and its asymptotic thin layer version.[15] The results are interpreted using Floquet waves. The effect of random deviation of the laminated structure periodicity has also been assessed. We limit ourselves to two-dimensional approximation of the ultrasonic beam, which satisfactorily represents the main features of the reflected and transmitted fields.[27] We use the recursive stiffness matrix method due to its numerical stability and its efficient utilization of the composite structure periodicity and symmetry, which makes the CPU time proportional to log (N)[14,15] instead of N or N^2 where N is the total number of laminas (or the number of subdivisions of the lamina into sublayers[15]).

II. TIME-DOMAIN-BEAM MODEL FOR REFLECTION/TRANSMISSION FROM/THROUGH MULTIDIRECTIONAL COMPOSITE STRUCTURES

A. Time-domain beam model

This section describes a model for the through transmission mode of composite testing. We are using the well-developed wave-number integral-transfer representations for the ultrasonic transducers. However we adapted it to the novel experimental configuration. In the double-transmission experiment outlined in Fig. 2(a) and Fig. 3, the receiver is replaced by a back reflector. The ultrasonic signal is transmitted through the sample, and then reflected by the plane reflector back to the transducer. In the model we consider the back reflector as an infinite mirror with a possible, for generality, misalignment by angle Δ from the vertical plane as shown in Fig. 2(a); the back propagation paths are redrawn by dotted lines in Fig. 2(b). Accounting for the phase delay in fluid $e^{-i2(k_z'L+k_x'L\tan\Delta)}$ as shown in Fig. 2(b), the time-domain voltage output is

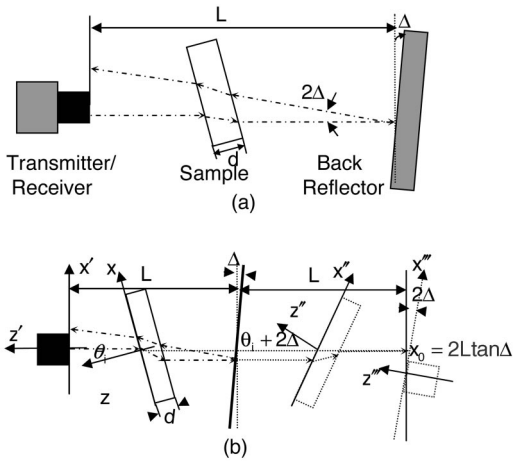


FIG. 2. Schematic diagram of the through-transmission method and coordinate systems. (a) Double-through transmission mode with a back reflector, (b) redraws the back propagation path in (a) considering the reflector as a mirror.

$$V_{out}(t)=\int_{-\infty}^{+\infty}F(\omega)e^{i\omega t}\,d\omega\int_{-\infty}^{+\infty}\Phi_t(k_x')\Phi_t(k_x''')T_t(k_x)\times T_b(k_x'')e^{-i2(k_z'L+k_x'L\tan\Delta)}\,dk_x', \tag{1}$$

where L is the distance between the surfaces of the transducer and back reflector. k_x', k_z' are the wave numbers in the coordinate systems associated with the surface of the transmitter (z' is the normal to the transmitter face and the origin is at its center), Δ is a reflector misalignment angle. k_x''', k_z''' are the wave numbers in the coordinate systems associated with the surface of the ''imaginary'' receiver (z''' is the normal to the transmitter face and the origin is at its center). The relation between wave numbers k_x, k_z, k_x'', k_z'', k_x''', k_z''' and k_x', k_z' can be found based on the relation of the coordinate systems (Fig. 2). For example, the $x'''-z'''$ coordinate system is rotated by the angle 2Δ from the $x'-z'$ coordinate system. In this case the wave numbers k_x''', k_z''' are related to k_x', k_z' as

$$k_x'''=k_x'\cos 2\Delta+k_z'\sin 2\Delta, k_z'''=k_z'\cos 2\Delta-k_x'\sin 2\Delta. \tag{2}$$

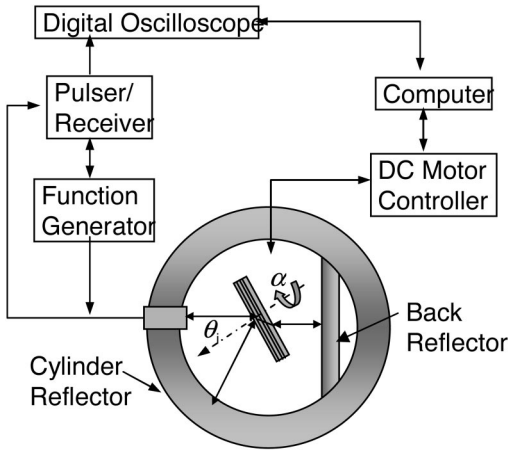


FIG. 3. The experimental setup of the ultrasonic measurement system.

In Eq. (1) $T_t(k_x)$ and $T_b(k_x'')$ are the transmission coefficients through the layer for the wave incident from fluid on the layer top or bottom respectively. $F(\omega)$ is the frequency response of the transmitter/receiver pair; $\Phi_t(k_x')$ and $\Phi_t(k_x''')$ are the angular responses of the transducer, which represent the directivity of the transducer. In our measurements, a flat immersion transducer has been used. The angular response of such a transducer may be approximated by a circular radiator using the first-order Bessel function; however the piezoceramic element in the transducer is bonded on its perimeter to the enclosure and therefore the displacement distribution on its surface is inhomogeneous and decays toward the edge. For this reason and for convenience of transducer response simulation (both analytical and numerical) the Gaussian beam has often been used.[28] As shown in Ref. 28, the output signals calculated using Bessel and Gaussian beams are nearly identical and agree with experiment when the appropriate beam width is used in the Gaussian beam approximation. In this study we have selected the angular responses $\Phi_t(k_x')$ of the flat immersion transducers to be approximated by a Gaussian function as

$$\Phi_t(k_x')=\frac{B_t}{\sqrt{2\pi}}e^{-(k_x'B_t)^2/2}, \tag{3}$$

where B_t is the beam width of the transmitter/receiver, which is experimentally determined by inverting the measured reflection signal from a homogeneous semispace such as a block of aluminum.

Similar to Eq. (1), one can formulate the time-domain angle beam models for reflection measurements.[25] It has been shown in Ref. 25 that the beam model based on the wave number integration and Gaussian beam approximation can well represent the experimental results for adhesive layer properties measurement in a bonded structure.

B. Reflection/transmission coefficient calculation using recursive stiffness matrix method

In Eq. (1), the plane wave transmission $T(\theta,\omega)$ coefficients are calculated exactly for the multilayered composite accounting for all multiple reflections between laminas. The reflection R and transmission T coefficients for an acoustic wave incident from fluid onto the immersed multidirectional layered composite are represented using the global compliance matrix as[13,14]

$$R=[(S_{11}^{33}-\Lambda)(S_{22}^{33}-\Lambda)-S_{21}^{33}S_{12}^{33}]/[(S_{11}^{33}+\Lambda)(S_{22}^{33}-\Lambda)-S_{21}^{33}S_{12}^{33}], \tag{4}$$

$$T=2\Lambda S_{21}^{33}/[(S_{11}^{33}+\Lambda)(S_{22}^{33}-\Lambda)-S_{21}^{33}S_{12}^{33}], \tag{5}$$

where S_{ij}^{33} are the (3,3) elements in the matrix $\mathbf{S}_{ij}(3\times3)$ of the total compliance matrix[14] and $\Lambda=\cos\theta_i/(i\omega\rho_f V_f)$, ρ_f is the fluid density, and V_f is the acoustic velocity in the fluid.

For layered semispaces, in some case it is advantageous to use the surface stiffness and compliance matrices.[6,13–15] The plane wave reflection coefficient R_{ss} for a fluid-loaded anisotropic solid semi-space is calculated by

$$R_{ss}=(S_{ss}^{33}-\Lambda)/(S_{ss}^{33}+\Lambda), \tag{6}$$

where S_{ss}^{33} is the (3,3) element in the 3×3 compliance matrix for a layered semispace or periodic semi-space.

For the fluid-loaded layered anisotropic structure, the dispersion equation for leaky Lamb modes is obtained by equating the denominator of the coefficients (4)–(6) to zero:

$$(S_{11}^{33}+\Lambda)(S_{22}^{33}-\Lambda)-S_{21}^{33}S_{12}^{33}=0;\quad S_{ss}^{33}+\Lambda=0. \tag{7}$$

For computation of the global compliance/stiffness matrix elements involved in the calculation of the reflection/transmission coefficients [Eqs. (4)–(6)] and solving the characteristic equations for guided waves we use the recently developed recursive stiffness matrix method (RSM),[13–15] which is unconditionally stable and has the efficiency of the transfer matrix method. Using the recursive stiffness matrix method, we obtain the total stiffness **K** and compliance **S** matrices for the *N*-layered anisotropic composite structure from the exact stiffness/compliance matrix of the laminas. Alternatively, the total stiffness and compliance matrices are calculated using the asymptotic recursive stiffness matrix method (Appendix A). This last method is simple and efficient and converges to the exact solution as the number of subdivisions of the thin layers increases (the method does not require finding the eigenvalues and eigenvectors of the Christoffel equation). To obtain the time-domain outputs [Eq. (1)], we perform numerical integration of plane wave reflection/transmission coefficients in the angle and frequency domains determined by the transducer bandwidth and angle response, which requires extensive computation. To optimize computational efficiency we take advantage of the composite's periodicity and symmetry[14] (Appendix B).

To facilitate discussion and interpretation of the experimental and simulated results, we will use in Sec. V the concept of Floquet waves. Simulation of their propagation in periodic anisotropic media by the stiffness matrix method has been discussed in Refs. 22 and 23. Combining the periodicity condition and cell stiffness matrix $\mathbf{K}^c$, one can obtain the characteristic equation for the Floquet wave in terms of the stiffness matrix as:

$$\det(e^{i\zeta h_c}\mathbf{K}_{21}^c-e^{-i\zeta h_c}\mathbf{K}_{12}^c+\mathbf{K}_{22}^c-\mathbf{K}_{11}^c)=0, \tag{8}$$

where ζ represents the Floquet wave number and h_c is the thickness of the periodic unit cell. In the wave number interval $-\pi<\zeta h_c<+\pi$, we can uniquely determine six Floquet wave numbers ζ.[22] Complex solutions represent nonhomogeneous Floquet waves and real solutions are propagating Floquet waves. The sign of the appropriate roots (wave numbers) for a given propagation direction are selected based on the energy propagation direction. The nonpropagating and propagating Floquet waves are described by stop and pass bands of the Floquet wave spectrum.

III. EXPERIMENTAL APPROACH

The experimental ultrasonic system is shown schematically in Fig. 3. It includes an ultrasonic goniometer allowing immersion double-through-reflection/transmission measurement.[26] In the goniometer, the sample can be rotated about two axes. The change of the incident θ_i and orientation angles α are digitally controlled, with measurements performed for different incident angles θ_i at each orientation angle α. The reflected or transmitted signals are amplified, digitized, averaged by a LeCroy 9400 digital oscilloscope, and collected by the computer through an IEEE-488 interface. The data can be further processed to collect frequency response and peak-to-peak amplitude information.

The sample used in the experiment is a multidirectional laminated graphite epoxy composite $[0/-45/90/45]_{2s}$ as shown in Fig. 1. The composite laminate with total thickness 3.1 mm has 16 laminas; the thickness of each lamina is 0.194 mm. As shown in Fig. 1(b), there is an excess of epoxy between the laminas forming thin bonding layers (6 μm). As we will discuss below the thin bond layer has a strong effect on the reflected signal at normal incidence.

Our objective is to investigate the reflection/transmission properties of the composite to optimize its inspection and characterization. The ultrasonic reflection and transmission characteristics are measured as functions of the incident angles θ_i and orientation angles α of the composite sample at different frequencies. α is defined as the angle between the plane of incidence and the fiber direction of the top lamina as shown in Fig. 1.

IV. RESULTS AND DISCUSSIONS

A. Reflection response at normal incidence

In the normal direction z, all laminas are assumed to have identical properties (elastic constant C_{33} and density); therefore, the first step is to approximate the medium as a homogeneous layer. However, for a graphite/epoxy laminate, there is a very thin residual interfacial epoxy layer between each lamina (which does not contain fibers) and therefore the properties of this thin layer are different from those of the lamina. The presence of this thin residual epoxy layer at each interface transforms the structure into a periodic layered medium in a normal direction. The unit cell of this periodic structure is formed by two layers: lamina and thin epoxy interfacial layer. Figures 4(a) and (b) show experimental and calculated time domain reflection responses from the $[0/-45/90/45]_{2s}$ composite sample at normal incidence using two broadband transducers at center frequencies 5 MHz and 10 MHz. For center frequency 5 MHz, one can clearly observe the reflection echoes from the front and back surfaces of the sample. The reflection response is very similar to that from an isotropic layer with two parallel surfaces. The very low amplitude signals (close to baseline level) between the multiple reflections indicate the scattering noise from the interfaces between layers; however the influence by thin excess epoxy layers between different laminas is small. For the 10 MHz transducer, one can still clearly see the multiple reflections; however there is resonance excitation showing as a continuous signal between the echoes caused by the signals reflected from each lamina boundary. The impedance mismatch between the thin excess epoxy layer and the lamina causes the reflection. In our theoretical model, we introduce a 6μm excess epoxy layer between each lamina. The model predicts the reverberations and agrees very well with experi-

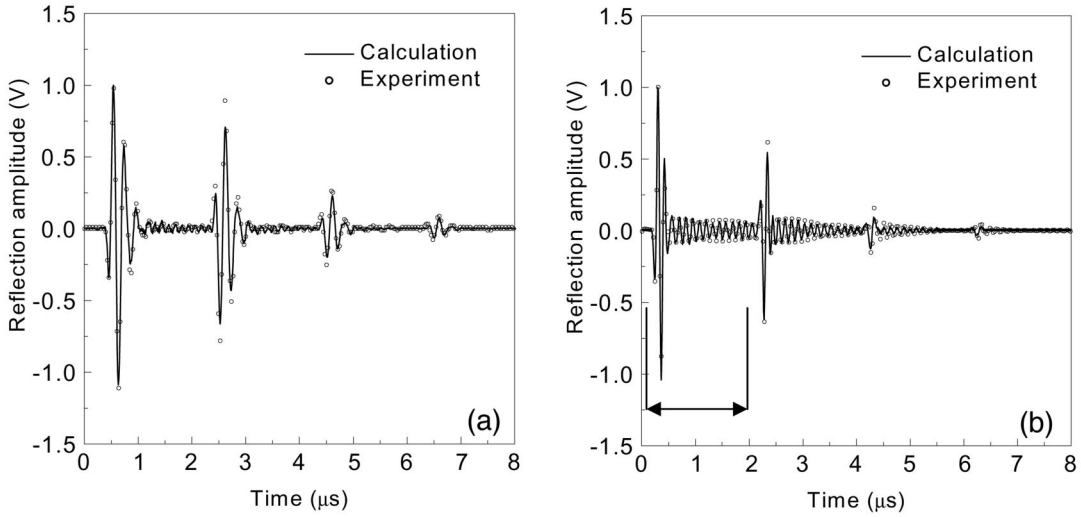


FIG. 4. Normal reflection time-domain signals from the multidirectional $[0/-45/90/45]_{2s}$ composites. (a) Ultrasonic pulse with center frequency 5 MHz, and (b) center frequency 10 MHz.

ment [Fig. 4(b)]. It is obvious that one can measure the lamina properties such as thickness from the reverberation signal.

When the wavelength λ is much larger than the cell thickness h_c ($\lambda > 2h_c$) the residual epoxy layer between laminas does not affect the reflection and transmission and the composite is acting as a homogeneous layer in the normal direction. When the wavelength becomes comparable to or smaller than the cell thickness ($\lambda < 2h_c$), due to the periodicity of the assembly of lamina and residual layer the laminate may be considered as a periodic medium. In this case, one may observe stop and pass bands. The theoretical stop and pass bands can be calculated using Eq. (8) for a wave propagating in the normal direction (z). In Fig. 5, the top curve (a) shows the calculated stop and pass bands for the longitudinal wave as function of frequency (bottom axis). The corresponding ratio of cell thickness and wavelength is shown on the top axis. One can see that stop bands appear when frequency is larger than 7.9 MHz ($\lambda > 2h_c$) and the width of the stop bands increases as frequency increases. Figure 5(b) shows the reflection spectrum for a wave incident from fluid onto a periodic semi-space calculated using Eq. (6). The solid line is calculated without attenuation, the dotted line calculated with attenuation ($C_{33} = 6.3 + 0.06i$ for

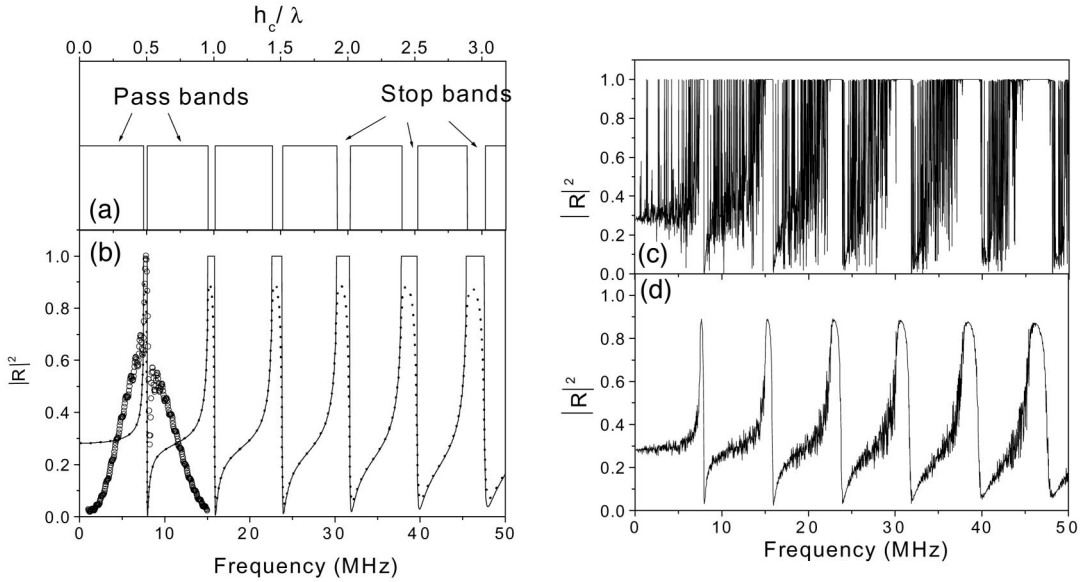


FIG. 5. (a) Floquet wave pass and stop bands at normal incidence for an infinite periodic medium. (b) Reflection coefficient for wave incident from fluid onto a semi-space periodic medium. The solid line is calculated without attenuation, the dotted line calculated with attenuation (C33 = 6.3 + 0.06i for epoxy layer). The open circles are the experimental spectrum of the gated signal shown in Fig. 4(b). The experimental spectrum is not deconvolved and includes the transducer response. The periodic medium is [0/−45/90/45] composite including a 6 μm epoxy layer at each interface. (c), (d) Calculated reflection spectrum for wave incident from fluid onto a semispace with random thickness of the residual epoxy layer between composite laminas, distributed uniformly between 5 and 7 μm. No attenuation in (c) and with attenuation in (d).

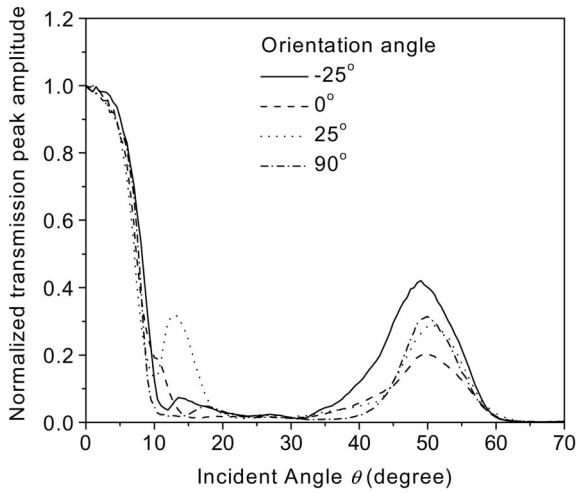


FIG. 6. Measured peak amplitude of double transmission signal versus incident angle α at different orientation angles α. (a) $\alpha=0°$, (b) $\alpha=25°$, (c) $\alpha=-25°$, (d) $\alpha=90°$. Center frequency of the input signal is 2.25 MHz.

epoxy layer) and the open circles are the spectrum corresponding to the gated time domain signal indicated in Fig. 4(b). It can be seen that, if attenuation is not considered, the reflection amplitude is equal to unity at the stop-bands zones and at the end of the stop zones the reflection coefficient suddenly drops almost to zero. As a result of the stop and pass bands phenomena in the frequency domain the periodic medium acts as a filter. One has to select the signal frequency very carefully according to the application to avoid or exploit these stop and pass band effects. As shown in Fig. 4(b), the received reverberation due to the first stop band may overlap the reflection from an inside defect in the laminas. However, gating these reverberations and obtaining the resonance frequency is used to evaluate the averaged thickness of the lamina with the interface residual epoxy layer.

The structure geometry discussed above assumed that the residual layer between each composite lamina has the same properties. Therefore the residual and composite layers form a perfect periodic structure. If the interface properties such as the layer thickness randomly vary, they form a random ''quasi-periodic'' medium. Figures 5(c), (d) show the reflection coefficient for a wave normally incident from fluid onto this ''quasi-periodic'' semi-space. In this calculation, we assume the thickness of the residual layer is a random variable with uniform distribution between 5 μm to 7 μm (6 ± 1 μm). As shown in these figures, the variation of the residual layer thickness has little effect on the stop and pass band positions. However, it introduces random sound reverberation and scattering in the pass bands. This reverberation is extremely high in the pass band for the interface layers without attenuation [Fig. 5(c)]; however it is significantly reduced if one adds attenuation for the epoxy layer [Fig. 5(d)]. The ''quasi-periodic'' media in this content have been studied in detail by Lu and Achenbach[29] where they considered a random imperfect interface modeled by spring boundary conditions between composite laminas. As was discussed by Lu and Achenbach,[29] due to the interface randomness the transmittivity in the pass zones decreases with frequency increase; in our case this is exhibited by noise level increases as one can observe in Fig. 5(d).

TABLE I. Properties of one lamina.

Elastic constants of one lamina (GPa)	
C_{11}	143.2-1.4i
$C_{22}=C_{33}$	15.8-0.5i
C_{12}	7.5-0.1i
C_{23}	8.2-0.1i
C_{55}	7.0-0.3i
Density (g/cm^3)	1.6
Thickness (mm)	0.194

B. Oblique incidence response

The interpretation by Floquet wave theory of the phenomena observed at normal incidence is quite straightforward. The oblique incidence discussed in this section is much more complicated. Figure 6 shows experimental results for the double through transmission amplitude from 0° to 80° incident angle at four orientation angles α: 0°, 25°, $-25°$, 90° for the same composite sample. The incident wave is a pulse with center frequency 2.25 MHz. From these figures one can observe that the transmission amplitude drops very fast when the incident angle changes from 0 to 10 degrees. Sharp amplitude drop in this range is nearly independent of the incident plane orientation angle α and is explained by the very large difference of the in-plane and out-of-plane lamina elastic anisotropy (Table I). At incident angles between 20 and 40 degrees, there is almost no transmitted energy. Then a very obvious transmission peak emerges around 50 degrees and its maximum position and amplitude changes for different orientation angles. It must be noted that at 25° orientation angle there is another transmission peak around 15° incident angle. The time domain signals at several incident angles for the top lamina orientation $\alpha=25°$ are shown in Figs. 7(a), 7(b). The theoretical results indicated by the solid lines are calculated using the double-through-transmission beam model [Eq. (1)] with beam size $B_t=6$ mm and material properties given in Table I. One can see that the agreement with experimental results is very good. At normal incidence (0°), the first transmitted and

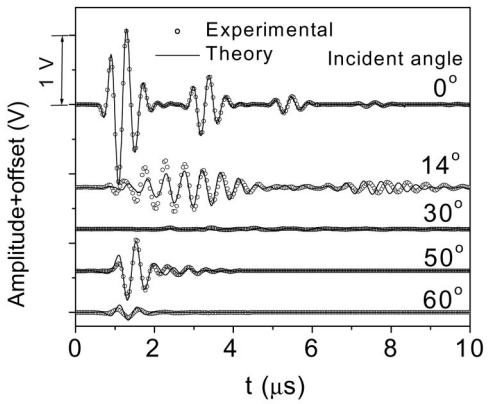


FIG. 7. Experimental (open circles) and theoretical (solid lines) time domain signal of different incident angles. The rotation angle α is 25° and center frequency of the signals is 2.25 MHz.

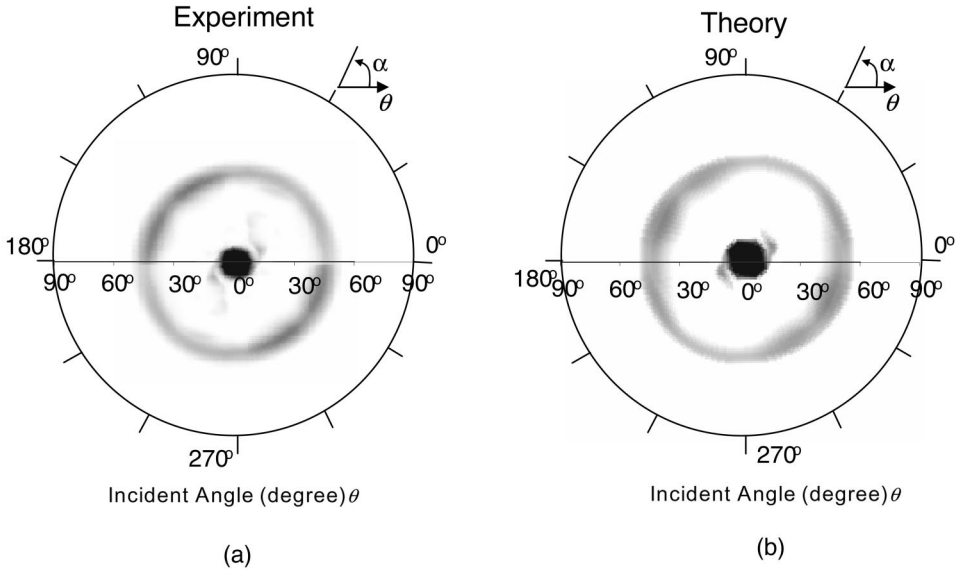


FIG. 8. Effects of rotation (α) and incident angles (θ) on the double transmission peak amplitude for the 3.1 mm $[0/-45/90/45]_{2s}$ composite. The transducer center frequency is 2.25 MHz. The brightness represents the transmission amplitude value. (a: experiment, b: theory.)

multiple reflection echoes are very clear and there are no reverberation signals due to reflections from the excess epoxy layers between the laminas. At 14° incident angle, the transmitted signal reaches the first maximum and clearly exhibits dispersion. The amplitude drops after 14° and becomes almost the same as the base level signal at 30°. From the 30° incident angle the transmitted signal amplitude increases again and reaches a second maximum at 50°. Spectrum analysis shows that the transmitted signal has a center frequency of about 2 MHz at 50° incident angle.

To further investigate the transmission properties of the composite, double-through-transmission measurements were carried out at orientation angles from 0° to 180° with steps of 1°. The transmission peak amplitude is represented in the image of Fig. 8(a). In this image, the radial direction represents the incident angle θ and the circumferential direction represents the orientation angle α. The gray level represents the amplitude; the darker the gray level, the higher the amplitude level. Figure 8(b) shows the corresponding theoretical transmission peak amplitude distribution calculated for the composite laminate using Eq. (1). By comparing both the calculation and experimental results, one can see the following: (1) The double-through-transmission amplitude has 180° rotation symmetry; (2) It has no reflection symmetry (α and $-\alpha$); (3) At incident angles close to normal incidence ($<7°$), the multidirectional composite is similar to an isotropic layer; (4) It is very interesting that both experiment and theoretical calculation agree very well that there is a transmission peak when $10°<\theta<20°$ and $10°<\alpha<75°$. The transmitted signals at this peak are clearly dispersion signals; (5) Larger transmission amplitude also appears at a larger incident angle between 45° to 55°. Within this range, the maximum transmission amplitude depends on rotation angle. Maximum values appear around rotation angle 22° and 60°.

For the pulse signal with 2.25 MHz center frequency, it is clearly shown from Figs. 7 and 8 that there is an incident angle range around 50° allowing the energy to be transmitted. But it is also very important to know whether this transmission peak exists when the frequency increases. Figure 9 compares the double transmission amplitude for both 2.25 MHz pulse and 5 MHz tone burst signal cases at orientation angle 45°. At 5MHz, very little energy is transmitted through the composite laminate around incident angle $\theta=50°$. Theoretical calculation also shows similar results. This puts a limit on using higher frequency signals for oblique measurements on this composite laminate.

Above we described experimentally and theoretically the transmission phenomena for quasi-isotropic composites with $[0/-45/90/45]$ cells. Here we will show theoretically the generality of these results for other quasi-isotropic composites which are made of four angles 0°, 90° and ±45°. There are $24(4\times3\times2\times1)$ possible lay-ups for the cells. However, these cells are related to each other by coordinate transformation. For example, $[45/90/-45/0]$ can be obtained by rotating $[0/45/90/-45]$ by 45°. $[0/-45/90/45]$ and $[0/45/90/-45]$ are symmetrical about 0°. Considering these prop-

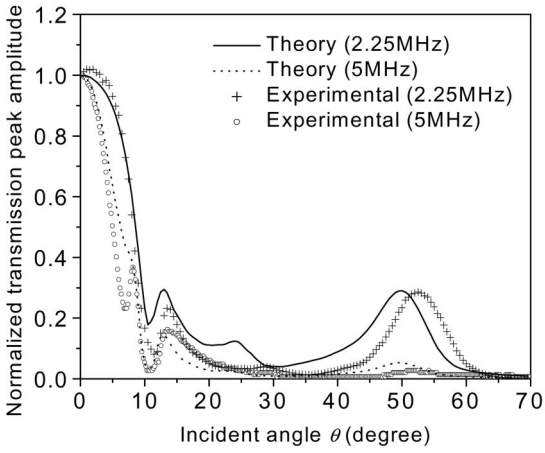
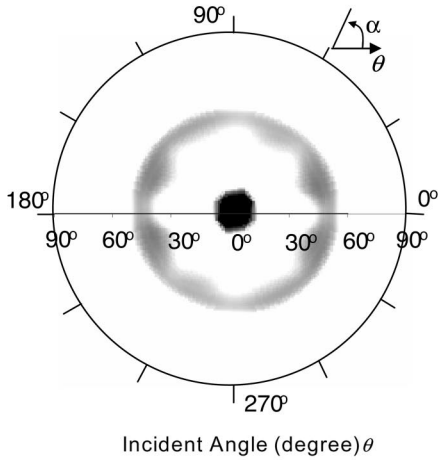


FIG. 9. Peak amplitude of double transmission signal versus incident angle θ at orientation angle $\alpha=45°$ for the $[0/-45/90/45]_{2s}$ composite for a tone burst with center frequency of 2.25 MHz and 5 MHz. The amplitude is normalized by the peak amplitude in normal incidence. The solid and dotted lines represent simulations and symbols (cross and open circles) are experimental results.



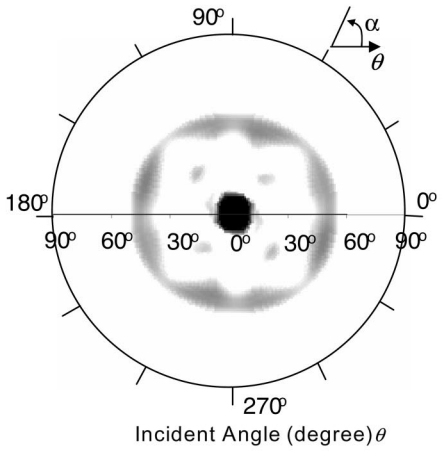


FIG. 10. Calculated transmission peak amplitude distribution as a function of orientation angle α and incident angle θ for an ultrasonic pulse with center frequency 2.25 MHz for different composite cells. (a) $[0/90/45/-45]_{2s}$, (b) $[0/45/-45/90]_{2s}$.

erties, there are only three independent cells $[0/-45/90/45]$, $[0/45/-45/90]$ and $[0/90/45/-45]$. The behaviors of other cells can be deduced from these three cells by symmetry operations. Therefore, we only need to investigate the transmission properties of the three independent cells versus orientation and incident angles. Figure 10 shows polar diagrams of the theoretical transmission peak amplitude distribution for composites $[0/90/45/-45]_{2s}$ and $[0/45/-45/90]_{2s}$. The parameters used to calculate this figure are the same as those for Fig. 8(b). Comparing with the results for the composite $[0/45/90/-45]_{2s}$ given in Fig. 8(b), one can see the main transmission features (four main divided ranges) are similar for these three composite cells.

These features listed above are important for composite evaluation and deserve better investigation. First, we consider a composite as an infinite repetition of cells [0/45/90/−45]. The relation between transmission peak amplitude distribution and Floquet wave pass and stop bands is discussed in Sec. V.

V. INTERPRETATION OF OBLIQUE THROUGH-TRANSMISSION SPECTRA

A. Floquet wave stop and pass bands

As mentioned before, multidirectional composites are usually made by a repetition of one unit cell. In order to interpret the general transmission features described above, let us first consider the composite as an infinite periodic medium. The waves propagating in such a medium are Floquet waves.[4,6,17–23] The relation between Floquet wave number ζ and cell stiffness matrix $\mathbf{K}^c$ is given by the characteristic Eq. (8). Due to the 2π periodicity of $e^{i\zeta h_c}$, the Floquet wave number ζ is not uniquely defined. In the wave number interval $-\pi<\zeta h_c<+\pi$ for given (ω,k_x,α), we can uniquely determine six Floquet wave numbers ζ_i $(i=1,...,6)$ as roots of Eq. (8). For generally anisotropic material all six roots may be different. Three of them, ζ_i $(i=1,...,3)$, correspond to waves propagating (energy flux direction) or decaying along the $+z$ direction and the other three ζ_i $(i=4,...,6)$ to waves propagating or decaying in the $-z$ direction. Since the z axis is a symmetry axis for the composites shown in Fig. 1, the wave numbers of the three Floquet waves propagating in the $+z$ direction have the same values as those propagating in the $-z$ direction with only change of sign $\zeta_i=-\zeta_{i+3}$ $(i=1,...,3)$. Therefore we need only consider the three independent Floquet waves. Real wave numbers correspond to propagating Floquet waves, while complex wave numbers are associated with waves whose amplitudes decay exponentially along either the positive or negative direction of the z-axis. The Floquet wave numbers are related to frequency ω, horizontal wave number k_x, orientation angle (α) and cell properties. The values (ω,k_x,α) for which ζ is real form the pass bands and those for which ζ is complex form the stop bands. In order to compare with the transmission peak amplitude distribution shown in Fig. 8, the horizontal wave number k_x can be transformed to an equivalent incident angle θ_i from water based on Snell's law $\sin(\theta_i)=k_xV_f/\omega$. Figure 11(a) shows the Floquet wave stop and pass bands in the frequency and incident-angle domains. The four gray levels represent the number of Floquet waves with real wave number. White represents the stop band for all three Floquet waves (complex wave numbers) and black corresponds to the pass band for all. As discussed in Ref. 22, in the low frequency range the multidirectional composite can be homogenized. The upper frequency bound f_h of the homogenization domain is given by

$$\frac{h_c}{\lambda_{\min}(\theta_i)}\cos\theta_i<0.5 \quad \text{or} \quad f_h<\frac{0.5V_{\min}(\theta_i)}{h_c\cos\theta_i}, \tag{9}$$

where $\lambda_{\min}$ and $V_{\min}=\omega/k_{\max}$ are the minimum wavelength and velocity at a given incident propagation angle θ_i among all the bulk waves in each lamina of the cell. h_c is the cell-thickness. At incident angles close to normal, since the properties in the z-direction are the same for all laminas the pass bands dominate (we neglect the residual epoxy layer in considering the oblique incident wave because the frequency used is less than 7.9 MHz). As one can see from Eq. (9) the first stop band appears at cell thickness half the wavelength. For larger incident angles, because waves are propagating almost completely along the x direction, the wave number in

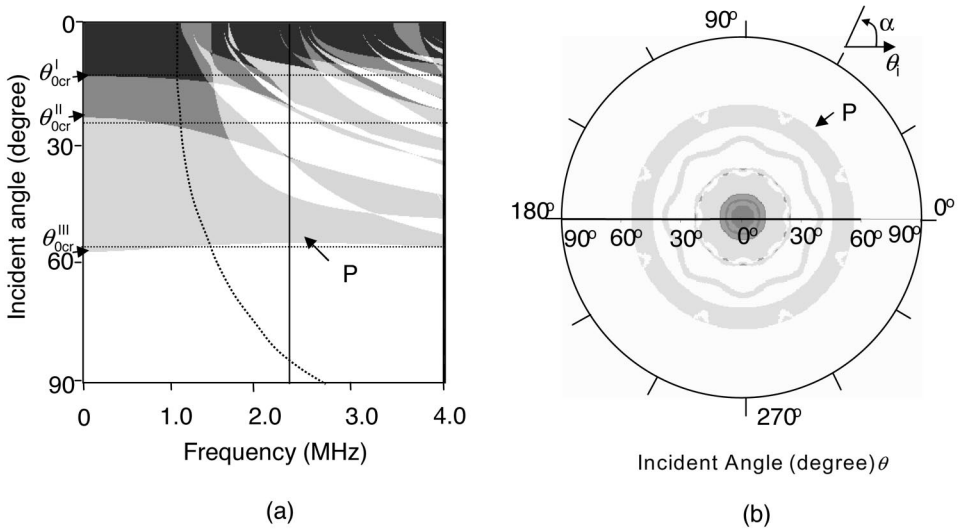


FIG. 11. Pass and stop bands of the three Floquet waves for an infinite periodic [0/−45/90/45] composite. P indicates the pass band at larger incident angle (45°–60°). (a) As a function of incident angle θ and frequency, the propagating plane is oriented along 0°. Arrows on the vertical axis indicate the three critical angles. The three dashed horizontal lines show the estimated critical angles and the dashed curve presents the estimated frequency upper bound of the homogenization domain. (b) As a function of incident angle θ and orientation α frequency is 2.25 MHz. (black: three propagating waves, gray: one or two propagating waves, white: stop band, no propagating waves permitted.) "P" indicates the pass band at larger incident angle at 2.25 MHz.

the z direction is smaller, i.e., $\cos\theta_i$ will be small in Eq. (9); therefore the homogenization domain is extended to much higher frequencies at larger incident angles as shown in Fig. 11(a) and one pass band still exists although total stop bands (white area) already appear at small and intermediate incident angles. At intermediate incident angles, a complicated pass and stop band spectrum exists and it is dominated by stop bands. Similar to the normal incidence case, these stop bands will behave as filters in both frequency and incident angle domains and lead to a significant distortion of time-domain transmitted signals, as shown in Fig. 7 at incident angle 14°.

Figure 11(b) shows the stop and pass bands of the three Floquet waves as a function of incident angle θ_i and rotation angle (α) at frequency 2.25 MHz for the [0/−45/90/45] cell. As can be seen, all three Floquet waves have a pass band at small incident angles (black circle at the center) at this frequency. The size of this pass band is dependent on frequency as shown in Fig. 11(a). Below the homogenization zone, the size is determined by the first critical angle of the Floquet wave. At higher frequency, the size of this pass band can be determined approximately by the smallest critical angle of the laminas in the cell. A pass band for one of the Floquet waves which is almost independent of rotation angle appears at incident angles from 40° to 60°. Comparing this figure with the transmission amplitude distribution shown in Fig. 9, one can see that the pass and stop bands of the Floquet wave determine the transmission amplitude distribution. Large transmission coefficient appears at incident angles within the pass band.

B. Floquet wave critical angles

Within the low frequency domain, the composite is considered as an effective homogeneous medium, where similar to homogeneous media, one can observe three critical angles as shown in Fig. 11(a) (they are indicated by arrows on the vertical axis). These critical angles determine the transmission range for each mode in the incident angle domain. For applications, it is important to estimate these angles for the wave transmission analysis in the homogenization domain and for an incident angle selection.

To find these angles, let us first consider for simplicity a periodic medium whose cell has only two isotropic layers as shown in Fig. 12(a). We also consider only the propagation of an SH wave and assume that the velocity V_1 in layer one

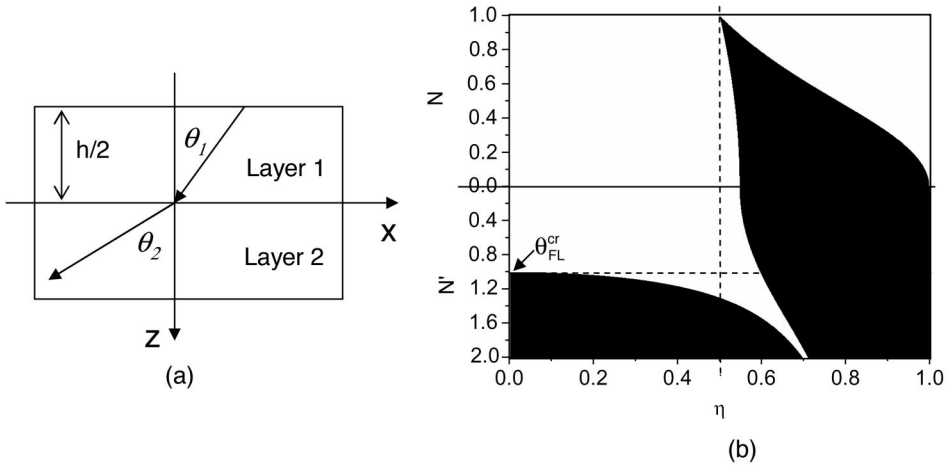


FIG. 12. (a) A simple model for homogenization domain estimation. The cell has only two layers and only the SH wave is considered. (b) Floquet wave stop and pass bands for the two layer isotropic model. The horizontal dashed line shows the critical angle $N'=1$. The dashed vertical line indicates the upper bound of the homogenization domain $\eta=0.5$.

is smaller than the velocity V_2 in layer two while the density and thickness $h/2$ are the same for both layers. Let us define nondimensional parameters $\eta=(k_{z1}h/2\pi)$, $N=(k_{z2}/k_{z1})$ (for real k_z) and $N'=[\mathrm{Im}(k_{z2})/k_{z1}]$, where k_{z1} and k_{z2} are the z-direction (normal) components of the wave number in layers one and two respectively. The first two Floquet wave pass and stop bands for this system are shown in Fig. 12(b). From this figure, one can see that the critical angle at zero frequency is given by $N'=1$.

To clarify the physical meaning of the critical angle criterion $N'=1$ at zero frequency, let us consider the static homogeneous periodic medium with the two-layer cell shown in Fig. 12. The static effective modulus G_{eff} is given by

$$G_{\mathrm{eff}}=(h_1+h_2)G_1G_2/(h_1G_2+h_2G_1). \tag{10}$$

h_1 and h_2 are the thicknesses and G_1 and G_2 are the moduli of the cell layers one and two respectively. If we assume thickness $h_1=h_2$ and density $\rho_1=\rho_2$ as discussed above, then Eq. (10) can be written in terms of the effective velocities for the periodic medium

$$V_{\mathrm{eff}}^2=2V_1^2V_2^2/(V_1^2+V_2^2). \tag{11}$$

Now, let us select any slow speed V_1 layer in the system and consider the media above and below it as effective media. One can find a critical angle θ_{Fl}^{cr} for this effective homogeneous medium V_{eff}. Applying Snell's law to the effective medium and to the slow and fast layers in the cell we have

$$\frac{\sin\theta_1}{V_1}=\frac{\sin\theta_2}{V_2}=\frac{1}{V_{\mathrm{eff}}}=\frac{\sqrt{V_1^2+V_2^2}}{\sqrt{2}V_1V_2}. \tag{12}$$

Thus at the Floquet wave critical angle, the k_z projection of the wave numbers in each layer of the cell satisfy the condition:

$$k_{z1}=\frac{\omega}{V_1}\sqrt{1-\sin^2\theta_1}=\frac{\omega}{V_2}\sqrt{\sin^2\theta_2-1}=ik_{z2};\ \text{ and }\ N'=1. \tag{13}$$

This is also applied to a periodic semi-space loaded by fluid with velocity V_0; the critical angle for the incident wave is given by $\theta_{Fl}^{cr}=\sin^{-1}(V_0/V_{\mathrm{eff}})$.

As frequency increases, energy becomes more and more localized in the slow layers because the wave in the fast layer is a nonpropagating wave and the decay rate increases with frequency. This decreases the contribution of the fast layer in Eq. (10) and the reduction of the effective modulus, so the critical angle will decrease with frequency as shown in Fig. 12(b). The above discussion provides a very clear interpretation of the critical angle for Floquet waves and shows that it is closely related to the behavior of the effective homogeneous medium: i.e., in the homogenization domain [Eq. (9)] the ultrasonic response is absolutely identical in the actual periodic medium and a properly-introduced effective medium.[22]

When a cell is built from two anisotropic laminas, the wave number k_z is not only dependent on incident angle (or the wave number projection k_x) but also on the incident plane orientation. In this case the three critical angles are given by $N'=[\mathrm{Im}(k_{z1}^{L,T})/k_{z2}^{L,T}]=1$, where L,T denotes quasi-longitudinal and quasi-transverse modes (the values of the critical angles in this approximation are shown in Fig. 11(a) by the horizontal dashed lines).

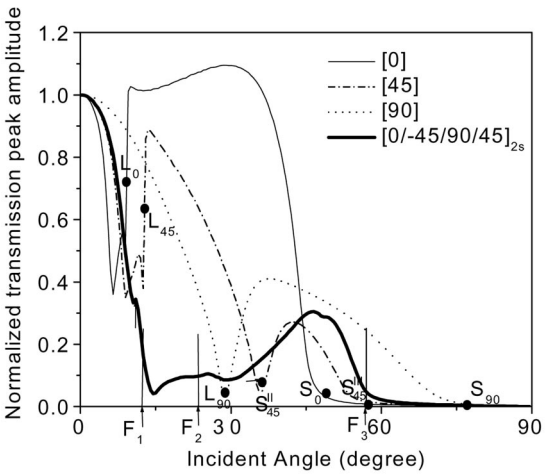


FIG. 13. Calculated transmission peak amplitude distribution as a function of incident angle θ for an ultrasonic pulse with center frequency 2.25 MHz for unidirectional composites $[0]_4$, $[90]_4$ and $[45]_4$, and quasi-isotropic composites $[0/-45/90/45]_{2s}$. The critical angles are marked "L_0" and "S_0" for the [0] unidirectional lamina, "L_{90}" and "S_{90}" for the [90] lamina and "L_{45}" and "S_{45}^{II}," "S_{45}^{III}" for the [45] lamina. "L" denotes the longitudinal wave and "S" transverse waves. "F_1," "F_2," and "F_3" are the Floquet wave critical angles for the quasi-isotropic composite.

To show the effect of the critical angles, we have compared the pulse peak transmission amplitude for the quasi-isotropic composite with those for a unidirectional composite of the same thickness and incident plane orientation 0°, 45° and 90° to the fiber direction in Fig. 13. The center frequency of the input signal is 2.25 MHz. The critical angles are marked "L_0" and "S_0" for the [0] unidirectional lamina, "L_{90}" and "S_{90}" for the [90] lamina and "L_{45}" and "S_{45}" for the [45] lamina on the corresponding transmissitivity curves (here "L" denotes the longitudinal wave and "S" shear waves critical angles). To compare with the experiment we simulated a tone burst signal; for this reason the transmittivity curves have a smooth transition through the critical angles. "F_1," "F_2" and "F_3" are the Floquet wave critical angles for the quasi-isotropic composite. For the unidirectional composites, initially the peak transmitted amplitude decreases with the incident angle and reaches the first minimum at an incident angle slightly less than the longitudinal critical angle. This minimum is formed when with increase of incident angle the transverse wave becomes the dominant energy carrier in the composite. Between the critical angles, there is always a peak in the transmission amplitude. For the quasi-isotropic composite, initially the amplitude also decreases and reaches a minimum around the first F_1 Floquet wave critical angle. There is also a transmittivity peak before the third F_3 Floquet wave critical angle. The first critical angle for the Floquet wave is above the critical angle L_0 and almost coincides with the L_{45} (longitudinal critical angle for 45° rotated unidirectional composite). F_3 almost coincides with the third critical angle S_{45}^{III} and is below the second critical angle S_{90}. The transmittivity peak for the quasi-isotropic composite is slightly above the transmittivity peak

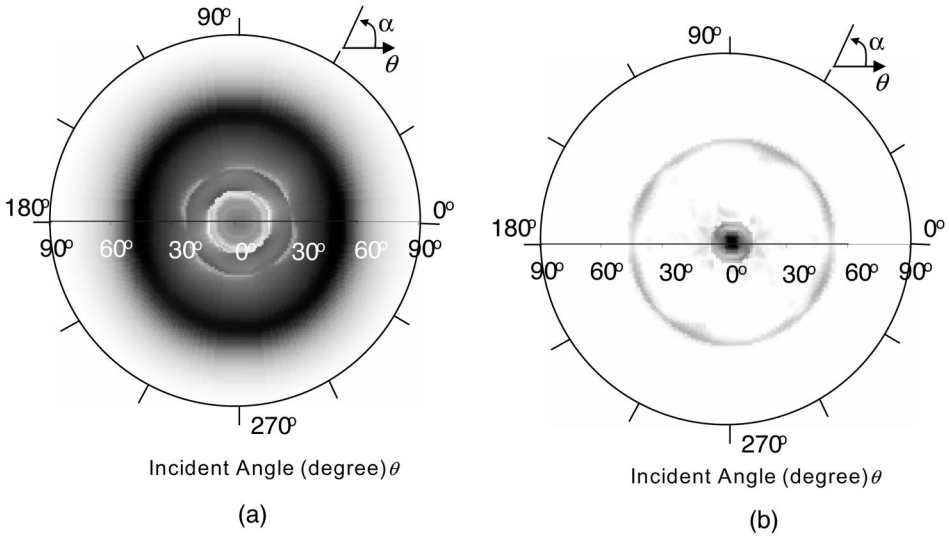


FIG. 14. Calculated transmission peak amplitude distribution as a function of orientation angle α and incident θ angles for the $[0/-45/90/45]_{2s}$ composite for an ultrasonic pulse with center frequencies (a) 0.5 MHz and (b) 5 MHz.

through the 45° composite corresponding to a slow quasi-transverse wave between the second S_{45}^{II} and the third S_{45}^{III} critical angles.

C. Examples for finite thickness composites

The Floquet wave stop and pass bands are characteristics of an infinite periodic medium. One can expect for a finite thickness composite laminate that the number of cells will affect the dependence of transmission amplitude on angle and frequency. As shown in Fig. 11(a), the structure of the Floquet wave stop and pass bands depends significantly on frequency. If the frequency of the incident signal is below the homogenization frequency (typically around 1 MHz for $[0/-45/90/45]$ composites), as discussed in Ref. 23, one can treat the laminate as a homogeneous layer. Figure 14(a) shows, as in Fig. 8, a computed transmission peak amplitude polar diagram for a $[0/-45/90/45]_{2s}$ composite, but at 0.5 MHz center frequency. One can see that the transmission amplitude is almost independent of orientation angle (α) due to the quasi-isotropic properties of this composite. A larger transmission amplitude at intermediate incident angle (20°–40°) also exists due to the absence of Floquet wave stop bands at this frequency. Figure 14(b) shows the transmission peak amplitude polar diagram at 5 MHz center frequency for the same composite. The complicated distribution of Floquet wave stop and pass bands prevents wide band signal transmission at this frequency and the transmission window at larger incident angle 50° observed for 2.25 MHz becomes much smaller.

Figures 15(a) and (b) show the transmission peak amplitude distribution for composites $[0/-45/90/45]_s$ and $[0/-45/90/45]_{8s}$ respectively at center frequency 2.25 MHz. Comparing these figures with Fig. 11(b), one can see that for the $[0/-45/90/45]_s$ composite the transmission window at larger incident angle becomes wider and its width depends strongly on the orientation angle; there are also transmission windows at intermediate incident angles with more complicated patterns. As the repetition number increases, as shown for the $[0/-45/90/45]_{8s}$ composites in Fig. 15(b), only the transmission window at larger incident angle exists. This

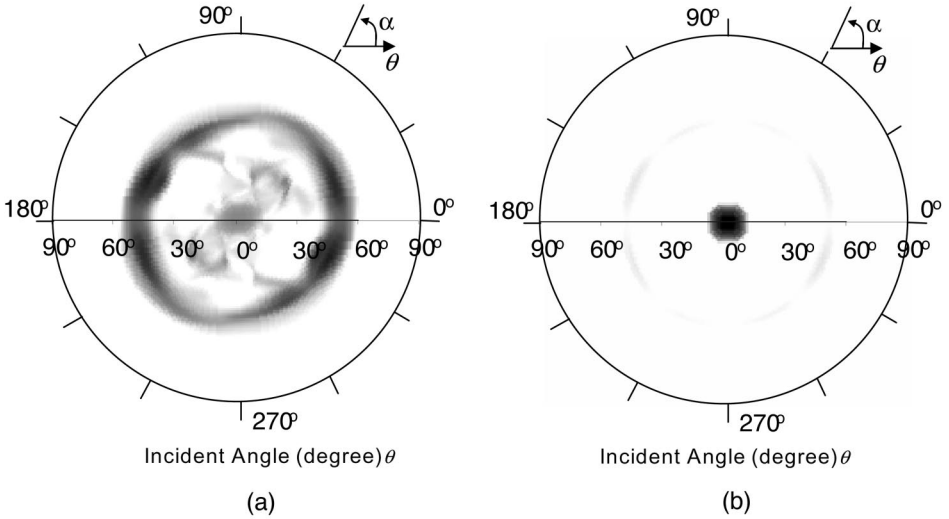


FIG. 15. Calculated transmission peak amplitude distribution as a function of orientation angle α and incident angle θ for an ultrasonic pulse with center frequency 2.25 MHz. Effect of number of cell repetitions on transmission amplitude. (a) $[0/-45/90/45]_s$, (b) $[0/-45/90/45]_{8s}$.

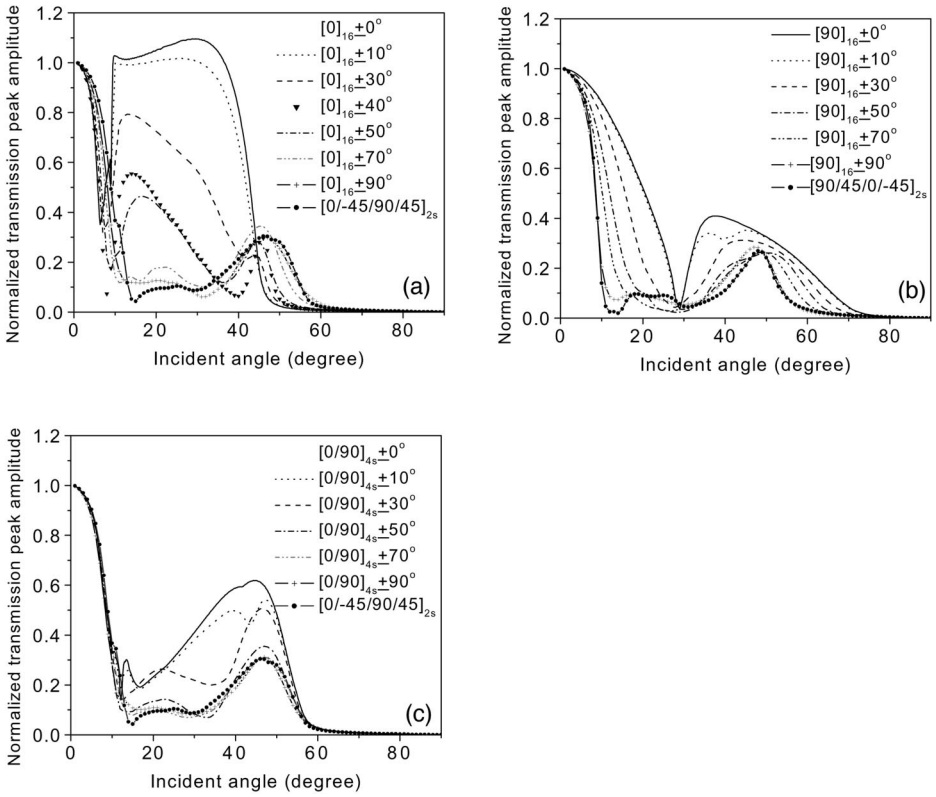


FIG. 16. Effect of lamina lay-up randomness on the averaged transmission pulse amplitude; for each angle (β) deviation range it is averaged over ten independent random angle deviation lay-up sets. The magnitudes of randomness are indicated in the legend by deviation angle range $\pm\beta°$ ($\pm 10°$, $\pm 30°, \pm 40°, \pm 50°, \pm 70°, \pm 90°$). $\pm 0°$ represents perfect lay-up and $\pm 90°$ corresponds to fully random lay-up. The center frequency of the ultrasonic pulse is 2.25 MHz and composite orientation angle $\alpha = 0°$ relative to the incident wave plane. The solid circles are the results for a quasi-isotropic composite $[0/-45/90/45]_{2s}$ with perfect lay-up. (a) Unidirectional composite $[0]_{16}$ with different randomness in the lamina lay-up. (b) Unidirectional composite $[90]_{16}$ with different randomness. (c) Cross-ply composite $[0/90]_{4s}$ with different randomness. In all cases the transmission angle spectra statistically converge to that of the regular quasi-isotropic composite.

window, as discussed above, corresponds to the pass band "P" of the Floquet wave spectra given in Fig. 11(b). For a larger number of repetitions, material damping also plays an important role and significantly reduces the size of this transmission window.

D. Effect of randomization in composite lay-up

Due to use of low frequencies for oblique incident wave the effect of residual epoxy layer, discussed in Sec. IV. A, is negligible; however the effect of lamina anisotropy is significant. In the previous discussion, we assumed perfect lay-up of the lamina orientation. To show the effect of possible random misalignments of the lamina orientation, we calculate the transmission pulse peak amplitude for three composites $[0]_{16}$, $[90]_{16}$, $[0/90]_{4s}$ with different randomness of the lamina lay-up as shown in Fig. 16. The transducer center frequency is 2.25 MHz. The random lay-ups are obtained by adding a random number for each lamina orientation angle in the original lay-up sequence. The random angle deviation number is uniformly distributed between $+\beta°$ and $-\beta°$. The value $\beta°$ represents the magnitude of the randomness ($\pm 0°$ corresponds to the perfect lay-up and $\pm 90°$ to the completely random lay-up). Ten independent random angle deviation sets were generated for each β deviation range. For each set of angle distributions the ultrasound transmittivity through the random laminate has been calculated and the result shown in Fig. 16 is the average transmittivity of the ten sets.

Figure 16(a) presents the results for a unidirectional composite $[0]_{16}$ with 16 laminas. For $[0]_{16}$ composite, when all laminas are oriented at $0°$, large transmission is observed between the first and second critical angles (see Fig. 13). If one introduces a random variation in the orientation of each of the 16 laminas, the transmission amplitude in this range decreases as randomness (β) increases. This is accompanied by the appearance of another transmission window at an incident angle slightly above the second critical angle (Fig. 13) and slight increase of the first critical angle with increase of randomness. Finally for completely random lay-up, the transmission amplitude becomes very similar to that of the regular $[0/-45/90/45]_{2s}$ quasi-isotropic composites. Figure 16(b) shows the results for the $[90]_{16}$ composite. In this case, the randomness lay-up reduces the transmittivity in the intermediate incident angle ($10°$ to $40°$) of the $[90]_{16}$ composite by reducing the first critical angle and formation of a wide stop band zone. Similarly the results for the $[0/90]_{4s}$ composite [Fig. 16(c)] show transmittivity reduction in this range of angles. From these figures, one can see that introduction of randomness in the lamina lay-up diminishes the wave transmittivity in the intermediate incident angle range and makes the transmission peak amplitude distribution similar to that of the $[0/-45/90/45]_{2s}$ quasi-isotropic composite.

It is very interesting that for all these cases with the increase of the range β of the random angle misalignment the averaged angle transmittivity response statistically converges to that for the regular quasi-isotropic lay-up $[0/-45/90/45]_{2s}$. This is illustrated by the histogram of normalized transmission amplitude for 10^5 completely random lay-up sets ($\beta = 90°$) at the transmittivity peak ($46°$ incident angle). The transmission amplitude of the regular composite $[0/-45/90/45]_{2s}$, averaged over orientation angle α from $0°$ to $360°$e is 0.291, is indicated by the vertical dashed line. It is positioned at the maximum of the distribution for the random lay-up composite (Fig. 17).

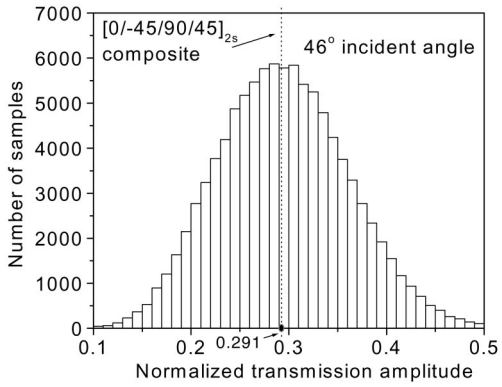


FIG. 17. Histogram of normalized transmission amplitude for completely random lay-up at 46° incident angle (10^5 composite lay-up sets, $\beta=90°$). The center frequency of the ultrasonic pulse is 2.25 MHz. The transmission pulse amplitude at maximum transmittivity (46°) for the regular $[0/-45/90/45]_{2s}$ composite is indicated by the vertical dashed line (it is averaged over orientation angle α from 0° to 360°).

At relatively lower frequency, 2.25 MHz, the lay-up randomness in the $[0/-45/90/45]_{2s}$ quasi-isotropic composite introduces only small deviations in the angle transmittivity spectrum relative to the composite with perfect alignment. As shown in Figs. 8 and 10, at this frequency the transmittivity of quasi-isotropic composites slightly varies with composite orientation angle α and has some regular transmittivity patterns in the intermediate incident angle range. The transmittivity of a composite with full random lay-up is independent of composite orientation to the incident plane. The transmittivity through such a random composite in relation to a perfect lay-up quasi-isotropic composite will be averaged over composite orientation angle α.

VI. CONCLUSION

Both modeling and experimental results show complicated features of wave propagation in a 16-layer quasi-isotropic composite $[0/-45/90/45]_{2s}$. Good agreement is observed between theoretical and experimental results. While for static mechanical properties this material is transversely isotropic with isotropy axis perpendicular to the composite surface, the ultrasonic response exhibits monoclinic symmetry at frequency above 2 MHz. In this frequency range the ultrasonic wave transmission decays very fast with increasing deviation of the incident angle from the normal direction. A transmission window is found at about 50° at central frequency 2.25 MHz. This makes it possible to use obliquely incident waves for composite evaluation. However the ultrasonic behavior is highly frequency dependent and scattering is significant for wave propagation at higher frequencies. These transmission phenomena have been interpreted in terms of Floquet waves. The effect of random deviation of the laminated structure periodicity has also been discussed: (a) randomness of the thickness of the residual epoxy layer and (b) randomness in the misalignment angle in lamina lay-up in the composite structure. The variation of the residual layer thickness does not change the overall stop and pass band structure at normal incidence, but introduces random scattering in the pass bands. With increase of randomness in the lamina angle misalignment the averaged angle transmittivity response statistically converges to that for the regular quasi-isotropic lay-up $[0/-45/90/45]_{2s}$.

ACKNOWLEDGMENT

The authors are grateful to Q. Xie who assisted with collecting the experimental data shown in Fig. 8(a).

APPENDIX A. ASYMPTOTIC METHOD: RECURSIVE ASYMPTOTIC STIFFNESS MATRIX METHOD (RASM) FOR WAVE PROPAGATION IN MULTILAYERED COMPOSITES

Wang and Rokhlin described[15] an asymptotic recursive stiffness matrix method based on a simple second-order solution for elastic wave interaction with a thin anisotropic layer.[30] For a monoclinic homogeneous layer, the second-order asymptotic stiffness matrix can be written as

$$\mathbf{K}_{11}=\begin{bmatrix} -K_{s1}-bC_{11}+M & -K_{s3}-bC_{16} & (C_{55}-C_{13})a \\ -K_{s3}-bC_{16} & -K_{s2}-bC_{66}+M & (C_{45}-C_{36})a \\ (C_{13}-C_{55})a & (C_{36}-C_{45})a & -K_{n3}-bC_{55}+M \end{bmatrix},$$

$$\mathbf{K}_{22}=\begin{bmatrix} K_{s1}+bC_{11}-M & K_{s3}+bC_{16} & (C_{55}-C_{13})a \\ K_{s3}+bC_{16} & K_{s2}+bC_{66}-M & (C_{45}-C_{36})a \\ (C_{13}-C_{55})a & (C_{36}-C_{45})a & K_{n3}+bC_{55}-M \end{bmatrix},$$

$$\mathbf{K}_{12}=\begin{bmatrix} K_{s1}-bC_{11}+M & K_{s3}-bC_{16} & (C_{55}+C_{13})a \\ K_{s3}-bC_{16} & K_{s2}-bC_{66}+M & (C_{45}+C_{36})a \\ (C_{13}+C_{55})a & (C_{36}+C_{45})a & K_{n3}-bC_{55}+M \end{bmatrix},$$

$$\mathbf{K}_{21}=\begin{bmatrix} -K_{s1}+bC_{11}-M & -K_{s3}+bC_{16} & (C_{55}+C_{13})a \\ -K_{s3}+bC_{16} & -K_{s2}+bC_{66}-M & (C_{45}+C_{36})a \\ (C_{13}+C_{55})a & (C_{36}+C_{45})a & -K_{n3}+bC_{55}-M \end{bmatrix}, \quad \text{(A1)}$$

where $K_{s1}=C_{55}/h$, $K_{s2}=C_{44}/h$, $K_{s3}=C_{45}/h$, $K_{n3}=C_{33}/h$, $b=k_x^2h/4$, $a=ik_x/2$, $M=\rho\omega^2h/4$; C_{IJ} are the elastic constants in compact form.

For computations the lamina is subdivided into a set of thin layers and the stiffness matrices of the lamina and of the total layered structure are obtained from the thin layer stiffness matrix (A1) using the recursive algorithm.[14] The solution has been demonstrated to be efficient and convergent to the exact solution.[15]

APPENDIX B. PERIODICITY AND SYMMETRY

For materials with structural periodicity, which is the case for most composites, the recursive method allows very efficient computation of the total stiffness matrix. Let us introduce a repetitive cell consisting of n different anisotropic layers; we define it as $[\theta_1/\theta_2/\ldots\ldots\theta_{n-1}/\theta_n]_{ps}$. Here θ_i indicates a layer in the cell p with the ply orientation angle θ_i. The index p indicates the number of cells in the layered structure without symmetry or half of the layers for the system with symmetry; the index s indicates the structure symmetry. Using the recursive algorithm,[14] one first calculates the compliance/stiffness matrix of the layered cell. The global compliance/stiffness matrix is calculated accounting for the repetition of the cells. If $p=2^i$, then only i recursive operations are required to obtain the total compliance/stiffness with p repetitive cells. If the structure is symmetric about its centerline, the total compliance/stiffness matrix $\mathbf{M}^B(\mathbf{M}^B_{ij})$ of the bottom half of the structure is related to that of the top half structure $\mathbf{M}^T(\mathbf{M}^T_{ij})$ as

$$\mathbf{M}^B_{ij}=-\mathbf{I}_2\mathbf{M}^T_{(3-i)(3-j)}\mathbf{I}_2,\quad (i,j=1,2), \tag{B1}$$

where

$$\mathbf{I}_2=\begin{bmatrix}1&0&0\\0&1&0\\0&0&-1\end{bmatrix}.$$

Therefore, once the compliance/stiffness matrix for the top half structure $[\theta_1/\theta_2/\ldots\ldots\theta_{n-1}/\theta_n]_p$ has been calculated, the stiffness matrix for the bottom half can be immediately determined by Eq. (B1).

[1] J. D. Achenbach, "Wave Propagation in Fiber Reinforced Composites," Ch. 8 in *Composite Materials*, Vol. 2, *Mechanics of Composite Materials*, edited by G. P. Sendeckyi (Academy Press, New York, 1974).

[2] A. K. Mal and T. C. T. Ting, Eds., *Wave Propagation in Structural Composites*, ASME_AMD, Vol. 90 (ASME, New York, 1988).

[3] S. I. Rokhlin, T. K. Bolland, and L. Adler, "Reflection and refraction of elastic waves on a plane interface between two generally anisotropic media," J. Acoust. Soc. Am. **79**, 906–915 (1986).

[4] A. H. Nayfeh, *Wave Propagation in Layered Anisotropic Media* (North-Holland, New York, 1995).

[5] A. H. Nayfeh and D. E. Chimenti, "Elastic wave propagation in fluid-loaded multiaxial anisotropic media," J. Acoust. Soc. Am. **89**, 542–554 (1991).

[6] S. I. Rokhlin and L. Wang, "Ultrasonic waves in layered anisotropic media: Characterization of multidirectional composites," Int. J. Solids Struct. **39**, 4133–4149 (2002).

[7] A. H. Nayfeh, "The general problem of elastic wave propagation in multilayered anisotropic media," J. Acoust. Soc. Am. **89**, 1521–1531 (1991).

[8] D. E. Chimenti and A. H. Nayfeh, "Ultrasonic reflection and guided wave propagation in biaxially laminated composite plates," J. Acoust. Soc. Am. **87**, 1409–1415 (1990).

[9] A. K. Mal, "Wave propagation in layered composite laminates under periodic surface loads," Wave Motion **9**, 231–238 (1988).

[10] L. Wang, S. I. Rokhlin, and N. N. Hsu, "Time resolved line focus acoustic microscopy of composites," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1998), Vol. 18B, pp. 1321–1328.

[11] S. K. Datta, *Comprehensive Composite Materials*, Vol. 1, edited by T. W. Chou (Elsevier, New York, 2000), Chap. 1.18, pp. 511–558.

[12] G. R. Liu and Z. C. Xi, *Elastic Waves in Anisotropic Laminates* (CRC Press LLC, Boca Raton, FL, 2001).

[13] L. Wang and S. I. Rokhlin, "Stable reformulation of transfer matrix method for wave propagation in layered anisotropic media," Ultrasonics **39**, 407–418 (2001).

[14] 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).

[15] L. Wang and S. I. Rokhlin, "Recursive asymptotic stiffness matrix method for analysis of surface wave devices on layered piezoelectric media," Appl. Phys. Lett. **81**, 4049–4052 (2002).

[16] D. E. Chimenti, "Guided waves in plates and their use in materials characterization," Appl. Mech. Rev. **50**, 247–284 (1997).

[17] A. M. Braga and G. Herrmann, "Floquet waves in anisotropic periodically layered composites," J. Acoust. Soc. Am. **91**, 1211–1221 (1992).

[18] C. Potel and J. Belleval, "Propagation in an anisotropic periodically multilayered medium," J. Acoust. Soc. Am. **93**, 2669–2676 (1993).

[19] C. Potel, J. Belleval, and Y. Gargouri, "Floquet waves and classical plane waves in an anisotropic periodically multilayered medium: Application to the validity domain of homogenization," J. Acoust. Soc. Am. **97**, 2815–2836 (1995).

[20] P. J. Shull, D. E. Chimeti, and S. K. Datta, "Elastic guided waves and the Floquet concept in periodically layered plates," J. Acoust. Soc. Am. **95**, 99–105 (1994).

[21] A. Auld, D. E. Chimenti, and P. J. Shull, "Shear horizontal wave propagation in periodically layered composites," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 319–326 (1996).

[22] L. Wang and S. I. Rokhlin, "Floquet wave homogenization of periodic anisotropic media," J. Acoust. Soc. Am. **112**, 38–45 (2002).

[23] L. Wang and S. I. Rokhlin, "Floquet wave ultrasonic method for determination of single lamina moduli in multi-direction composites," J. Acoust. Soc. Am. **112**, 916–924 (2002).

[24] D. E. Chimenti and S. I. Rokhlin, "Relation between leaky Lamb modes and reflection coefficient zeros for a fluid-coupled elastic layer," J. Acoust. Soc. Am. **88**, 1603–1612 (1990).

[25] L. Wang, B. Xie, and S. I. Rokhlin, "Determination of embedded layer properties using adaptive time-frequency domain analysis," J. Acoust. Soc. Am. **111**, 2644–2653 (2002).

[26] S. I. Rokhlin, Q. Xie, Y. Liu, and L. Wang, "Ultrasonic study of quasi-isotropic composites," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1999), Vol. 18B, pp. 1249–1256.

[27] O. I. Lobkis and D. E. Chimenti, "Three-dimensional transducer voltage in anisotropic materials characterization," J. Acoust. Soc. Am. **106**, 36–45 (1999).

[28] O. I. Lobkis, A. Safaeinili, and D. E. Chimenti, "Precision ultrasonic reflection studies in fluid-coupled plates," J. Acoust. Soc. Am. **99**, 2727–2736 (1996).

[29] Y. Lu and J. D. Achenbach, "Effects of random deviations in interface properties on the propagation of ultrasound in thick composites," J. Acoust. Soc. Am. **90**, 2576–2585 (1992).

[30] S. I. Rokhlin and W. Huang, "Ultrasonic wave interaction with a thin anisotropic layer between two anisotropic solids: II. Second-order asymptotic boundary condition," J. Acoust. Soc. Am. **94**, 3405–3420 (1993).

Capillary forces in the acoustics of patchy-saturated porous media

Yaroslav Tserkovnyak[a)]
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138

David Linton Johnson
Schlumberger-Doll Research, Old Quarry Road, Ridgefield, Connecticut 06877-4108

(Received 10 January 2003; revised 2 August 2003; accepted 2 September 2003)

A linearized theory of the acoustics of porous elastic formations, such as rocks, saturated with two different viscous fluids is generalized to take into account a pressure discontinuity across the fluid boundaries. The latter can arise due to the surface tension of the membrane separating the fluids. We show that the frequency-dependent bulk modulus $\tilde{K}(\omega)$ for wavelengths longer than the characteristic structural dimensions of the fluid patches has a similar analytic behavior to the case of a vanishing membrane stiffness and depends on the same parameters of the fluid-distribution topology. The effect of the capillary stiffness can be accounted for by renormalizing the coefficients of the leading terms in the low-frequency limit of $\tilde{K}(\omega)$. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1621009]

PACS numbers: 43.20.Jr [ANN] Pages: 2596–2606

I. INTRODUCTION

Using the Biot theory[1] for the acoustics of fluid-saturated porous media as a starting point, it was possible[2] to formulate a simple theory for the frequency-dependent bulk modulus, $\tilde{K}(\omega)$, of a composite consisting of a homogeneous porous frame heterogeneously saturated with two different viscous Newtonian fluids. It is assumed that saturation occurs in ''patches'' of 100% saturation of one or the other of the fluids. In the limit when the two fluids are identical, the theory reduces to the low-frequency regime of the Biot theory. The mismatch of the local properties of the composite across the patch boundaries leads to a dispersion and attenuation mechanism[2] for acoustic waves, which is physically different from that of the Biot theory. In particular, the propagatory modes of the Biot theory, as applied to a fully mono-saturated sample, are predicted to be nondispersive and non-attenuative in the relevant regime, $\omega \ll \omega_B$, where ω_B is the Biot crossover frequency. In the case of patchy saturation, on the other hand, the mismatch of the elastic properties at the patch boundaries leads to a redistribution of the fluids between the patches when the system is perturbed. Because of the viscosity, this mechanism governs energy dissipation and, therefore, the attenuation and dispersion of the propagatory modes. There is a crossover frequency, ω_c, below which the pore pressure can equilibrate across the patches and above which it does not.

This mechanism is particularly effective when there is a large contrast between the two fluids as occurs when one of them is a gas. In such a system, which was further studied in Ref. 3 by analyzing measurements of Ref. 4, the gas saturated patches of the sample may often be approximated as vacuum. The relevant formulas become quite a bit simpler in this case.

a)Electronic mail: yaroslav@physics.harvard.edu

While the fluid redistribution across the patch boundaries was of a central importance in Ref. 2, its theory failed to account for the membrane surface tension which can arise, for example, due to the pinning of the boundary edges inside the pores. The latter will result in the formation of a meniscus when the fluids are pushed between different patches, which in turn leads to a pressure drop between the patches due to the capillary forces. The present paper addresses this phenomenon. It is important to note that various additional capillary effects, such as the fractal nature of the fluid distribution in the pore space, can be important in realistic porous materials but lie outside the scope of this work.

Johnson[2] suggested that the patchy-saturation effect may be a dominant mechanism for the low-frequency ($\omega \ll \omega_B$) attenuation and dispersion in carbonate rocks, but can be dwarfed by microscopic ''squirt'' mechanisms in materials like sandstones. We indeed later showed[3] that the sonic and even ultrasonic measurements[4–6] on various limestones can be successfully described by the patchy-saturation effect. In particular, we demonstrated there how one can extract information about the characteristic patch sizes and shapes in partially water saturated limestones by measuring the velocity and attenuation of acoustic waves. Extending our theory to take into account the capillary forces is necessary to improve our understanding of the acoustics in patchy-saturated materials. While the effect of the membrane tension at the boundary is but one of the important effects, investigating its effects can be a starting point for further studies of the capillary effects in porous media.

The manuscript is structured as follows. In Sec. II we formulate the assumptions and limitations of the theory. Then, in Sec. III we in turn discuss the static, low-frequency, and high-frequency regimes of the theory in the general case of the patchy saturation by two arbitrary fluids. In Sec. IV an alternative formulation of the static and low-frequency limits is presented which is simple and physically transparent, al-

beit only valid for the case when one of the fluids is a gas. We compare our findings with exact numerical calculations in Sec. V for two simple geometries: that of the concentric spheres and the periodic slab geometries. Finally, in Sec. VI a simple theoretical model for $\tilde{K}(\omega)$ is presented enabling us to connect the low- and high-frequency asymptotes of the theory. Our results are then summarized in Sec. VII. Two somewhat technical discussions, on the capillary correction to the high-frequency limit and analytic structure of $\tilde{K}(\omega)$, are given separately from the main text, in Appendices A and B, respectively.

II. GENERALITIES

As in Ref. 2, we consider a porous rock fully saturated with two different fluids, one having global saturation S_1 and the other $S_2=1-S_1$. Each point of the sample is saturated with one fluid or the other and the local elastic properties of the composite are assumed to be governed by the usual Biot parameters of the rock skeleton and the locally present fluid. The only exception is the patch and the sample boundaries which deserve special treatment (in the following). While we will talk about various limits of the frequency dependence of the bulk modulus $\tilde{K}(\omega)$, the Biot theory is always assumed to be in its low-frequency regime. The Biot crossover frequency[1]

$$\omega_B=\frac{\eta\phi}{k\rho_f\alpha_\infty} \tag{1}$$

(using the notation of Ref. 7), defined in terms of the fluid density ρ_f, the viscosity η, the tortuosity of the pore space α_∞, the dc permeability k, and the porosity ϕ, is thus the upper frequency limit of our theory, i.e., it is assumed that $\omega\ll\omega_B$. Biot equations of motion within each fluid patch[1] then simplify to

$$\nabla\cdot\boldsymbol{\tau}=0, \tag{2}$$

$$\nabla p_f=\frac{i\omega\phi\eta}{k}(\mathbf{U}-\mathbf{u}), \tag{3}$$

where all quantities vary as $\mathbf{u}(\mathbf{r})e^{-i\omega t}$. Here, p_f is the acoustic component of the pore pressure, $\mathbf{U}$ and $\mathbf{u}$ are the displacement of the fluid and solid phases, respectively, and $\boldsymbol{\tau}$ stands for the total stress due to both (solid and fluid) phases. Note that Eqs. (2) and (3) do not contain the inertia terms, which is a relevant approximation as the attenuation/dispersion mechanism of our theory is governed by the slow diffusive mode. The pore pressure p_f and the total stress $\boldsymbol{\tau}$ are related to the deformation strain in the solid phase, $\epsilon_{ij}=(1/2)[u_{i,j}+u_{j,i}]$, and to that in the fluid phase, $E_{ll}=U_{l,l}$, by[1]

$$\boldsymbol{\tau}=[(P+Q-2N)\epsilon_{ll}+(R+Q)E_{ll}]\delta_{ij}+2N\epsilon_{ij}, \tag{4}$$

$$p_f=-\frac{1}{\phi}[Q\epsilon_{ll}+RE_{ll}], \tag{5}$$

where the coefficients P, Q, and R depend on the bulk moduli of the solid phase, K_s, the porous frame, K_b, and the pore fluid, K_f, as well as the porosity and the shear modulus of the porous skeleton, N (see, e.g., Ref. 8). The fast compressional and shear velocities of the Biot theory are nondispersive and nonattenuative whereas the slow compressional wave is diffusive in the low-frequency limit considered here. Furthermore, we assume that the wavelengths of the fast and shear waves are larger than the characteristic fluid-patch dimension L, i.e., $\omega\ll\omega_x$, where

$$\omega_x=\frac{2\pi V_{\mathrm{sh}}}{L} \tag{6}$$

with V_{sh} being the (slower) shear-wave velocity. The latter assumption will allow us to define the effective bulk modulus $\tilde{K}(\omega)\overset{\mathrm{def}}{=}-V(P_e/\delta V)$, where $(\delta V/V)e^{-i\omega t}$ is the oscillatory fractional volume change in response to an external normal stress, $\boldsymbol{\tau}\cdot\hat{\mathbf{n}}=-P_e\hat{\mathbf{n}}e^{-i\omega t}$, $\hat{\mathbf{n}}$ being the outward normal of the surface. Finally, the sample is considered to be isotropic and homogeneous, except for the fluid patches.

Due to capillary forces, the fluid pore pressure is in general discontinuous across patch boundaries. Absent any applied stress, the pressure discontinuity across the boundary, i.e., the capillary pressure, $P_f^{(1)}-P_f^{(2)}$, is related to the mean curvature of the interface within each pore and to the surface tension via the Laplace–Young equation. In a completely different context than we are considering in this article, one is led to the equations of two-phase flow in which the partial saturation is generally treated as a well-defined function of the capillary pressure. See, e.g., Bear.[9] These concepts make no distinction between partial saturations that are patchy and those in which each pore may be partially saturated. Indeed, $P_f^{(1)}$ and $P_f^{(2)}$ are each generally considered to be macroscopic quantities defined at each material point throughout the system.

In the present context, however, we assume the fluid patches are perturbed from equilibrium by the application of an additional small-amplitude acoustic wave. The changes in the pore pressures are small compared to the equilibrium values: $|p_f(\mathbf{r})|\ll|P_f^{(j)}|$, $\{j=1,2\}$, where $p_f(\mathbf{r})$ is the acoustic component of the pore pressure. On a microscopic level, the meniscus bends in and out, relative to its rest position. On a macroscopic level, there is an oscillatory displacement of fluid relative to the solid frame, at the boundary of the patch. This effect is conventionally captured by postulating the following boundary condition at the interface between two patches:

$$p_f^{(1)}-p_f^{(2)}=W\phi(\mathbf{U}-\mathbf{u})\cdot\hat{\mathbf{n}}. \tag{7}$$

Here, $p_f^{(j)}$ is the acoustic pore pressure on the jth side of the boundary surface and $\hat{\mathbf{n}}$ is an outward normal from patch 1 to patch 2. $\mathbf{U}\cdot\hat{\mathbf{n}}$ and $\mathbf{u}\cdot\hat{\mathbf{n}}$ represent the components of the fluid and the solid displacements, respectively, that are normal to the patch surface; each of these entities is separately continuous across the patch boundary. Thus, Eq. (7) simply states that the additional amount by which the fluid interface bulges is proportional to the additional capillary pressure induced by the acoustic wave; the spring constant, W, is often called the ''membrane stiffness.'' Indeed, Nagy and Blaho[10] have shown that Eq. (7) follows from the Laplace–Young equation applied to the case of pores that are circular cylinders. In that example they showed that $W=\sigma/k$, where σ is the surface tension between the two fluids. This specific example illustrates the point that surface tension manifests itself as a

stiffer membrane in smaller pores than in larger ones. More generally, the membrane stiffness is predicted to take the form of

$$W = s\sigma/k, \tag{8}$$

where s is a shape factor which becomes much less than one as the pore structure becomes more irregular.[10]

Our previous results (Johnson,[2] and Tserkovnyak and Johnson[3]) were specific to the case $W \equiv 0$. Here, we treat W as a real-valued, frequency-independent, phenomenological parameter of the theory, to be measured independently in a given context. Nagy and Blaho[10] have done just that for various porous media. See also Ref. 11. Furthermore, Nagy[12] successfully used Eq. (7) to analyze his data on surface acoustic modes of fully saturated porous media. In a somewhat different context, Eq. (7) is used by Liu and Johnson[13] to describe the effect of an elastic mudcake on tube waves in permeable formations. The approximate validity of Eq. (7) is established on firm theoretical and experimental grounds. It is the essential purpose of the present article to investigate the effects of a nonzero value for W on the acoustic properties of patchy-saturated samples.

Equation (7) can be combined with Eq. (3) to generalize the latter to hold at the interfaces as well as inside the patches:

$$\nabla p_f + W\phi[\hat{\mathbf{n}}\cdot(\mathbf{U}-\mathbf{u})]\hat{\mathbf{n}}\delta(R) = \frac{i\omega\phi\eta}{k}(\mathbf{U}-\mathbf{u}), \tag{9}$$

where $R(\mathbf{r})$ is the shortest distance from a given point $\mathbf{r}$ to the patch boundary and $\delta(R)$ is the Dirac's δ function. In addition, we assume the no net flow condition over the surface of the sample:

$$\int dS[\mathbf{U}-\mathbf{u}]\cdot\hat{\mathbf{n}} = 0, \tag{10}$$

which is valid either for sealed outer boundary or for the periodic boundary conditions. (The integration is taken either over the surface of the sample, for the sealed-pore boundary condition, or over a unit cell, for the periodic boundary conditions.) Equation (10) can be converted to a volume integral using Gauss' theorem:

$$\langle \epsilon_{ll} \rangle = \langle E_{ll} \rangle. \tag{11}$$

To summarize this section, the present theory supercedes that of Ref. 2 in that the assumption of vanishing capillary forces is now relaxed. The capillarity is accounted for by considering the pressure drop across interfaces which is governed by membrane stiffness, see Eq. (7). The theory of Ref. 2 is thus recovered in the limit when $W \to 0$.

III. HIGH- AND LOW-FREQUENCY LIMITS

A. Static regime

In the static limit, it is possible to get an exact closed-form analytic expression for the compressive modulus, K_0, defined by Eqs. (2)–(7). Hill[14] has developed a well-known exact result for the bulk modulus of a (nonporous) composite in which the shear modulus is everywhere constant, though the bulk modulus varies from point to point. See also Ref. 15. We extend that theory to the present case by looking for solutions of the form

$$u_i = \alpha x_i + \chi_{,i}, \quad U_i = \alpha x_i + \psi_{,i}. \tag{12}$$

It is understood that χ as well as the normal displacement $\chi_{,i}n_i$, are continuous across the patch boundary; similarly for ψ and $\psi_{,i}n_i$. Inasmuch as the first terms in Eq. (12) are special cases of the second terms, one is free to choose any value for α, including zero. It is slightly convenient, however, to choose

$$3\alpha = \langle \epsilon_{ll} \rangle \equiv \langle E_{ll} \rangle. \tag{13}$$

Equation (4) can now be written, within any patch, as

$$\tau_{ij,j} = [(P+Q)\chi_{,jj} + (R+Q)\psi_{,jj}]_{,i} \overset{\text{set}}{=} 0. \tag{14}$$

This equation is satisfied if the Laplacians of χ and of ψ are constant within each patch:

$$\chi_{,jj} = \begin{cases} a_1, & \mathbf{r} \in \text{patch } 1 \\ a_2, & \mathbf{r} \in \text{patch } 2 \end{cases}, \tag{15}$$

$$\psi_{,jj} = \begin{cases} b_1, & \mathbf{r} \in \text{patch } 1 \\ b_2, & \mathbf{r} \in \text{patch } 2 \end{cases}. \tag{16}$$

The four constants, $\{a_1, a_2, b_1, b_2\}$, are determined by various ancillary conditions. First, the volumetric strains within the jth patch may be written as

$$\epsilon_{ll}^{(j)} = 3\alpha + a_j, \quad E_{ll}^{(j)} = 3\alpha + b_j. \tag{17}$$

Therefore, Eq. (13) gives two conditions on the constants:

$$S_1 a_1 + S_2 a_2 = 0, \quad S_1 b_1 + S_2 b_2 = 0. \tag{18}$$

Next, there is the condition that the traction, $\boldsymbol{\tau}\cdot\hat{\mathbf{n}}$, is continuous across the boundary between patches. Hill[14] has pointed out an identity satisfied by the discontinuity across the patch boundary of the second derivatives of χ which can be rewritten for our purposes as

$$[\chi_{,ij}^{(1)} - \chi_{,ij}^{(2)}]n_j = (a_1 - a_2)n_i, \tag{19}$$

where the superscript on $\chi^{(k)}$ refers to the kth patch. Therefore, the discontinuity in the traction is

$$\begin{aligned}[\tau_{ij}^{(1)} - \tau_{ij}^{(2)}]n_j = \{&[(P_1+Q_1-2N)(3\alpha+a_1) \\ &+ (R_1+Q_1)(3\alpha+b_1)] \\ &- [(P_2+Q_2-2N)(3\alpha+a_2) \\ &+ (R_2+Q_2)(3\alpha+b_2)] + 2N(a_1-a_2)\}n_i.\end{aligned} \tag{20}$$

(P_k, Q_k, R_k refer to P, Q, R evaluated with respect to the saturating fluid in patch k.) In order for Eq. (20) to vanish across the patch interface, it is necessary that the quantity in curly braces vanishes:

$$\begin{aligned}&(P_1+Q_1)(3\alpha+a_1) + (R_1+Q_1)(3\alpha+b_1) \\ &\quad = (P_2+Q_2)(3\alpha+a_2) + (R_2+Q_2)(3\alpha+b_2).\end{aligned} \tag{21}$$

This, then, is the third condition on $\{a_1, a_2, b_1, b_2\}$.

The fourth and final condition follows from Eq. (7). The pore pressure within the kth patch is given by Eqs. (5) and (17):

$$p_f^{(k)} = -\frac{1}{\phi}[R_k(3\alpha + b_k) + Q_k(3\alpha + a_k)]. \tag{22}$$

We see that the pore pressure is constant within each patch, as required by Eq. (3) in the static limit. The left-hand side of the boundary condition (7) is constant over the interface between the two patches. Consider the case of a finite volume with sealed boundaries. Integrating Eq. (7) over the surface between the two patches, one finds

$$[p_f^{(1)} - p_f^{(2)}]A = W\phi \int_A dA(\mathbf{U} - \mathbf{u})\cdot\hat{\mathbf{n}}, \tag{23}$$

where A is the area of the interface between the patches. This integration can be extended over the entire bounding surface of fluid 1, say, because there is no flow across the outer boundary:

$$[p_f^{(1)} - p_f^{(2)}]A = W\phi \int_{A_1} dA(\mathbf{U} - \mathbf{u})\cdot\hat{\mathbf{n}}. \tag{24}$$

Now one may use Gauss' theorem:

$$[p_f^{(1)} - p_f^{(2)}]A = W\phi \int_{V_1} dV(E_{ll} - \epsilon_{ll}) = W\phi S_1 V[E_{ll}^{(1)} - \epsilon_{ll}^{(1)}]. \tag{25}$$

The volumetric strains and the pore pressure can be related to the unknown constants by means of Eqs. (17) and (22), respectively. The result for Eq. (25) is then

$$R_1(3\alpha + b_1) + Q_1(3\alpha + a_1) - [R_2(3\alpha + b_2) + Q_2(3\alpha + a_2)] = \tilde{W}(a_1 - b_1)S_1, \tag{26}$$

where

$$\tilde{W} = W\phi^2 \frac{V}{A} \tag{27}$$

in terms of the sample volume, V, and the bounding area between the patches, A. If the derivation is repeated for the case of periodic boundary conditions, Eq. (26) still holds; in that case, V is the volume of a unit cell and A is the interface area between the patches within the unit cell. In either case, if the derivation is repeated by closing the surface over fluid 2, instead of fluid 1, the right-hand side of Eq. (26) is replaced by $-\tilde{W}(a_2 - b_2)S_2$, which is equivalent via Eq. (18).

The four linear equations, (18), (21), and (26), can be inverted analytically to get the ratios $\{a_1, a_2, b_1, b_2\}/\alpha$; in practice, this is more easily done numerically. The effective static modulus is simply related to these constants. Following Ref. 14, we identify the external pressure, P_e, with the average compressive stress in the system:

$$P_e = -\tfrac{1}{3}\langle \tau_{ll} \rangle = -\tfrac{1}{3}[S_1 \tau_{ll}^{(1)} + S_2 \tau_{ll}^{(2)}], \tag{28}$$

where the compressive stress in the kth patch follows from Eq. (4):

$$\tau_{ll}^{(k)} = 3[(P_k + Q_k - \tfrac{4}{3}N)a_k + (R_k + Q_k)b_k + 3K_{\mathrm{BG}}^{(k)}\alpha]. \tag{29}$$

Here, $K_{\mathrm{BG}}^{(k)} = P_k + 2Q_k + R_k - (4/3)N$ is the Biot–Gassmann modulus for the kth patch. The effective static modulus for the system, $K_0 = -P_e/\langle \epsilon_{ll} \rangle$, is thus given by

$$K_0 = \frac{1}{3}\left\{ S_1\left[\left(P_1 + Q_1 - \frac{4}{3}N\right)\frac{a_1}{\alpha} + (R_1 + Q_1)\frac{b_1}{\alpha} + 3K_{\mathrm{BG}}^{(1)}\right] + S_2\left[\left(P_2 + Q_2 - \frac{4}{3}N\right)\frac{a_2}{\alpha} + (R_2 + Q_2)\frac{b_2}{\alpha} + 3K_{\mathrm{BG}}^{(2)}\right]\right\}. \tag{30}$$

Within the context of the model, Eq. (30) is exact. It depends upon the usual Biot parameters, the saturation values, the membrane stiffness, W, and it also depends upon the surface to volume ratio of the patch interface, A/V. There are three interesting limiting cases of Eq. (30): (1) If the two fluids have identical bulk moduli, then

$$\lim_{K_f^{(1)} = K_f^{(2)}} K_0 = K_{\mathrm{BG}}^{(1)} = K_{\mathrm{BG}}^{(2)}, \tag{31}$$

independent of the value of the membrane stiffness parameter, W. When the sample is compressed, the pore pressure rises equally in the two patches with the result that there is no relative flow between fluid and solid. (2) If the membrane stiffness is large enough, then each fluid is locked within its own patch and the static modulus is equal to the Biot–Gassmann–Hill result:

$$\lim_{W \to \infty} K_0 = K_{\mathrm{BGH}}, \tag{32}$$

where

$$\frac{1}{K_{\mathrm{BGH}} + (4/3)N} = \frac{S_1}{K_{\mathrm{BG}}^{(1)} + (4/3)N} + \frac{S_2}{K_{\mathrm{BG}}^{(2)} + (4/3)N}. \tag{33}$$

(3) If the membrane stiffness is zero, then the static modulus is given by the Biot–Gassmann–Woods result as was proved previously:[2]

$$\lim_{W \to 0} K_0 = K_{\mathrm{BGW}}, \tag{34}$$

where $K_{\mathrm{BGW}} = K_{\mathrm{BG}}(K_f^{\mathrm{eff}})$ and K_f^{eff} is an effective fluid modulus given by the Wood's law expression:

$$\frac{1}{K_f^{\mathrm{eff}}} = \frac{S_1}{K_f^{(1)}} + \frac{S_2}{K_f^{(2)}}. \tag{35}$$

In Fig. 1 we show the behavior of K_0 as a function of $\tilde{W}$. The parameter set[16] is based on that of Ref. 2. The sample is saturated with water, $S_w = 0.7$, and with air, $S_a = 0.3$. If the membrane stiffness is small, then Eq. (34) holds whilst if it is large, Eq. (32) is valid. The crossover occurs when $\tilde{W}$ is approximately equal to the modulus of the stiffer fluid, water. In Fig. 2 we show the saturation dependence of the static modulus, K_0, as well as that of K_{BGW} and K_{BGH}. Even when the membrane-stiffness parameter, W, is independent of the saturation, the parameter $\tilde{W}$ will in general have a saturation dependence reflecting the change in the interfacial

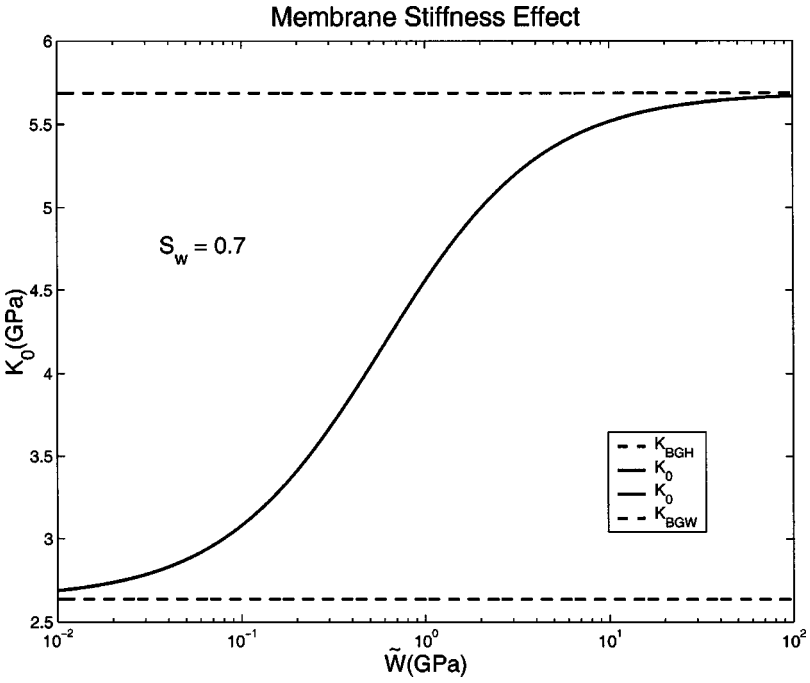


FIG. 1. Dependence of the static modulus on $\tilde{W}$ for a sample saturated with water, $S_w=0.7$ and air, $S_a=0.3$. Parameters are listed in Ref. 16. The two results for the static modulus, Eqs. (30) and (52), are seen to be nearly identical in this case in which one of the fluids is gas. $K_{\mathrm{BGH}}>K_{\mathrm{BGW}}$.

area. Nonetheless, for this plot, $\tilde{W}$ is held constant at a value $\tilde{W}=1$ GPa.

B. Low-frequency limit

Here we wish to find the behavior of $\tilde{K}(\omega)$ as it approaches K_0. Using the method developed in Ref. 2, we consider two equivalent expressions for the average energy dissipation per cycle. The macroscopic expression for the dissipation power averaged over one cycle is

$$\overline{P}=\frac{1}{2}\mathrm{Real}\int dS\,\frac{\partial \mathbf{u}^*}{\partial t}\cdot\boldsymbol{\tau}\cdot\hat{\mathbf{n}}=-\frac{1}{2}\mathrm{Real}\left[i\omega V\frac{|P_e|^2}{\tilde{K}(\omega)}\right], \quad (36)$$

where the integral is taken over the bounding surface of the sample. The microscopic expression is

$$\overline{P}=-\frac{1}{2}\mathrm{Real}\int dV\phi\left(\frac{\partial \mathbf{U}^*}{\partial t}-\frac{\partial \mathbf{u}^*}{\partial t}\right)\cdot\boldsymbol{\nabla} p_f. \quad (37)$$

In Eq. (37) the integration is taken over the volumes of each patch, specifically excluding the bounding surfaces of the patches; $\boldsymbol{\nabla} p_f$ has a δ-function contribution on the patch surface, see Eq. (9). One may then substitute Eq. (3) into Eq. (37) to arrive at the following microscopic expression for the dissipation:

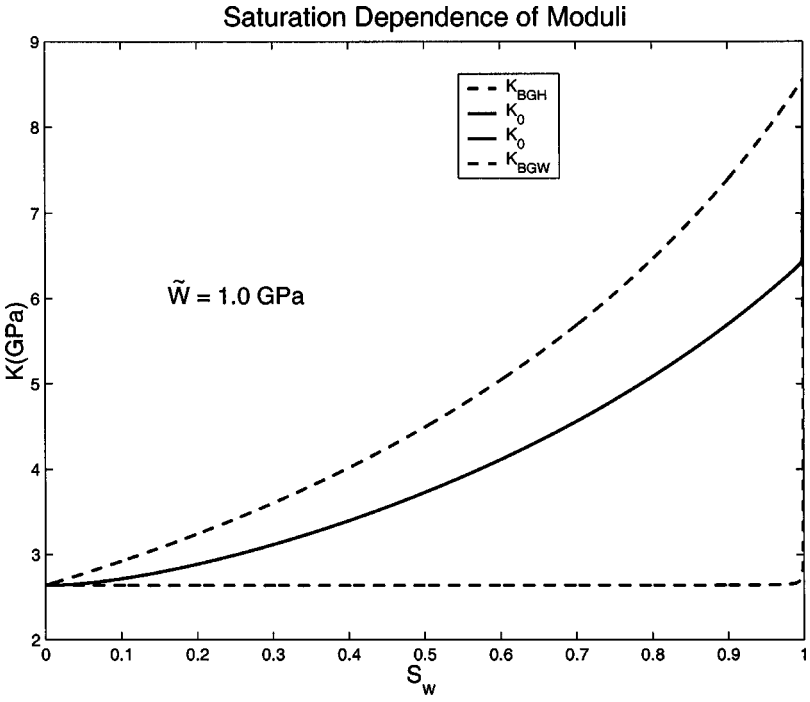


FIG. 2. The static moduli as a function of the water saturation, S_w. Parameter set as in Fig. 1. $\tilde{W}=1$ GPa is held constant. Again, K_0 is calculated in two different ways. $K_{\mathrm{BGH}}\geqslant K_{\mathrm{BGW}}$.

$$\overline{P} = -\frac{1}{2}\mathrm{Real}\int dV \phi i\omega(\mathbf{U}^* - \mathbf{u}^*)\cdot\left[\frac{i\omega\phi\eta}{k}(\mathbf{U}-\mathbf{u})\right]$$
$$= \frac{\omega^2\phi^2}{2k}\int dV\eta(\mathbf{r})|\mathbf{U}(\mathbf{r})-\mathbf{u}(\mathbf{r})|^2. \quad (38)$$

In this expression, the integrand does not contain any δ-function contributions. It is immediately clear that, as expected, the low-frequency behavior of $\overline{P}$ is quadratic in frequency. Therefore, the low-frequency expansion of $\tilde{K}(\omega)$ is given by

$$\lim_{\omega\to 0}\tilde{K}(\omega) = K_0[1 - i\omega T + O(i\omega)^2], \quad (39)$$

where the parameter T is determined by substituting Eq. (39) into Eq. (36) and equating it to Eq. (38):

$$T = \frac{\phi^2 K_0}{kV|P_e|^2}\int dV\eta(\mathbf{r})[\mathbf{U}^{(0)}(\mathbf{r}) - \mathbf{u}^{(0)}(\mathbf{r})]^2, \quad (40)$$

which is expressed in terms of the static displacements, $\mathbf{U}^{(0)}$ and $\mathbf{u}^{(0)}$.

We know from Eq. (12) that the displacements in the static limit are given by the gradient of certain functions. Therefore, we may define an auxiliary function, Φ, by

$$\nabla\Phi = -\eta(\mathbf{r})\frac{\mathbf{U}^{(0)} - \mathbf{u}^{(0)}}{P_e}. \quad (41)$$

Within each patch $\Phi = \eta[\chi - \psi]/P_e + \text{const}$. The function Φ is continuous across a patch boundary (otherwise there would be a δ-function contribution to the displacements); the normal component of the vector $-\nabla\Phi/\eta(\mathbf{r})$ is also continuous; and Φ is either periodic or $\nabla\Phi = 0$ (in the case of the sealed-pore boundary condition) on the outer boundary. The function Φ satisfies the differential equation

$$\nabla\cdot\left\{\frac{-1}{\eta(\mathbf{r})}\nabla\Phi\right\} = \frac{E_{ll}^{(0)}(\mathbf{r}) - \epsilon_{ll}^{(0)}(\mathbf{r})}{P_e} \overset{\text{def}}{=} g(\mathbf{r}). \quad (42)$$

This equation is analogous to an electrical conductivity problem in which $\Phi(\mathbf{r})$ plays the role of a potential, $\eta(\mathbf{r})$ is the local resistivity, and $g(\mathbf{r})$ is the distributed current source. Electrical neutrality is guaranteed by Eq. (11): $\int dV g(\mathbf{r}) = 0$.

In terms of Φ, Eq. (40) may be written as

$$T = \frac{\phi^2 K_0}{kV}\int dV\frac{1}{\eta(\mathbf{r})}(\nabla\Phi)^2, \quad (43)$$

which may be integrated by parts:

$$T = \frac{\phi^2 K_0}{kV}\int dV g(\mathbf{r})\Phi(\mathbf{r}). \quad (44)$$

(The surface term vanishes identically for either the sealed or the periodic boundary condition.) The source function, $g(\mathbf{r})$, is constant within each patch but in general is discontinuous across the patch interfaces. From Eqs. (12), (15), and (17) one has

$$g(\mathbf{r}) = \begin{cases} (b_1 - a_1)/P_e, & \mathbf{r}\in\text{patch } 1 \\ (b_2 - a_2)/P_e, & \mathbf{r}\in\text{patch } 2 \end{cases}. \quad (45)$$

Thus, the determination of the parameter T which governs the low-frequency approach to the static limit, Eq. (39), requires solving the differential equation (42) in which the piecewise constant sources, Eq. (45), are known by virtue of the solution for the static regime, and imposing one of our two no net-flow boundary conditions. This situation is only slightly different from that encountered when the surface-membrane effect is ignored altogether.[2]

C. High-frequency limit

Here we consider frequencies that are high in the sense that the pore pressure does not have time to equilibrate from patch to patch but we are still in the regime where $\omega \ll \omega_B$ and $\omega \ll \omega_x$. In this high frequency limit, $\omega\to\infty$, the fluids within each patch are locked with the solid frame, $\epsilon_{ll} = E_{ll}$. This conclusion is independent of the interface-stiffness parameter, W. We have a situation which satisfies the assumptions of Ref. 14: The shear modulus is a global constant and the bulk modulus is constant within each patch. Hence,

$$K_\infty \overset{\text{def}}{=} \lim_{\omega\to\infty}\tilde{K}(\omega) = K_{\mathrm{BGH}} \quad (46)$$

for any value of W, where K_{BGH} is given by Eq. (33). Recall that we found the same result for the static compressive modulus, K_0, when the patch boundary is locked by the large capillary tension, $W\to\infty$, see Eq. (32).

For a large but finite frequency, the correction to K_∞ is provided by the diffusive modes at the patch boundary.[2] In terms of a local coordinate x normal to the interface, the pore pressure near the boundary, $x = 0$, is given by

$$p_f = \begin{cases} p_{f1} + A_1 e^{-iq_1 x}, & x < 0 \quad (\text{patch } 1) \\ p_{f2} + A_2 e^{iq_2 x}, & x > 0 \quad (\text{patch } 2) \end{cases}. \quad (47)$$

Here, p_{fk} is the pore pressure inside the kth patch in the limit of infinite frequency and

$$q_k = \sqrt{i\omega/D_k} \quad (48)$$

is the slow-mode wave vector in terms of the diffusion coefficient D_k; see Appendix A. In the case of a finite membrane stiffness, $p_f(\mathbf{r})$ is in general discontinuous across the patch boundary. We can, nevertheless, show (see Appendix A) that the pore pressure discontinuity

$$\delta p_f \overset{\text{def}}{=} |p_f^{(1)} - p_f^{(2)}| = O(\omega^{-1/2}) \ll \Delta p_f \overset{\text{def}}{=} |p_{f1} - p_{f2}| = O(\omega^0),$$

for large frequencies, and thus the leading correction to K_∞ can be found by setting $\delta p_f = 0$, i.e., assuming continuity of the pressure, Eq. (47). Requiring, in addition, continuity of the fluid flux, we can then find coefficients $A_{1,2}$ in Eq. (47) and reproduce the result of Ref. 2:

$$\lim_{\omega\to\infty}\tilde{K}(\omega) = K_{\mathrm{BGH}}[1 - (i\omega)^{-1/2}G + \cdots], \quad (49)$$

where

$$G = \frac{kK_{\mathrm{BGH}}}{\eta_1\sqrt{D_1} + \eta_2\sqrt{D_2}}\left(\frac{\Delta p_f}{P_e}\right)^2\frac{A}{V} \quad (50)$$

in terms of the would-be drop in the pore pressure, Δp_f, relative to the applied external stress, P_e:

$$\frac{\Delta p_f}{P_e}=\frac{(R_2+Q_2)[K_1+(4/3)N]-(R_1+Q_1)[K_2+(4/3)N]}{\phi S_1 K_1[K_2+(4/3)N]+\phi S_2 K_2[K_1+(4/3)N]}. \tag{51}$$

Notice that the membrane-stiffness parameter does not enter any of Eqs. (49)–(51), so that the capillary forces not only have vanishing contribution to the high-frequency compressive modulus (46) but also to the leading correction to it, Eq. (49).

IV. LIQUID AND VACUUM

In Sec. III we have seen that whereas the two terms in the high-frequency limit, Eq. (49), are identical to those derived earlier when $W=0$, the two terms which describe the low-frequency limit, Eq. (39), are now considerably more complicated. The prescription for finding the static bulk modulus, K_0 [Eq. (30)], and the coefficient of the leading low-frequency correction, T [Eq. (44)], follows from the solution of the system of equations for parameters $\{a_1,a_2,b_1,b_2\}$; no simple analytic form is deduced and no transparent physical interpretation is readily available. We therefore offer an alternative derivation, which is both simple and physically appealing, albeit valid only for the case when one of the two fluids is a gas taken here to be the vacuum, and the other is a liquid whose properties are similar to water or oil. (A gas at standard temperature and pressure is four orders of magnitude more compressible than water or oil and so the approximation is a good one except for gas saturations less than $\sim 10^{-4}$.)

As in Sec. III, the starting point is Hill's theory:[14] If we denote the bulk modulus of the patch saturated with the liquid by K', then the static modulus of the composite is given by

$$\frac{1}{K_0+(4/3)N}=\frac{S}{K'+(4/3)N}+\frac{1-S}{K_b+(4/3)N}, \tag{52}$$

where S is the fluid saturation and K_b is the bulk modulus of the dry frame, as before. K' can be found by extending the Biot theory: Since the liquid does not bear any shear stresses, the static compressive modulus of the liquid-saturated patch must be given by the Biot–Gassmann formula,

$$K'=K_{\mathrm{BG}}(K_f'), \tag{53}$$

for some effective fluid bulk modulus, K_f'. This effective modulus characterizes the ''effort of squeezing'' the liquid out of the pore space. In particular, two limiting cases are obvious. (1) K_f' vanishes with the membrane stiffness:

$$\lim_{W\to 0} K_f'=0 \tag{54}$$

and (2) it reduces to the bare fluid modulus for infinitely large stiffness parameter:

$$\lim_{W\to\infty} K_f'=K_f. \tag{55}$$

If by inducing the pore pressure p_f, the fluid-filled pore-space volume is reduced by δV, then K_f' is defined by

$$\frac{1}{K_f'}=\frac{1}{p_f}\frac{\delta V}{\phi SV}. \tag{56}$$

If the fluid volume squeezed out of the pore space into menisci formed at the patch surface is $\delta V'$, then

$$p_f A=W\delta V', \tag{57}$$

which follows from Eq. (24). Finally, we have

$$\frac{1}{K_f}=\frac{1}{p_f}\frac{\delta V-\delta V'}{\phi SV} \tag{58}$$

for the bare bulk modulus of the pore fluid. Combining Eqs. (56)–(58) we then arrive at

$$\frac{1}{K_f'}=\frac{1}{K_f}+\frac{A}{W\phi SV}, \tag{59}$$

which satisfies Eqs. (54) and (55) in the two limits given by Eqs. (54) and (55). As a consequence, in the limit $W\to 0$, $K'=K_b$ (which is a specific case of a general result holding for arbitrary two fluids: $K'=K_{\mathrm{BGW}}$), and for $W\to\infty$, $K'=K_{\mathrm{BGH}}$. Equations (52), (53), and (59) complete the derivation for the static limit. This result for the static modulus is also plotted in Figs. 1 and 2 where it is seen to be identical to that computed from the full formalism.

It is as straightforward to find the parameter T governing the low-frequency behavior. T is given by Eq. (44) after solving Eq. (42), when the function $g(\mathbf{r})$ is known. In the fluid-filled patch, the volumetric strain of the solid can be found using Hill's theory once again [see also Eq. (35) of Ref. 2]:

$$\epsilon_{ll}=-\frac{[K_b+(4/3)N]P_e}{K'K_b+(4/3)N[SK'+(1-S)K_b]}. \tag{60}$$

Using Eq. (5) and

$$p_f A=W\phi SV(E_{ll}-\epsilon_{ll}), \tag{61}$$

provided by Eq. (25), we then obtain

$$E_{ll}-\epsilon_{ll}=-\frac{Q+R}{\tilde{W}S+R}\epsilon_{ll}, \tag{62}$$

which together with Eq. (60) defines the value of g, Eq. (42), inside the fluid patch:

$$g=\left(\frac{Q+R}{\tilde{W}S+R}\right)\frac{[K_b+(4/3)N]}{K'K_b+(4/3)N[SK'+(1-S)K_b]}. \tag{63}$$

Here, quantities Q and R are evaluated using the bare fluid bulk modulus, K_f. Defining an auxiliary function $\tilde{\Phi}$ in the space occupied by the fluid patch, V_f, by

$$\nabla^2\tilde{\Phi}=-1,\quad \mathbf{r}\in V_f \tag{64}$$

subject to the patch boundary condition $\tilde{\Phi}=0$ as well as the usual no net-flow condition on the external surface of the sample, we find for the parameter T, Eq. (44),

$$T=\frac{\eta S\phi^2 K_0 g^2}{k}l_f^2=\frac{l_f^2}{D_T}, \tag{65}$$

where the coefficient D_T has the dimensions of a diffusion constant and is given by

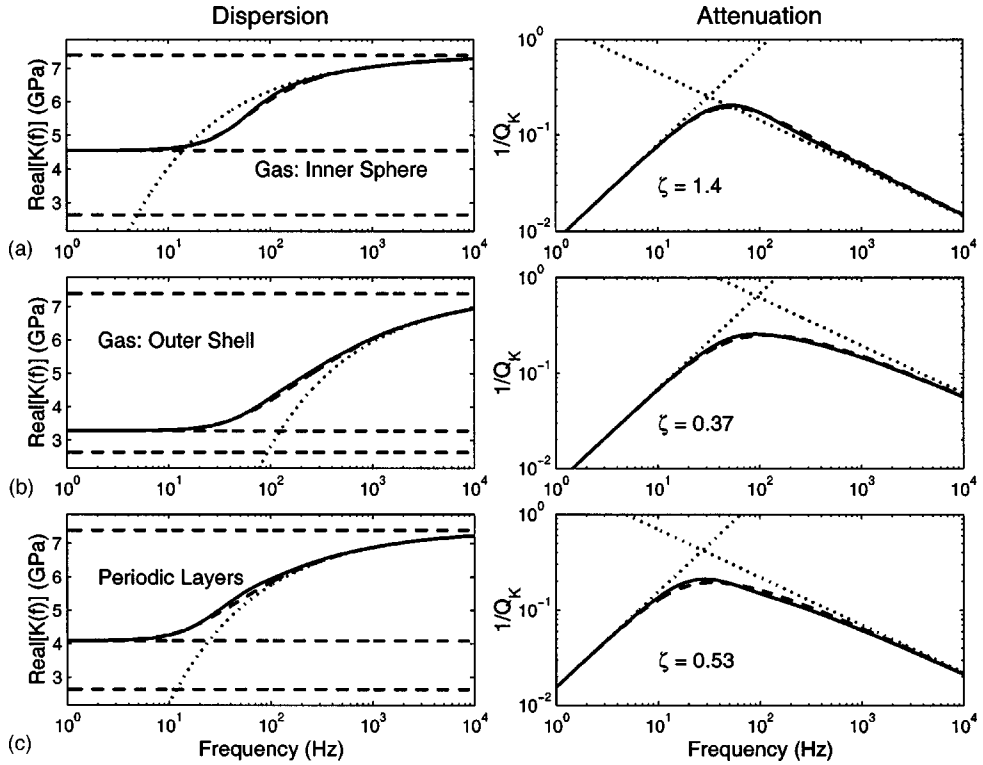


FIG. 3. Dispersion (Real$[\tilde{K}(\omega)]$) and attenuation $[1/Q_K = -\text{Imag}\,\tilde{K}(\omega)/\text{Real}\,\tilde{K}(\omega)]$ due to the patchy-saturation effect for an air/water combination, $S_a = 10\%$, in three different situations. Solid curves are numerically exact solutions, dotted curves are the high- and low-frequency limiting expressions, and dashed curves represent the analytic expression, Eq. (70). Horizontal dashed lines are (in order of decreasing value) K_{BGH}, $K_0(W = 30\ \text{GPa/m})$, and $K_0(W = 0)$. (a) Top row: The gas is located in an inner sphere, $R_a = 4.642$ cm, surrounded by a shell of water, $R_b = 10$ cm. (b) Middle row: The roles are reversed, $R_a = 9.655$ cm, $R_b = 10$ cm. (c) Bottom row: A periodic slab geometry, $L_a = 2$ cm, $L_w = 18$ cm. The other material parameters are listed in Ref. 16. In all cases, the membrane stiffness is $W = 30$ GPa/m. The values of ζ, Eq. (72), are shown.

$$D_T \overset{\text{def}}{=} \frac{k}{\eta S \phi^2 K_0 g^2}. \tag{66}$$

The Poisson size of the fluid patch, l_f, is defined by

$$l_f^2 \overset{\text{def}}{=} \frac{1}{V_f}\int_{V_f} dV \tilde{\Phi}. \tag{67}$$

The sole effect of the membrane stiffness, W, in this limit, is thus to renormalize the coefficient D_T from that which held when $W = 0$. We have verified numerically that the results derived in this section are identical to those presented in Sec. III when one of the fluids is taken to be a vacuum.

V. EXAMPLES

It is straightforward to repeat the calculations of Ref. 2 for $\tilde{K}(\omega)$ in which the patchy geometries are either that of periodic slabs or concentric spheres. In the slab geometry, region 1 is a layer of thickness L_1 and region 2 is a layer of thickness L_2, periodically repeated. Here, $S_1 = L_1/(L_1 + L_2)$ and $A/V = 2/(L_1 + L_2)$. In the concentric-spheres geometry, region 1 is a sphere of radius R_a surrounded by region 2 of outer radius R_b, so that $S_1 = (R_a/R_b)^3$ and $A/V = 3R_a^2/R_b^3$.

Equations (B.4) and (B.8) in Ref. 2 are still correct as written for the slab and the sphere geometries, respectively. However, the pore pressure is no longer continuous across the patch boundary but it obeys Eq. (7) instead, leading to a slightly different matrix equation to be solved. The results of such numerical calculations are presented in Fig. 3, using the parameter set of Ref. 16 at a water saturation $S_w = 0.9$. The horizontal dashed lines in the left column are K_{BGH}, K_0, and K_{BGW} from high to low, respectively. At low frequencies, we see that the numerically exact results do indeed approach K_0, as per Eq. (39). K_{BGW} has the same value in all three geometries, because the saturation values were chosen to be the same, and similarly for K_{BGH}. The values for K_0, however, are different in the three cases because the surface to volume ratios for the different patch geometries are different. The parameter $\tilde{W}$ in Eq. (26) has a different value in each case and so, therefore, do the values of K_0.

As we mentioned earlier, the low-frequency parameter T is determined by a differential equation (42), which is, in form, identical to that which existed when $W = 0$.[2] This means that the previously derived results for the slab and the sphere geometries are virtually unchanged. For the sphere, we have from Eq. (40) of Ref. 2

$$\begin{aligned} T = \frac{K_0 \phi^2}{30 k R_b^3}\{&[3\eta_2 g_2^2 + 5(\eta_1 - \eta_2) g_1 g_2 - 3\eta_1 g_1^2] R_a^5 \\ &- 15\eta_2 g_2 (g_2 - g_1) R_a^3 R_b^2 \\ &+ 5 g_2 [3\eta_2 g_2 - (2\eta_2 + \eta_1) g_1] R_a^2 R_b^3 - 3\eta_2 g_2^2 R_b^5\}. \end{aligned} \tag{68}$$

Similarly, the expression appropriate to the slab geometry is Eq. (41) of Ref. 2:

$$T=-\frac{K_0\phi^2}{6k(L_1+L_2)}\{\eta_1 g_1^2 L_1^3+3\eta_1 g_1 g_2 L_1^2 L_2 + 3\eta_2 g_1 g_2 L_1 L_2^2+\eta_2 g_2^2 L_2^3\}. \quad (69)$$

Of course, the correct value for the static modulus, K_0, Eq. (30), and the correct values for g_j, Eq. (45), must be used. Subject to this clarification, the low-frequency attenuation implied by the second term in Eq. (39) is shown as a dotted line; we see that the numerically computed results asymptote correctly in this limit. It was noted in Sec. III that the high-frequency asymptote for $\tilde{K}(\omega)$ is unaffected by a nonzero value for W. This limit is also plotted as dotted lines in Fig. 3.

VI. THEORETICAL MODEL FOR $\tilde{K}(\omega)$

A key observation by Johnson[2] was that all singularities and branch cuts of the bulk modulus $\tilde{K}(\omega)$ must lie on the negative imaginary axis, when analytic continuation into the complex plane is performed. This stems from the diffusive nature of the relaxation mechanism. In Appendix B, we show that this conclusion stays valid for a nonzero value for the membrane stiffness, leading us to suggest that the simple analytical formula proposed earlier[2] may also work in the present context. We wish to have a function which has singularities only on the negative imaginary axis and which has the limiting properties of Eqs. (39) and (49). That is, we propose the following function should work well as compared against exact numerical results:

$$\tilde{K}(\omega)=K_\infty-\frac{K_\infty-K_0}{1-\zeta(1-\sqrt{1-i\omega\tau/\zeta^2})}, \quad (70)$$

where consistency with Eqs. (39) and (49) requires

$$\tau=\left[\frac{K_\infty-K_0}{K_\infty G}\right]^2 \quad (71)$$

and

$$\zeta=\frac{(K_\infty-K_0)^3}{2K_0K_\infty^2TG^2}=\frac{(K_\infty-K_0)}{2K_0}\frac{\tau}{T}. \quad (72)$$

In Eq. (70) the branch cut in the definition of the square root function, $\sqrt{Z}$, is taken to be along the negative real z axis.

For the specific cases of the slab and concentric spheres geometries, the parameter T is computed directly from Eqs. (68) and (69) as the case may be. G is evaluated using Eq. (50). The resulting predictions from Eq. (70) are shown as dashed curves in Fig. 3, where one can see that it does an excellent job of interpolating throughout the entire frequency range.

VII. CONCLUSIONS

In summary, we have extended the theory of Johnson[2] to consider effects of the capillary forces on the acoustics of porous media saturated with two different fluids. Using a well-known expression for the pore-pressure discontinuity across the patch boundary, Eq. (7), we have shown that the finite membrane stiffness, W, does not affect the analytic structure of the frequency-dependent bulk modulus $\tilde{K}(\omega)$. In addition, the high-frequency asymptote of $\tilde{K}(\omega)$ does not depend on W, as long as W is sufficiently small [see Eqs. (A10) and (A11)], so that the high-frequency limit of our theory can be reached at all. The effect of the capillary forces is thus to rescale the static bulk modulus, K_0, and the low-frequency correction parameter, T, Eqs. (30) and (44), respectively. Furthermore, there are no new geometrical parameters introduced by the membrane stiffness: For example, when one of the fluids is a gas, the two relevant topological parameters are the sample-volume to patch-surface ratio and the Poisson size of the fluid patch, l_f, Eq. (67). In the limit $W\to 0$, results of Ref. 2 are recovered. In the limit $W\to\infty$, $\tilde{K}(\omega)\equiv K_{\mathrm{BGH}}$, Eq. (33), for all frequencies.

ACKNOWLEDGMENT

This work was supported in part by NSF Grant No. DMR 02-33773.

APPENDIX A: CAPILLARY CORRECTION IN THE HIGH-FREQUENCY LIMIT

In this appendix we show that the pore pressure discontinuity δp_f across the patch boundary, see Eq. (47), scales as $\omega^{-1/2}$ at high frequencies, $\omega\to\infty$, so that the membrane stiffness does not modify the high-frequency limit. The pressure equilibration, Eq. (47), is provided by the slow longitudinal waves on both sides of the patch boundary. In order to find equations of motion for the slow wave inside either patch, we take divergence of Eqs. (2) and (3) and use relations (4) and (5):

$$\nabla^2[(P_k+Q_k)\epsilon_{ll}^{(k)}+(R_k+Q_k)E_{ll}^{(k)}]=0, \quad (A1)$$

$$\nabla^2[Q_k\epsilon_{ll}^{(k)}+R_kE_{ll}^{(k)}]=\frac{i\omega\phi^2\eta_k}{k}[\epsilon_{ll}^{(k)}-E_{ll}^{(k)}]. \quad (A2)$$

Looking for solutions of the form

$$\epsilon_{ll}^{(k)}=c_1^{(k)}e^{-i(q_kx+\omega t)},$$
$$E_{ll}^{(k)}=c_2^{(k)}e^{-i(q_kx+\omega t)}, \quad (A3)$$

we obtain from Eq. (A1) the ratio of the amplitudes of the solid and fluid phases

$$\frac{c_1^{(k)}}{c_2^{(k)}}=-\frac{R_k+Q_k}{P_k+Q_k} \quad (A4)$$

and then from Eq. (A2) the diffusion coefficient

$$D_k\overset{\mathrm{def}}{=}\frac{i\omega}{q_k^2}=\frac{k}{\phi^2\eta_k}\frac{P_kR_k-Q_k^2}{P_k+2Q_k+R_k}. \quad (A5)$$

The above equations are valid on both sides of the patch boundary. The coefficients of pressure correction A_k, Eq. (47), at the interface are then given by

$$A_k=-\frac{1}{\phi}[Q_kc_1^{(k)}+R_kc_2^{(k)}], \quad (A6)$$

where we used Eq. (5) for the additional pressure exerted by the slow waves. The pressure drop δp_f across the interface is

related to the relative fluid displacement through Eq. (7), and using Eq. (3) we thus obtain

$$\delta p_f = (p_{f1} + A_1) - (p_{f2} + A_2) = -\frac{q_1 k W}{\eta_1 \omega} A_1. \tag{A7}$$

Finally, we require the fluid-flux continuity across the patch boundary:

$$\frac{q_1}{\eta_1} A_1 = -\frac{q_2}{\eta_2} A_2. \tag{A8}$$

Equations (A4), and (A6)–(A8) now constitute a linear system of six equations with six unknowns: $c_1^{(k)}$, $c_2^{(k)}$, and A_k. The resulting expression for δp_f, Eq. (A7), simplifies considerably if one patch is filled with a vacuum. In this case, the pressure discontinuity at the fluid boundary, relative to the difference in pore pressure relatively far from the interface, is given by

$$\frac{\delta p_f}{\Delta p_f} = \frac{1}{1 + \eta\omega/Wkq} = \frac{1}{1 + \sqrt{\omega/i\omega_W}}, \tag{A9}$$

where the crossover frequency

$$\omega_W \stackrel{\text{def}}{=} \frac{W^2 k^2}{\eta^2 D} \tag{A10}$$

is defined in terms of quantities corresponding to the fluid patch. When $\omega \gg \omega_W$, $\delta p_f \propto \omega^{-1/2}$ and the pressure drop across the interface, δp_f, is negligible in comparison with the pressure difference inside the patches, Δp_f, Eq. (51).

For consistency of our theory, we require that this limit can be reached while we are still in the low-frequency regime of the Biot theory and the propagatory modes have wavelengths longer than the typical patch dimensions, i.e., we need

$$\omega_W \ll (\omega_B, \omega_x). \tag{A11}$$

Equations (A10) and (A11) imply that our high-frequency limit is valid only for sufficiently small membrane-stiffness parameter W. With the parameter set used to generate Fig. 3 we have $\omega_W/(2\pi) = 4.5\,\text{Hz}$. However, our theory makes no assumption about the value of ω_W as compared against the diffusive crossover frequency and, indeed, the theory works quite well even if W is so large that ω_W is greater than the crossover implied by the intersection of the high- and low-frequency asymptotes.

APPENDIX B: ANALYTIC STRUCTURE OF $\tilde{K}(\omega)$

Here we generalize Appendix C of Ref. 2 to take into account the additional term due to the finite membrane stiffness W in Eq. (9). We want to investigate analytic properties of the real-valued causal response function $\hat{K}(t)$ in the frequency domain:

$$\tilde{K}(\omega) \stackrel{\text{def}}{=} \int_0^\infty \hat{K}(t) e^{i\omega t} dt. \tag{B1}$$

It is clear that $\tilde{K}(\omega)$ is analytic everywhere in the upper-half complex ω plane, and we will further show that all the zeros, singularities, and branch cuts lie on the negative imaginary axis.

In the following we consider an eigenvalue problem in which all quantities vary as $\mathbf{u}(\mathbf{r}) e^{-i\omega_n t}$. Assume

$$\int dS \mathbf{u} \cdot \boldsymbol{\tau}^* \cdot \hat{\mathbf{n}} = 0, \tag{B2}$$

as would be in the case of a zero, singularity, or branch point of $\tilde{K}(\omega)$.[2] We can then show using Eq. (2) and the Biot relations between stresses and strains, Eqs. (4) and (5), that $\int dV p_f \tau_{ll}^*$ is real-valued. This is true since all these equations are unaffected by the capillary forces, and we thus reproduce Eq. (C5) of Ref. 2:

$$\int dV p_f \tau_{ll}^* = \int dV \frac{1}{K_b/K_s - 1} \times \left[\frac{1}{3} |\tau_{ll}|^2 + \frac{3K_b}{2N} (\mathbf{D}\mathbf{D}^*)_{ll} \right], \tag{B3}$$

where $D_{ij} = \tau_{ij} - (1/3)\tau_{ll}\delta_{ij}$ is the deviatoric part of the stress tensor. Finally, multiplying Eq. (9) by $\mathbf{U}^* - \mathbf{u}^*$ and integrating it over the entire volume of the sample, we obtain

$$i\omega_n \int dV \frac{\eta\phi}{k} |\mathbf{U} - \mathbf{u}|^2 = \int dV p_f (\epsilon_{ll}^* - E_{ll}^*) + W\phi \int dV |\hat{\mathbf{n}} \cdot (\mathbf{U} - \mathbf{u})|^2 \delta(R). \tag{B4}$$

As the volumetric strains, ϵ_{ll} and E_{ll}, are given by linear combinations of p_f and τ_{ll} [via inverting Eqs. (4) and (5)], the right-hand side of Eq. (B4) is real-valued and, therefore, ω_n is pure imaginary. This completes the proof.

[1] 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).
[2] D. L. Johnson, "Theory of frequency dependent acoustics in patchy-saturated porous media," J. Acoust. Soc. Am. **110**, 682–694 (2001).
[3] Y. Tserkovnyak and D. L. Johnson, "Can one hear the shape of a saturation patch?" Geophys. Res. Lett. **29**, 12 (2002).
[4] T. Cadoret, "Effet de la saturation eau/gaz sur les propriétés acoustiques des roches; étude aux fréquences sonores et ultrasonores," Ph.D. dissertation, L'Université de Paris VII, Paris, France, 1993.
[5] T. Cadoret, D. Marion, and B. Zinszner, "Influence of frequency and fluid distribution on elastic wave velocities in partially saturated limestones," J. Geophys. Res. B [Solid earth] **100**, 9789–9803 (1995).
[6] T. Cadoret, G. Mavko, and B. Zinszner, "Fluid distribution effect on sonic attenuation in partially saturated limestones," Geophysics **63**, 154–160 (1998).
[7] 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**, 115–125 (1994).
[8] D. L. Johnson, "Recent developments in the acoustic properties of porous media," in *Frontiers in Physical Acoustics XCIII*, edited by D. Sette (North-Holland Elsevier, New York, 1986), pp. 255–290.
[9] J. Bear, *Dynamics of Fluids in Porous Media* (Dover, New York, 1972).
[10] P. B. Nagy and G. Blaho, "Experimental measurements of surface stiffness on water-saturated porous solids," J. Acoust. Soc. Am. **95**, 828–835 (1994).
[11] 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).
[12] P. B. Nagy, "Observation of a new surface mode on a fluid-saturated permeable solid," Appl. Phys. Lett. **60**, 2735–2737 (1992).

[13] H.-L. Liu and D. L. Johnson, ''Effects of an elastic membrane on tube waves in permeable formations,'' J. Acoust. Soc. Am. **101**, 3322–3329 (1997).
[14] R. Hill, ''Elastic properties of reinforced solids: Some theoretical principles,'' J. Mech. Phys. Solids **11**, 357–372 (1963).
[15] G. Milton, *The Theory of Composites* (Cambridge University Press, Cambridge, 2001).

[16] Values of input parameters for the calculation of patchy-saturation effects (from Ref. 2): $\phi=0.284$, $K_s=35$ GPa, $K_b=2.637$ GPa, $N=1.740$ GPa, $k=10^{-13}\,\mathrm{m}^2$, $K_f\,(\mathrm{water})=2.25$ GPa, $\eta\,(\mathrm{water})=10^{-3}$ Pa s, $K_f\,(\mathrm{air})=10^5$ Pa, $\eta\,(\mathrm{air})=10^{-5}$ Pa s, $\sigma\,(\mathrm{water-air})=7.3\times10^{-2}$ N/m. From Eq. (8) we can then find $W=30$ GPa/m using $s=0.04$ and from Eq. (27) $\tilde{W}=1$ GPa using $V/A=0.4$ m.

Investigation of the convergence of the mixed displacement-pressure formulation for three-dimensional poroelastic materials using hierarchical elements

S. Rigobert
Laboratoire des Sciences de l'Habitat, DGCB URA CNRS 1652, Ecole Nationale des Travaux Publics de l'Etat, 69518 Vaulx-en-Velin Cedex, France

N. Atalla
Groupe d'Acoustique de l'Université de Sherbrooke, Department of Mechanical Engineering, University de Sherbrooke, Sherbrooke, Quebec J1K 2R1, Canada

F. C. 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 1 October 2001; accepted for publication 12 August 2003)

Recently, a mixed pressure displacement {**u**, **P**} formulation based on Biot's poroelasticity equations has been presented for porous materials. This model leads to a reduction of the number of degrees of freedom required for the modeling of three-dimensional porous media in comparison to classical displacement–displacement {**u**, **U**} formulations. In this paper, an extension of the {**u**, **P**} formulation based on hierarchical elements is presented. First, a variant of the weak integral form of the {**u**, **P**} formulation is presented and its numerical implementation using hierarchical elements is detailed, together with the application of boundary and loading conditions. Numerical results are presented to show the accuracy and performance of the present approach. In particular, the importance of correctly capturing the coupling effects between the two phases is highlighted. [DOI: 10.1121/1.1616579]

PACS numbers: 43.20.Jr, 43.50.Gf, 43.55.Wk [VWS] Pages: 2607–2617

I. INTRODUCTION

The behavior of porous material is classically modeled using empirical formulas for locally reacting materials,[1] or with accurate one-dimensional (1D) analytical methods based on Biot poroelasticity theory.[2–5] At low frequencies, where the modal behavior of the system is important, several two-dimensional (2D) and three-dimensional (3D) finite elements models have been developed. The first models use, with different variations, the displacement fields in the solid and fluid phase of the porous material as variables.[6–10] They are referred to as {**u**, **U**} type formulations. While accurate, the {**u**, **U**} formulations lead to large frequency dependent systems and thus to cumbersome calculations for complex structures. To alleviate the problem, Atalla *et al.*[11] presented an exact mixed displacement-pressure {**u**, **P**} finite element formulation in three dimensions for a porous material, based on Biot's poroelastic equations. Debergue *et al.*[12] presented the boundary and coupling conditions for this new formulation. Atalla *et al.*' model presents the calculation of the response of the porous material in the form of a coupled fluid-structure problem. This formulation has been proven to give accurate results when compared to the {**u**, **U**} formulation. More importantly, by reducing the number of degrees of freedom involved (i.e., from six degrees of freedom per node to four) and by simplifying the coupling conditions between the porous medium and elastic or fluid media, the approach leads to important gains in memory and computation time requirements.

While the {**u**, **P**} formulation allows for a more efficient modeling to porous elastic media compared to the {**u**, **U**} formulation, yet both techniques suffer from important computational costs in practical applications. Indeed, because of the biphasic nature of poroelastic elements, mesh criterion used for elements describing monophasic media (solid, fluid) is *a priori* not valid for poroelastic elements. Using the {**u**, **U**} formulation, Dauchez *et al.*[13] showed that linear poroelastic elements verify convergence rate of linear monophasic elements, according to each type of Biot wave. Classical mesh criterion, i.e., six linear elements per wavelength, provides a necessary condition to get reliable results. However, for real 3D deformations, the classical criterion gives indications for a minimal mesh but is insufficient because of locking of 3D linear elements and discrepancies in the wavelengths of the displacement fields of the two phases. An important refinement of the mesh is necessary to get satisfactory results. In particular, indicators related to fluid motion have been found very sensitive. As a consequence, the minimal number of elements required is difficult to predict. By nature, convergence of poroelastic elements is rather slower than convergence of either equivalent solid or fluid elements, because of the presence of two different scale phenomena. The discrepancy can be lowered when the behavior of the porous material is dominated by the motion of one phase. As a consequence, this slow convergence leads to large computational costs. Debergue[14] found similar results using the {**u**, **P**} formulation. Several recent attempts to alleviate the computing cost of the poroelastic formulations, using linear elements, have been unsuccessful. For example,

Sgard *et al.*[15,16] have investigated the use of a selective modal analysis to decrease the size of large finite elements models involving poroelastic materials. The method used a dual uncoupled basis of undamped modes associated with the skeleton *in vacuo* and the fluid phase occupying the volume of the porous material to approximate the {**u**, **P**} variables. However, this method converges poorly due to the importance of the damping brought in by the porous material. The same authors also proposed a mixed wave-finite element approach dedicated to the modeling of porous materials.[17] Starting from the weak {**u**, **P**} formulation for a porous material, the fields for each phase of the porous material were expanded at each node of the finite element mesh in terms of a finite number of wave functions. These functions are chosen to be plane waves solution of Biot's poroelasticity equations. This approach gives accurate results for a single porous material and leads to a considerable reduction of the size of the system to be solved. However, this system is no longer sparse and its construction is time consuming. Another way to solve the computing cost problem, would be to design specific models dedicated to particular applications. For example, Dauchez *et al.*[18] pointed out that most of the extra-energy dissipation related to the foam coating of a plate was due to structural damping in the solid phase of the porous medium. Thus, the porous material could be reduced to its solid phase in such a configuration. However, this assumption can not be generalized, especially for a coating by a glasswool where viscous effects become quickly predominant as frequency increases.[19]

In order to solve the convergence problem, it is proposed in this paper to investigate the performance of the {**u**, **P**} formulation using high-order hierarchical elements. It is known that these elements allow for higher order polynomial representation and thus eliminate problems such as locking.[20–24] Because of the higher order of approximation of hierarchical elements, a more accurate representation of the different type of deformations and coarser meshes, compared to classical finite elements, can be used. Therefore, a significant reduction of the number of degrees of freedom required for a correct modeling is expected. In the first part of the paper, a variant of the {**u**, **P**} formulation is recalled and the theory of hierarchical elements in this context is introduced. In particular, the treatment of the boundary conditions and excitations is detailed. Validation results are then presented to show the accuracy of the present approach. The performance of the hierarchical poroelastic formulation in terms of the needed number of degrees of freedom is underlined, and can be increased by the choice of different interpolation orders for the basis functions in the solid and fluid phase. The interest of this latter feature is shown and guidelines for the convergence of hierarchical poroelastic elements are presented from the study of a particular material.

II. THEORY

This section concentrates on the forced response of a single isotropic porous material. Biot–Allard equations of poroelasticy are considered for the description of the porous medium. The displacement of the solid phase and the interstitial pressure are chosen as variables.

A. The mixed displacement-pressure formulation

The starting point of the modeling of the porous medium is the weak {**u**, **P**} formulation (Atalla *et al.*).[11] The mixed displacement-pressure formulation for a poroelastic material occupying a volume Ω_p reads

$$\begin{aligned}
&\int_{\Omega_p} [\tilde{\underline{\underline{\sigma}}}^S(\mathbf{u}) : \underline{\underline{\epsilon}}^S(\delta\mathbf{u}) - \tilde{\rho}\omega^2 \mathbf{u}\cdot\delta\mathbf{u}]\, d\Omega \\
&\quad + \int_{\Omega_p} \left[\frac{h^2}{\omega^2 \tilde{\rho}_{22}} \nabla P \cdot \nabla \delta P - \frac{h^2}{\tilde{R}} P\, \delta P \right] d\Omega \\
&\quad - \tilde{\gamma} \int_{\Omega_p} \delta(\nabla P \cdot \mathbf{u})\, d\Omega - \int_{\partial\Omega_p} h \left(1 + \frac{\tilde{Q}}{\tilde{R}} \right) \\
&\quad \times \delta(P u_n)\, dS - \int_{\partial\Omega_p} [\tilde{\underline{\underline{\sigma}}}^t(\mathbf{u}) \cdot \mathbf{n}] \cdot \delta\mathbf{u}\, dS \\
&\quad - \int_{\partial\Omega_p} h(U_n - u_n)\, \delta P = 0, \qquad (1)
\end{aligned}$$

where $\partial\Omega_p$ stands for the boundary of the poroelastic domain. **u** is the solid phase displacement, **U** and P are the displacement and the pressure of the fluid in the pores, respectively. $\tilde{\underline{\underline{\sigma}}}^S$ and $\underline{\underline{\epsilon}}^S$ are, respectively, the stress and strain tensor of the solid phase *in vacuo*. The total stress tensor in the porous material $\tilde{\underline{\underline{\sigma}}}^t$ is related to $\tilde{\underline{\underline{\sigma}}}^S$ by the relation $\tilde{\underline{\underline{\sigma}}}^S = \tilde{\underline{\underline{\sigma}}}^t - h[1 + (\tilde{Q}/\tilde{R})] P \underline{\underline{1}}$. $\tilde{\rho}$ and $\tilde{\rho}_{22}$ are the apparent density of the solid and the fluid phase, h is the porosity, $\tilde{\gamma}$ is a coupling factor between the two phases, $\tilde{Q}$ and $\tilde{R}$ are poroelastic coefficients. The expressions of the latter quantities can be found in Ref. 11.

Instead of using directly Eq. (1), a variant is presented here. The first surface integral in Eq. (1) can be turned into a volume integral using the second Green formula and given the following mathematical relation:

$$\nabla \cdot (a\mathbf{b}) = \nabla a \cdot \mathbf{b} + a \nabla \cdot (\mathbf{b}). \qquad (2)$$

Hence, assuming that the porous material is isotropic,

$$\begin{aligned}
I &= -\int_{\partial\Omega_p} h \left(1 + \frac{\tilde{Q}}{\tilde{R}} \right) \delta(u_n P)\, dS \\
&= -\int_{\Omega_p} \nabla \left(h \left(1 + \frac{\tilde{Q}}{\tilde{R}} \right) \delta(u_n P) \right) d\Omega \\
&= -h \left(1 + \frac{\tilde{Q}}{\tilde{R}} \right) \int_{\Omega_p} \delta(\mathbf{u} \cdot \nabla P)\, d\Omega \\
&\quad - h \left(1 + \frac{\tilde{Q}}{\tilde{R}} \right) \int_{\Omega_p} \delta(P \nabla \mathbf{u})\, d\Omega. \qquad (3)
\end{aligned}$$

Using Eq. (3) in Eq. (1) the following expression of the weak {**u**, **P**} formulation is obtained

$$\int_{\Omega_p} [\tilde{\underline{\underline{\sigma}}}^S(\mathbf{u}) : \tilde{\underline{\underline{\varepsilon}}}^S(\delta\mathbf{u}) - \tilde{\rho}\omega^2 \mathbf{u}\cdot\delta\mathbf{u}] d\Omega$$
$$+ \int_{\Omega_p} \left[\frac{h^2}{\omega^2 \tilde{\rho}_{22}} \nabla P \cdot \nabla \delta P - \frac{h^2}{\tilde{R}} P\, \delta P \right] d\Omega$$
$$- \left[\tilde{\gamma} + h\left(1 + \frac{\tilde{Q}}{\tilde{R}} \right) \right] \int_{\Omega_p} \delta(\nabla P \cdot \mathbf{u}) d\Omega$$
$$- h\left(1 + \frac{\tilde{Q}}{\tilde{R}} \right) \int_{\Omega_p} \delta(P \nabla \cdot \mathbf{u}) d\Omega$$
$$- \int_{\partial\Omega_p} [\tilde{\underline{\underline{\sigma}}}^t(\mathbf{u}) \cdot \mathbf{n}] \cdot \delta\mathbf{u} - \int_{\partial\Omega_p} h(U_n - u_n)\, \delta P\, dS = 0. \quad (4)$$

The weak form of the mixed pressure-displacement formulation [Eq. (4)] presents two interesting features. First, the symmetric coupling terms between the two phases are expressed in terms of volume integrals over the domain Ω_p. Second, the boundary integral terms involve the total stress tensor and the displacement flux at the boundary of the domain. Therefore, the application of coupling conditions with another domain, as well as boundary conditions and excitations, is achieved in a simpler manner compared to Eq. (1).[12]

B. Hierarchical elements

In the following, the approximation of Eq. (4) using 8-nodes parallelipipedic volume hierarchical elements is considered. The so-called ''blending function method'' is generally used to map this parent element into the geometrical element.[22] Any point of the discretized subdomain with global coordinates (x,y,z) is located on a parent element by a set of local coordinates (ξ,η,ζ). The relation between local and global coordinates is given by

$$\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \sum_{i=1}^{8} \begin{Bmatrix} X_i \\ Y_i \\ Z_i \end{Bmatrix} \cdot \mathcal{N}_i(\xi,\eta,\zeta) + \sum_{j=1}^{12} \begin{Bmatrix} l_j^x \\ l_j^y \\ l_j^z \end{Bmatrix} \cdot l_j^1 \cdot l_j^2. \quad (5)$$

In Eq. (5), $\langle X_i, Y_i, Z_i \rangle$ stands for the global coordinates of node i. The functions $\mathcal{N}_i$ are the classic linear basis functions used in finite elements and are related to node i on the parent element. In order to consider complex geometries for an element, additional functions related to the edge j are taken into account. These functions correspond to the term on the right-hand side of Eq. (5). l_j^1 and l_j^2 are linear functions depending on a single local coordinate (different for l_j^1 and l_j^2) ξ, η, or ζ. l_j^x, l_j^y, and l_j^z are functions depending on the third local coordinate and used for the description of edge j. Further details are given in Ref. 22.

The next step is the interpolation of the fields on the mesh provided for the subdomain. On a given element, the variable q is approximated by

$$q(\xi,\eta,\zeta) = \sum_i \mathcal{N}_i(\xi,\eta\zeta) q_i^{\mathrm{ph}} + \sum_j \mathcal{G}_j(\xi,\eta,\zeta) q_j^{\mathrm{gen}}. \quad (6)$$

In Eq. (6), $\mathcal{N}_i$ are called node modes and are identical to those used in Eq. (5). The associated amplitudes q_i^{ph} stand for the physical value of q at node i (the superscript ph means physical). Functions $\mathcal{G}_j$ are additional shape functions classified in three categories: edge modes, face modes, and internal modes. The associated generalized amplitudes q_j^{gen} (the superscript gen means generalized) do not have simple physical meaning. Like node modes, edge and face modes are associated to a geometrical entity of the mesh. For example, an edge mode on an element is associated with a particular edge j of this element. Its value is 0 on all the other edges of the element. As for internal modes, their value is 0 on all the faces of the element. By considering each category of modes explicitly, Eq. (6) rewrites

$$q(\xi,\eta,\zeta) = \sum_i \mathcal{N}_i(\xi,\eta\zeta) q_i^{\mathrm{ph}} + \sum_j \mathcal{E}_j(\xi,\eta,\zeta) q_j^{\mathrm{gen}} + \sum_k \mathcal{F}_k(\xi,\eta,\zeta) q_k^{\mathrm{gen}} + \sum_t \mathcal{I}_t(\xi,\eta,\zeta) q_t^{\mathrm{gen}}, \quad (7)$$

where $\mathcal{E}_j$ are edge modes, $\mathcal{F}_k$ are face modes, and $\mathcal{I}_t$ are internal modes. These functions are chosen to make complete polynomials of ascending order p. Namely, the basis functions of the hierarchical variables are constructed using Legendre polynomials. The number of basis functions depends on the interpolation order p. The selection process, detailed in Ref. 22, leads to the expression of the generalized shape functions; they are given in the Appendix for completeness. Note that the present approach considers a mapping of the parent element into the geometrical element, using only the functions $\mathcal{N}_i$ in Eq. (5).

C. Numerical implementation

As previously stated, in the present approach, the weak integral formulation given by Eq. (4) is discretized using 8-node volume elements. The displacement of the solid phase $\mathbf{u}$ and the pressure in the pores P are expanded in terms of node modes and hierarchical shape functions. Hence, the field variables in a given element are written in the following form:

$$\mathbf{u}^e = [N_s]\{u_n\}^e \quad \text{and} \quad p^e = [N_f]\{p_n\}^e, \quad (8)$$

where $[N_s]$ and $[N_f]$ are the interpolation matrices on the considered element e. $\{u_n\}^e$ and $\{p_n\}^e$ stand for the physical and generalized amplitudes associated with the displacement of the solid phase and the pressure in the pores, respectively.

1. Discretization of the weak form

Substituting Eq. (8) into Eq. (4), one gets

$$\int_{\Omega_p} \tilde{\underline{\underline{\sigma}}}^S(\mathbf{u}) : \tilde{\underline{\underline{\varepsilon}}}^S(\delta\mathbf{u}) d\Omega \Rightarrow \langle \delta u_n \rangle [K] \{u_n\}, \quad (9)$$

$$\int_{\Omega_p} \tilde{\rho}\mathbf{u}\cdot\delta\mathbf{u}\, d\Omega \Rightarrow \langle \delta u_n \rangle [\tilde{M}] \{u_n\}, \quad (10)$$

$$\left[\tilde{\gamma}+h\left(1+\frac{\tilde{Q}}{\tilde{R}.}\right)\right]\int_{\Omega_p}\delta\mathbf{u}\cdot\nabla P\,d\Omega\Rightarrow\langle\delta u_n\rangle[\tilde{C}_1]\{P_n\}, \tag{11}$$

$$h\left(1+\frac{\tilde{Q}}{\tilde{R}}\right)\int_{\Omega_p}P\nabla\cdot\delta\mathbf{u}\,d\Omega\Rightarrow\langle\delta u_n\rangle[\tilde{C}_2]\{P_n\}, \tag{12}$$

$$\frac{h^2}{\tilde{\rho}_{22}}\int_{\Omega_p}\nabla P\cdot\nabla\delta P\,d\Omega\Rightarrow\langle\delta P_n\rangle[\tilde{H}]\{P_n\}, \tag{13}$$

$$\frac{h^2}{\tilde{R}}\int_{\Omega_p}P\,\delta P\,d\Omega\Rightarrow\langle\delta P_n\rangle[\tilde{Q}]\{P_n\}, \tag{14}$$

where $\{u_n\}$ and $\{P_n\}$ represent the solid and fluid phase physical and generalized degrees of freedom of the whole mesh. $[K]$ and $[\tilde{M}]$ are, respectively, the stiffness and mass matrix associated to the solid phase *in vacuo*. $[\tilde{H}]$ and $[\tilde{Q}]$ are, respectively, the kinetic and compressional energy matrices of the fluid phase. $[\tilde{C}_1]$ and $[\tilde{C}_2]$ stand for the coupling between the two phases of the porous medium. The notation $[\tilde{\ }]$ indicates that the matrices are complex valued and frequency dependent. Note that compared to the implementation of the {**u**, **P**} formulation presented in Ref. 11, two new matrices have to be computed. In the theory of hierarchical elements, the basis functions on an element are built using Legendre polynomials. The computation of the matrices in Eqs. (9)–(14) is performed with the commonly used Gauss–Legendre integration scheme and enough integration points are chosen to ensure a correct approximation. This step is time consuming and is the main drawback of hierarchical elements.[24] To alleviate that problem, Hinnant[25] proposed the ''vector quadrature'' integration scheme. However, this optimization procedure is not used in the present approach.

Substituting Eqs. (9)–(14) in the weak formulation leads to the following system to be solved:

$$\begin{pmatrix}[K]-\omega^2[\tilde{M}] & -[\tilde{C}_1]-[\tilde{C}_2]\\ -\omega^2[\tilde{C}_1]^T-\omega^2[\tilde{C}_2]^T & [\tilde{H}]-\omega^2[\tilde{Q}]\end{pmatrix}\begin{Bmatrix}u_n\\ P_n\end{Bmatrix}=\begin{Bmatrix}F_s\\ F_p\end{Bmatrix}, \tag{15}$$

where the right-hand side of the equation denotes the loading vector for the porous material. The application of various loading conditions is described in the following paragraph. The coupled system given by Eq. (15) is similar to the one obtained in classical finite elements.

2. Application of boundary and loading conditions

The boundary conditions and the loading terms involved in the {**u**, **P**} formulation were presented by Debergue *et al.*[12] Their application in the context of the theory of hierarchical elements is detailed hereafter. In the present paper, three boundary conditions are considered. The porous material can be (i) bonded onto a rigid wall, (ii) guided, or (iii) free (i.e., not subject to any boundary condition). In case (i), the displacement vector of the solid phase is set to 0. In case (ii), the normal component to the interface of the solid phase displacement is set to 0. In case (iii), the fluid pressure is set to 0 on the interface (note that this condition is approximate and its accuracy is discussed by Debergue[12]). Therefore, for an element on the boundary, applying the boundary conditions mentioned above amounts to setting to 0 one or several components of the fields. This is achieved by constraining the amplitudes of the modes of the nodes, edges and faces lying on the boundary and related to the considered component of the fields. Practically, the coefficients of the matrices in Eqs. (9)–(14) relative to these degrees of freedom are not assembled. Note that, by construction, the value of internal modes are worth 0 on each face of the element and thus are not subject to the above mentioned conditions. In the configurations studied in this paper, two kinds of loads have been considered. First, the load is a rigid piston motion imposed on one side of the porous material. This leads to the following condition on the interface:

$$\mathbf{u}\cdot\mathbf{n}=u_0,\quad (u_n-U_n)=0, \tag{16}$$

where **n** is the normal to the interface, and u_0 the amplitude of the piston motion. Second, the porous medium is submitted to an acoustical excitation, which can be modeled as an imposed surface pressure of amplitude p_0. The condition at the interface is then

$$p=p_0,\quad \tilde{\underline{\sigma}}^t\cdot\mathbf{n}=-p_0\mathbf{n}. \tag{17}$$

To be concise, the two considered kinds of excitations lead to an imposed value q_0 of a component q of the fields in the porous material. Considering the expression of q given by Eq. (6) the following conditions must be imposed on the interface:

$$\{q^{\mathrm{ph}}\}=\{q_0\}\quad\text{and}\quad\{q^{\mathrm{gen}}\}=0. \tag{18}$$

Equation (18) means that the values of q have to be set to q_0 at the nodes on the interface. Besides, the amplitudes of the hierarchical shape functions related to the edges or faces located on the interface are set to 0. By definition, the amplitudes of the internal modes are null on the sides of an element. The conditions of imposed degrees of freedom expressed by Eq. (18) are taken into account in the discretized weak formulation using discrete Lagrange multipliers.

In addition, when a surface pressure with amplitude p_0 is imposed, Eq. (17) leads to an elementary force vector given by

$$\{F_s\}_{\mathrm{elem}}=-\int_{\partial\Omega p_{\mathrm{elem}}}p_0\mathbf{n}\cdot[N_s]dS, \tag{19}$$

where $[N_s]$ is the column vector containing the shape function related to the face if the considered element subject to the excitation. Note that only node modes, edge modes with interpolation order $p=2$, and face modes with interpolation order $p=4$ are taken into account in $[N_s]$. In fact, the other modes of analytical expression $\mathcal{G}$ verify the following relation by construction:

$$\int_{\partial\Omega p_{\mathrm{elem}}}\mathcal{G}\,dS=0. \tag{20}$$

The proof for this assertion is given in the Appendix. Note that for all the excitations considered in this paper, $\{F_p\}=0$ in Eq. (15).

3. Vibroacoustics indicators

In the result section, two vibroacoustic indicators are considered. For the solid phase of the porous material, the mean square velocity along the three directions is computed. Practically, the mean square velocity along direction i is given by

$$\langle v_i^2 \rangle = \frac{\omega^2}{2\Omega_p} \int_{\Omega_p} |v_i|^2 \, d\Omega = \frac{\omega^2}{2\Omega_p} \langle u_n^{\mathrm{ph}} \; u_n^{\mathrm{gen}} \rangle^* [M_i] \begin{Bmatrix} u_n^{\mathrm{ph}} \\ u_n^{\mathrm{gen}} \end{Bmatrix}, \tag{21}$$

where $\langle u_n^{\mathrm{ph}} \; u_n^{\mathrm{gen}} \rangle$ is the row vector containing the physical and generalized amplitudes relative to the solid phase displacement along direction i. $[M_i]$ contains the coefficients of a partition of the mass matrix $[\tilde{M}]$ divided by $\tilde{\rho}$ and related to the appropriate degrees of freedom. In Eq. (21), (*) means complex conjugate, and Ω_p is the volume of the poroelastic domain.

For the fluid phase, the mean square pressure is computed using the following formula:

$$\langle P^2 \rangle = \frac{1}{2\Omega_p} \int_{\Omega_p} |P|^2 \, d\Omega = \frac{1}{2\Omega_p} \langle P_n^{\mathrm{ph}} \; P_n^{\mathrm{gen}} \rangle^* [Q_1] \begin{Bmatrix} P_n^{\mathrm{ph}} \\ P_n^{\mathrm{gen}} \end{Bmatrix}, \tag{22}$$

where $\langle P_n^{\mathrm{ph}} \; P_n^{\mathrm{gen}} \rangle$ is the row vector of the physical and generalized amplitudes related to the pressure in the pores. $[Q_1]$ is the compression matrix of the fluid phase $[\tilde{Q}]$ divided by $h^2/\tilde{R}$.

III. NUMERICAL EXAMPLES

In the following, the validation of the $\{u,P\}$ formulation using hierarchical elements is numerically assessed. The present approach is compared to a finite element code developed at the Université de Sherbrooke and based on classical finite elements. This latter code has been validated both numerically and experimentally, elsewhere.[11,12] The accuracy of the results obtained by the use of hierarchical elements is underlined. The reduction of the number of degrees of freedom provided by the use of high order polynomials for the basis functions is highlighted, thus illustrating the performances of hierarchical poroelastic elements. An emphasis is made on this latter aspect by showing the benefit of using different interpolation orders for the basis functions of each phase. Finally, the convergence of the hierarchical poroelastic elements is investigated in the case of a particular material with several boundary conditions and excitations.

The studied problem consists of a single porous material with a rear face bonded onto a rigid wall (see Fig. 1). The boundary conditions on the lateral faces are either (i) free, or (ii) bonded. Two kinds of loads are used in the following, depending on the boundary conditions on the edges of the porous medium. In case (i), the porous medium is subjected to a rigid piston motion with a 10^{-3} m amplitude normal displacement [hereby referred to as configuration (i)]. In case (ii) a surface pressure of amplitude 1 Pa is imposed on the front face [hereby referred to as configuration (ii)]. A 5 cm thick porous material with lateral dimensions 0.35 m $\times$0.22 m is considered. Among the vibroacoustic indicators presented in the preceding section, only the mean square velocity along z axis (denoted $\langle V_z^2 \rangle$) and the mean quadratic pressure (denoted $\langle P^2 \rangle$) are studied. Actually the mean quadratic velocities along the lateral dimensions of the porous material are of limited interest in vibroacoustic applications.

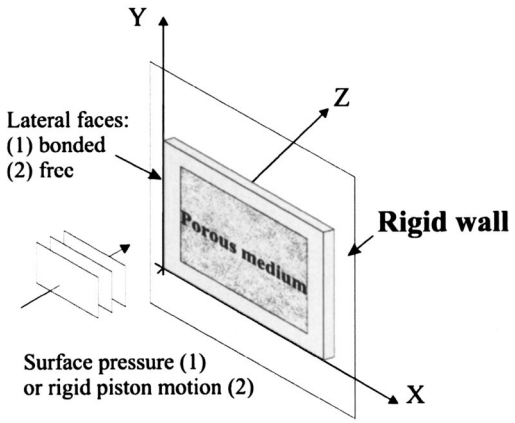


FIG. 1. Configuration of the problem.

A. Validation

In this subsection, the accuracy and the performance of the present approach is numerically assessed. In the following, the results obtained with hierarchical elements and finite elements are compared to a reference value. In the present paper, this reference is obtained by increasing considerably the interpolation orders for a fixed mesh when hierarchical elements are used, in order to be sure that convergence is reached. The same results can be obtained with classical finite elements but require a very refined mesh and thus huge computational resources. Hence, the classical finite element implementation has not been used for obtaining the reference values. In the following, the interpolation orders for the basis functions given a fixed mesh for the present approach or the classical finite element mesh are said to be acceptable when the response of the porous material is predicted within a tolerance from the reference response. This tolerance is fixed to 0.5 dB from the reference curve and 5% in the location of resonance peaks.

The first validation test consists in configuration (i). The material studied is a wool, UGW3, with characteristics given in Table I. $\langle P^2 \rangle$ is computed as a function of frequency and is actually the indicator which proves to have the most difficulties to converge. The results obtained with the classical code and the present approach with different meshes are presented in Fig. 2. Note that the maximum difference between the different meshes, including classical finite elements, and the reference curve is less than 0.5 dB which is within the defined convergence criterion. For the modeling with hierarchical elements, parameters for the mesh, namely the number of elements and the order of the basis functions, are given in the following form $n_x * n_y * n_z$ elements $(p_s - p_f)$ where n_x, n_y, n_z denote the number of elements used in directions x, y, z, respectively, and p_s, p_f denote the interpolation order of

TABLE I. Characteristics of the materials.

Material	h	σ kN s/m^4	α_∞	Λ μm	Λ' μm	ρ_S kg/m^3	N kPa	ν	η
FM2	0.9	25	7.8	226	226	300	286	0.4	0.265
UGW3	0.95	25	1.4	93.2	93.2	600	21	0	0.05

the basis functions for the solid phase and the fluid phase, respectively. Figure 2 shows that the present approach leads to accurate results for all tested meshes. To test the ability of the hierarchical poroelastic element to model different materials in various configurations a foam, FM2, with characteristics given in Table I is considered in configuration (ii). $\langle V_z^2 \rangle$ is computed as a function of frequency; this indicator has the most difficulties to converge. The results are presented in Fig. 3. Once again, accurate results are obtained with the hierarchical poroelastic elements for different meshes.

The excellent results for the two previous validation tests show the validity of the present approach. In addition, hierarchical poroelastic elements prove to be very performant. Tables II and III summarize the number of degrees of freedom (dof) required to ensure convergence, either with classical finite elements or hierarchical elements. Each of the two configurations studied above are considered. For the present approach, the number of degrees of freedom is given for different meshes and the corresponding required interpolation orders. Actually, the present approach allows for an important decrease of the number of dof required to satisfy the convergence criterion defined previously (0.5 dB from the reference curve and 5% accuracy in the location of resonance peaks). Moreover, the best performance is achieved using a coarse mesh and a high interpolation order for the basis functions. This trend has already been pointed out for the modeling of elastic materials using hierarchical elements in static problems.[22]

As seen in Tables II and III, the interpolation order for the basis functions can be chosen to be different for the two phases of the porous material. Actually, the solid and the fluid phase exhibit different physical behaviors depending on the studied configuration. This feature of the hierarchical poroelastic elements is of great interest since convergence is ensured by choosing the adequate interpolation order for each phase separately. On the other hand, convergence in classical finite elements is ensured by refining the mesh, i.e., a refined description of both phases of the porous material. Hence, the former approach provides an efficient way for the reduction of the number of degrees of freedom required for a correct modeling of the material.

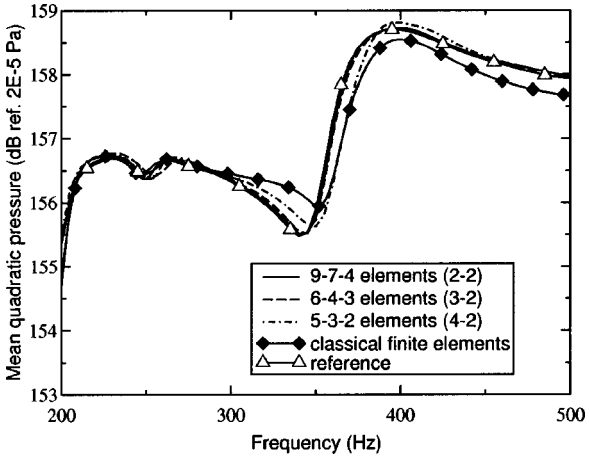


FIG. 2. $\langle P^2 \rangle$ for glasswool UGW3 in configuration (i).

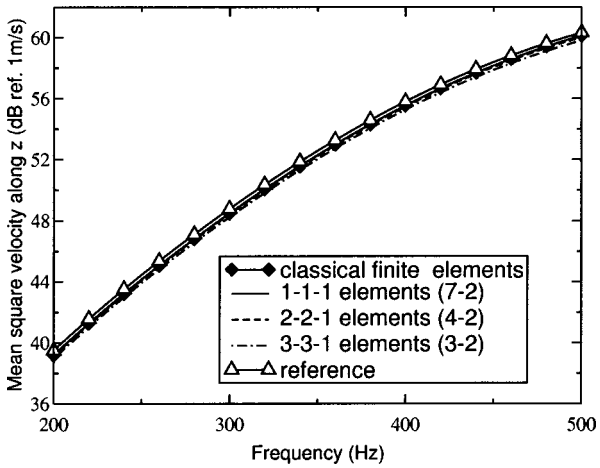


FIG. 3. $\langle V_z^2 \rangle$ for a foam FM2 in configuration (ii).

B. Convergence of the vibroacoustic indicators

In the following, foam FM2 is studied in configurations (i) and (ii). The geometry of the porous sample is identical to the one considered previously. The frequency range of interest is [0 Hz; 500 Hz]. Biot's theory of poroelasticity[2] indicates that two compression waves (denoted P_1 and P_2 in the following) and a shear wave (denoted S) can propagate simultaneously in a porous medium. For each kind of wave, the ratio μ of the fluid phase displacement to the solid phase displacement can be computed.[3] These ratio at 500 Hz for foam FM2 are given in Table IV. Table IV indicates that waves P_1 and S propagate in both phases. On the contrary, wave P_2 propagates mainly in the fluid phase, so that the two phases of the porous material can be considered as decoupled as far as wave P_2 is concerned. This point strongly influences the behavior of the porous materials according to the configuration studied. Foam FM2 has been previously studied in configuration (ii). Excellent values for $\langle V_z^2 \rangle$, represented in Fig. 3, can be obtained with low interpolation order for the basis functions of the fluid phase. This means that a coarse description of the fluid phase suffices for the convergence of this indicator related to the solid phase. Indeed, this porous material in configuration (ii) is subjected to an acoustical excitation acting on both phases. The pressure in the pores and the stress tensor in the solid phase *in vacuo* are imposed at the interface where the material is excited. Because wave P_2 mainly propagates in the fluid phase, the fluid phase has a weak influence on the solid phase and thus does

TABLE II. Convergence of $\langle V_z^2 \rangle$ for FM2 in configuration (ii).

Nb of elements interpolation order (solid–fluid)	Classical code	1–1–1 7–2	2–2–1 4–2	3–3–1 3–2
Number of dof	2144	70	144	244

TABLE III. Convergence of $\langle P^2 \rangle$ for UGW3 in configuration (i).

Nb of elements interpolation order (solid–fluid)	Classical code	4–3–2 3–2	5–3–2 3–2	6–4–3 3–2	9–7–4 2–2
Number of dof	5214	895	1094	2397	4828

not require a very accurate description for the convergence of $\langle V_z^2 \rangle$: the motion in the solid phase is due to wave P_1 and S. In comparison, indicator $\langle P^2 \rangle$ for foam FM2 in the same configuration is represented in Fig. 4. Different interpolation orders have been used with a fixed $2\times2\times1$ elements mesh for the present approach. Compared to $\langle V_z^2 \rangle$, $\langle P^2 \rangle$ requires a noticeable lower interpolation order for the basis functions in the solid phase. This indicates that a coarse representation of the solid phase suffices for the convergence of this latter indicator. The influence of waves P_1 and S on the motion of the fluid phase is not significant. Actually the motion of the fluid phase is mainly due to wave P_2.

Next, consider foam FM2 in configuration (i). As previously, $\langle P^2 \rangle$ is plotted as a function of frequency in Fig. 5. It can be observed that a higher interpolation order for the basis functions of the solid phase in comparison with configuration (ii) is required for the convergence of $\langle P^2 \rangle$. In this configuration, the porous is submitted to a rigid piston motion. For that kind of excitation, the solid phase is directly excited and its motion is strongly transmitted to the fluid when the amplitude ratios μ are important for waves P_1 and S. Hence, in configuration (i) the motion of the fluid phase is mainly due to these latter waves. In conclusion, the two phases of a porous material can exhibit very different behaviors according to the excitation. This induces a convergence which highly depends on the configuration studied for the considered vibroacoustic indicators, namely $\langle V_z^2 \rangle$ or $\langle P^2 \rangle$. Hence, it is difficult to conclude about the convergence of a set of indicators in a general manner.

C. Investigation of the convergence of hierarchical poroelastic elements

For 3D deformations, the classical convergence criterion gives indications for a minimal mesh but is insufficient because of locking of 3D linear elements and discrepancies of the fields in the two phases of the porous material. As stated in Sec. III A, the use of the theory of hierarchical elements together with the {**u**, **P**} formulation enables one to solve these latter problems and to get smaller linear systems to solve.

In this section, the convergence of hierarchical poroelastic elements is investigated for a particular material, wool UGW3. The aim of this study is to determine the interpolation order for the basis functions in the two phases of the porous material that allows for convergence given the element size. Initially, the motivation for this study was to derive a convergence criterion that could be reused for other porous materials. However, it appears that the results for wool UGW3 cannot be directly transposed and that only a convergence study performed on different materials would enable one to draw general trends for the convergence of hierarchical poroelastic elements. This task implies a huge amount of calculations. Therefore, only preliminary results are given here. This convergence study is however helpful to highlight the influence of the coupling between the two phases of wool UGW3 as it is shown thereafter and to confirm the observations made on foam FM2 in Sec. III B.

TABLE IV. Amplitude ratios for foam FM2 at 500 Hz.

Kind of wave	P_1	P_2	S
Amplitude ratio μ	$0.998+0.161\times i$	$-10.43+1.70\times i$	$0.97-1.14\times i$

In the following, the mesh in the lateral dimensions of a 3D porous sample is studied. As a consequence, the number of elements along the thickness is chosen large enough so that only the lateral mesh influences the results. For a chosen indicator, the interpolation order in both phases can be represented as a function of the element size. In order to have nondimensional data in the abscissa, the element size is scaled to the wavelength λ of wave P_1, P_2 or S at the highest frequency in the frequency range of interest. These wavelengths at 500 Hz are given in Table V. As these wavelengths are rather short, a wide range of element size to wavelength ratio is provided. The study that follows is carried out for each of the two configurations (i) and (ii) depicted in Sec. III A and for each vibroacoustic indicator ($\langle V_z^2 \rangle$ and P^2) according to remarks in Sec. III B.

Configuration (i): The interpolation orders for the solid phase (respectively, fluid phase) basis functions that ensure the convergence of the two indicators $\langle P^2 \rangle$ and $\langle V_z^2 \rangle$ are plotted as a function of the ratio of the element size to λ_{P_1}, where λ_{P_1} denotes the wavelength of wave P_1 at 500 Hz. The results are represented in Fig. 6 (respectively, Fig. 7). In these figures, fit curves are provided for clarity. Note that the wavelength of another kind of wave, e.g., S wave or P_2, can be chosen to scale the element size simply by multiplying the

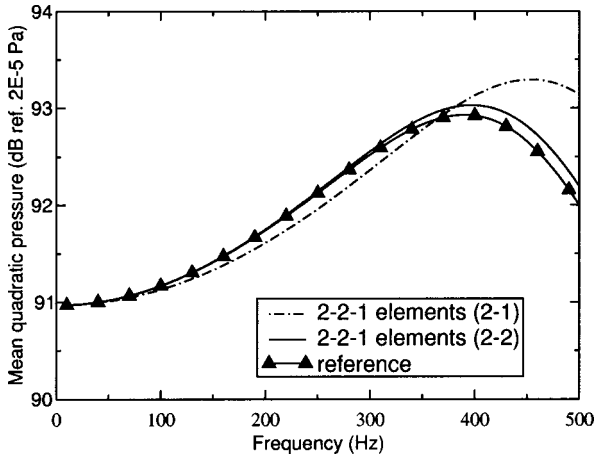


FIG. 4. $\langle P^2 \rangle$ for foam FM2 in configuration (ii).

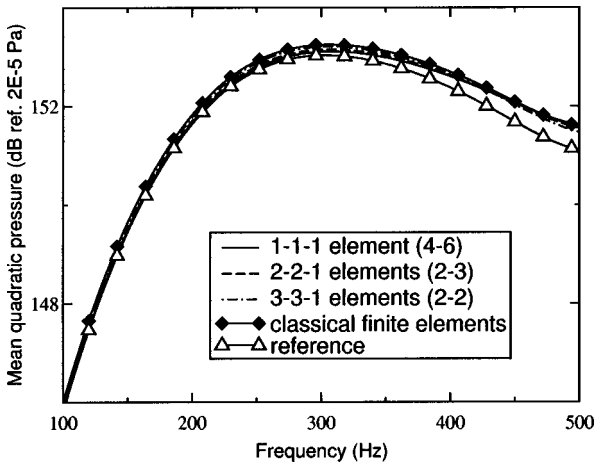


FIG. 5. $\langle P^2 \rangle$ for foam FM2 in configuration (i).

abscissa in Figs. 6 and 7 by the factor λ_{P_1}/λ_S or $\lambda_{P_1}/\lambda_{P_2}$.

Considering a fixed mesh, it is seen in Figs. 6 and 7 that the convergence of indicator $\langle P^2 \rangle$ is not solely governed by the interpolation order for the basis functions in the fluid phase. Indeed, a sufficiently high interpolation order for the basis functions in the solid phase is required to ensure the convergence of $\langle P^2 \rangle$. This interpolation is even higher than the one needed for the convergence of $\langle V_z^2 \rangle$. The opposite is observed regarding the interpolation order for the basis functions in fluid phase. This is because of important coupling phenomena between the two phases when the solid phase of the porous material is directly excited.

Configuration (ii): Like in configuration (i), the interpolation orders required for the convergence of the two indicators $\langle P^2 \rangle$ and $\langle V_z^2 \rangle$ are plotted as a function of the ratio of the element size to λ_{P_1}. These interpolation orders for the solid and the fluid phase are represented in Figs. 8 and 9, respectively. As previously, fit curves are provided for clarity. When considering the basis functions in the solid phase, Fig. 8 shows that for a fixed element size to λ_{P_1}, the interpolation order required for the convergence of $\langle P^2 \rangle$ is as high as the one for $\langle V_z^2 \rangle$, except for low size element to wavelength ratio. For the basis functions of the fluid phase, the interpolation order ensuring the convergence of $\langle V_z^2 \rangle$ is smaller than the one ensuring the convergence of $\langle P^2 \rangle$. Hence, from these observations, coupling phenomena between the two phases of the porous material appear to influence weakly the response of the porous material. This latter phenomenon can be explained by the fact that the porous material has bonded edges and is submitted to an acoustical excitation. To sum up the results found in this particular configuration, given a fixed mesh, the interpolation orders for the basis functions in the two phases of the porous material chosen to ensure simultaneously the convergence of $\langle P^2 \rangle$ and $\langle V_z^2 \rangle$ exhibit the following feature: regarding the basis

TABLE V. Wave numbers for the different kinds of waves in wool UGW3 at 500 Hz.

Kind of wave	P_1	P_2	S
Wavelength (m)	7.237×10^{-2}	1.962×10^{-1}	5.199×10^{-2}

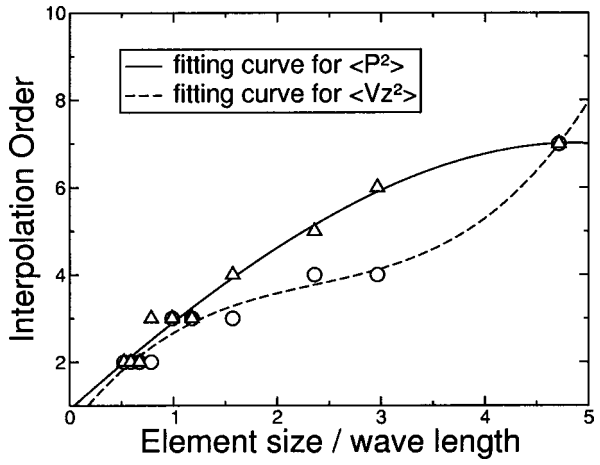


FIG. 6. Convergence criterion for the solid phase. Configuration (i). Given an element size for the mesh, an interpolation order in the solid phase that suffices for the convergence of $\langle P^2 \rangle$ (respectively, $\langle V_z^2 \rangle$) is determined by numerical experiments. △ (respectively, ○) stands for the corresponding couple of values interpolation order-element size to wavelength ratio.

functions in the solid phase (respectively, fluid phase), the choice of the interpolation order ensuring the convergence of $\langle V_z^2 \rangle$ (respectively, $\langle P^2 \rangle$) allows for the convergence of both indicators.

IV. CONCLUSION

The implementation of the {**u**, **P**} formulation for porous materials using the theory of hierarchical elements has been presented. The use or high order polynomials for the basis functions enables one to solve some of the problems met with linear poroelastic elements, among them numerical locking and the presence of two different scale phenomena. Hence, hierarchical poroelastic elements allows for the prediction of the forced response of a porous material using a reduced number of degrees of freedom. The reduction of the number of unknowns is all the more effective than a coarse mesh with high interpolation orders is provided. Besides, the

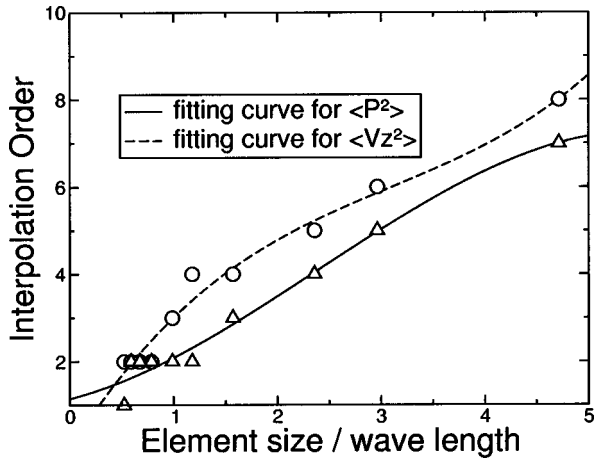


FIG. 7. Convergence criterion for the fluid phase. Configuration (i). Given an element size for the mesh, an interpolation order in the fluid phase that suffices for the convergence of $\langle P^2 \rangle$ (respectively, $\langle V_z^2 \rangle$) is determined by numerical experiments. △ (respectively, ○) stands for the corresponding couple of values interpolation order-element size to wavelength ratio.

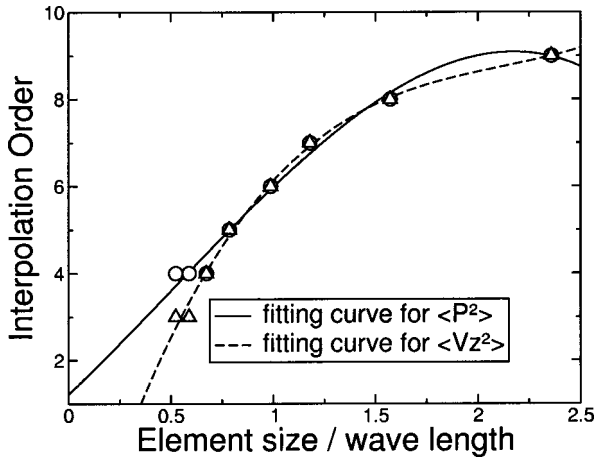


FIG. 8. Convergence criterion for the solid phase. Configuration (ii). Given an element size for the mesh, an interpolation order in the solid phase that suffices for the convergence of $\langle P^2 \rangle$ (respectively, $\langle V_z^2 \rangle$) is determined by numerical experiments. ○ (respectively, △) stands for the corresponding couple of values interpolation order-element size to wavelength ratio.

performance of hierarchical elements is enhanced by the use of different interpolation orders for the basis functions of each phase separately.

Considering a particular material, a preliminary study has been conducted in order to link the interpolation order of the basis functions in the two phases of the porous material to the ratio of the element size to the wavelength. It has been found that for a mechanical excitation, each phase happens to influence strongly the vibratory state of the other phase, showing the importance of the coupling between the two phases of the porous material. Regarding interpolation order for the basis functions in the solid phase, this implies, for example, to chose an interpolation order for the convergence of $\langle P^2 \rangle$ larger than the one for $\langle V_z^2 \rangle$. This interaction phenomenon proves to be of least importance for an acoustical excitation. In this case, considering the basis functions in the solid phase, the interpolation order for the convergence of $\langle V_z^2 \rangle$ is found to be as large as for $\langle P^2 \rangle$. To derive more

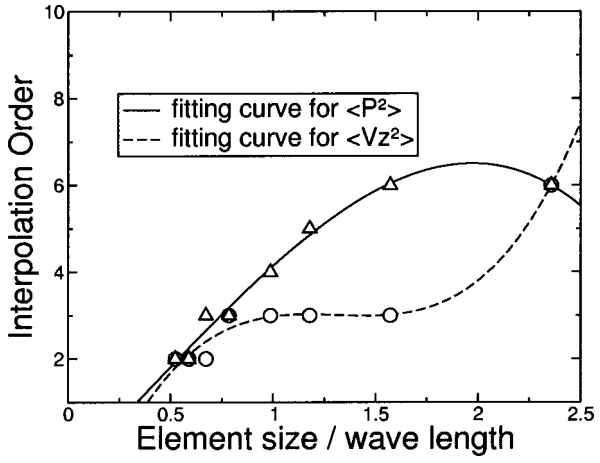


FIG. 9. Convergence criterion for the fluid phase. Configuration (ii). Given an element size for the mesh, an interpolation order in the fluid phase that suffices for the convergence of $\langle P^2 \rangle$ (respectively, $\langle V_z^2 \rangle$) is determined by numerical experiments. △ (respectively, ○) stands for the corresponding couple of values interpolation order-element size to wavelength ratio.

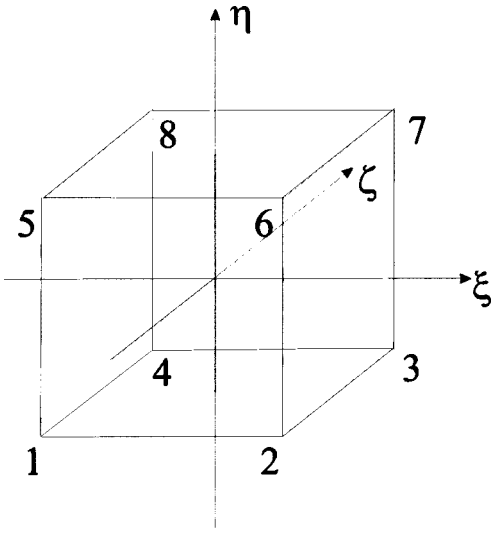


FIG. 10. The reference element and the standard numbering of nodes.

general trends for the convergence of hierarchical poroelastic elements, further works involving the study of other kinds of materials are required. However, the presented examples show the difficulty of deriving a general meshing criterion for porous-elastic materials. More importantly, they highlight the importance of conducting a serious convergence study before any vibroacoustic study regarding the behavior of porous-elastic materials, based on using finite element codes.

APPENDIX

The shape functions involved in the approximation of the fields on one element [Eq. (6)] are defined on a parent element, which is an 8-noded brick element represented in Fig. 10. On that element, the geometrical entities (nodes, edges, and faces) are assigned a number. From this follows the definition of the basis shape functions.

1. Node modes

A node mode of analytical expression $\mathcal{N}_i$, is related to node i on the parent element. $\mathcal{N}_i$ is a linear function of the local coordinates and has value 1 at node i and 0 at the other nodes. Node modes are given by

$$\mathcal{N}_i(\xi,\eta,\zeta)=\tfrac{1}{8}(1-\xi)(1-\eta)(1-\zeta),$$
$$\mathcal{N}_2(\xi,\eta,\zeta)=\tfrac{1}{8}(1+\xi)(1-\eta)(1-\zeta),$$
$$\mathcal{N}_3(\xi,\eta,\zeta)=\tfrac{1}{8}(1+\xi)(1+\eta)(1-\zeta),$$
$$\mathcal{N}_4(\xi,\eta,\zeta)=\tfrac{1}{8}(1-\xi)(1+\eta)(1-\zeta),$$
$$\mathcal{N}_5(\xi,\eta,\zeta)=\tfrac{1}{8}(1-\xi)(1-\eta)(1+\zeta),$$
$$\mathcal{N}_6(\xi,\eta,\zeta)=\tfrac{1}{8}(1+\xi)(1-\eta)(1+\zeta),$$
$$\mathcal{N}_7(\xi,\eta,\zeta)=\tfrac{1}{8}(1+\xi)(1+\eta)(1+\zeta),$$
$$\mathcal{N}_8(\xi,\eta,\zeta)=\tfrac{1}{8}(1-\xi)(1+\eta)(1+\zeta).$$

2. Edge modes

The edge i linking two nodes of the parent element is associated one or several edge modes. One particular edge

mode of analytical expression $\mathcal{E}_p$ has zero value on all the edges of the parent element except on edge j. On this particular edge, its expression is a polynomial with order p. The general expression of the edges modes for $p \geqslant 2$ are given by

$$\mathcal{E}_p^{12} = \tfrac{1}{4}\phi_p(\xi)(1-\eta)(1-\zeta),$$
$$\mathcal{E}_p^{23} = \tfrac{1}{4}\phi_p(\eta)(1+\xi)(1-\zeta),$$
$$\mathcal{E}_p^{34} = \tfrac{1}{4}\phi_p(\xi)(1+\eta)(1-\zeta),$$
$$\mathcal{E}_p^{41} = \tfrac{1}{4}\phi_p(\eta)(1-\xi)(1-\zeta),$$
$$\mathcal{E}_p^{15} = \tfrac{1}{4}\phi_p(\zeta)(1-\xi)(1-\eta),$$
$$\mathcal{E}_p^{26} = \tfrac{1}{4}\phi_p(\zeta)(1+\xi)(1-\eta),$$
$$\mathcal{E}_p^{37} = \tfrac{1}{4}\phi_p(\zeta)(1-\xi)(1+\eta),$$
$$\mathcal{E}_p^{48} = \tfrac{1}{4}\phi_p(\zeta)(1+\xi)(1+\eta),$$
$$\mathcal{E}_p^{56} = \tfrac{1}{4}\phi_p(\xi)(1-\eta)(1+\zeta),$$
$$\mathcal{E}_p^{67} = \tfrac{1}{4}\phi_p(\eta)(1+\xi)(1+\zeta),$$
$$\mathcal{E}_p^{78} = \tfrac{1}{4}\phi_p(\xi)(1+\eta)(1+\zeta),$$
$$\mathcal{E}_p^{85} = \tfrac{1}{4}\phi_p(\eta)(1-\xi)(1+\zeta),$$

where the superscript kl indicates that the considered edges is defined by node k and node l on the parent element. Functions ϕ_p are built using orthogonal Legendre polynomials L_p and are obtained by the formula

$$\phi_p(\xi) = \frac{1}{\sqrt{2(2j-1)}}(L_p(\xi) - L_{p-2}(\xi)) \quad \text{with} \quad p \geqslant 2. \tag{A1}$$

The relations verified by functions ϕ_p are given further on.

3. Face modes

Each face i of the parent element is associated with face modes which are zero on all the faces of the parent element except for face i. Considering the interpolation order $p \geqslant 4$, one or several couples of integers (k_1, k_2) verifying $k_1 + k_2 = p$ are constituted. Note that the couple (k_1, k_2) is different from the couple (k_2, k_1). Each couple is associated a mode related to face i. The general expression for this mode $\mathcal{F}_{k_1 k_2}$ is given by

$$\mathcal{F}_{k_1 k_2}^{1265} = \tfrac{1}{2}\phi_{k_1}(\xi)\phi_{k_2}(\zeta)(1-\eta),$$
$$\mathcal{F}_{k_1 k_2}^{2376} = \tfrac{1}{2}\phi_{k_7 1}(\eta)\phi_{k_2}(\zeta)(1-\xi),$$
$$\mathcal{F}_{k_1 k_2}^{1234} = \tfrac{1}{2}\phi_{k_1}(\xi)\phi_{k_2}(\eta)(1-\zeta),$$
$$\mathcal{F}_{k_1 k_2}^{4158} = \tfrac{1}{2}\phi_{k_1}(\eta)\phi_{k_2}(\zeta)(1+\xi),$$
$$\mathcal{F}_{k_1 k_2}^{3487} = \tfrac{1}{2}\phi_{k_1}(\xi)\phi_{k_2}(\zeta)(1+\eta),$$
$$\mathcal{F}_{k_1 k_2}^{5678} = \tfrac{1}{2}\phi_{k_1}(\xi)\phi_{k_2}(\eta)(1+\zeta),$$

where the superscript $klmn$ states for the face defined by the nodes k, l, m, and n.

4. Internal modes

For an interpolation order $p \geqslant 6$, sets of three integers (k_1, k_2, k_3) verifying $k_1 + k_2 + k_3 = p$ are considered. Note that the set (k_1, k_2, k_3) is different from the set (k_1, k_3, k_2) and (k_2, k_3, k_1). Each set is associated with an internal mode of analytical expression $\mathcal{I}_{k_1 k_2 k_3}$ given by

$$\mathcal{I}_{k_1 k_2 k_3} = \phi_{k_1}(\xi) \cdot \phi_{k_2}(\eta) \cdot \phi_{k_3}(\zeta).$$

5. Features of functions ϕ_k

The functions ϕ_k introduced in the preceding sections satisfy some interesting properties. First, by construction,

$$\phi_k(-1) = \phi_k(1) = 0. \tag{A2}$$

Also the first derivative of these functions are orthogonal, namely,

$$\int_{-1}^{1} \phi_m'(\xi)\phi_n'(\xi)\,d\xi = \delta_{mn}. \tag{A3}$$

Finally, let's consider the analytical expression of a particular edge mode with an interpolation order k, for example, $\mathcal{E}_k^{12}$. The integration of this mode on the parent element face pertaining to the nodes 1, 2, 3, 4 leads to

$$\int_{1234} \mathcal{E}^{12}\,dS = \int_{-1}^{1}(1-\eta)\,d\eta \cdot \int_{-1}^{1} \phi_k(\xi)\,d\xi. \tag{A4}$$

For $k > 2$, the function ϕ_k verify the relation

$$\int_{-1}^{1} \phi_k(\xi)\,d\xi = \sqrt{\frac{2}{2j+1}}\left[\frac{1}{2k+1}[\phi_{k+1}(1) - \phi_{k+1}(-1)] - \frac{1}{2k-3}[\phi_{k-1}(1) - \phi_{k-1}(-1)]\right]. \tag{A5}$$

Given Eq. (A2), the value of this integral is 0. Hence all the edge modes with an interpolation order $k > 2$ satisfy Eq. (20). The extension of this relation for the face modes with interpolation order $p > 4$ is straightforward.

[1] M. Delany and E. Bazley, "Acoustic properties for fibrous absorbent materials," Appl. Acoust. **3**, 105 (1970).
[2] M. Biot, "The theory of propagation of elastic waves in a fluid-saturated porous," J. Acoust. Soc. Am. **28**, 168 (1956).
[3] J. Allard, *Propagation of Sound in Porous Media. Modeling Sound Absorbing Materials* (Elsevier, New York, 1993).
[4] J. Bolton and E. Green, "Normal incidence sound transmission through double panel," Appl. Acoust. **39**, 25 (1993).
[5] U. Ingard, *Notes on Sound Absorption Technology* (Noise Control Foundation, 1994).
[6] Y. Kang and J. Bolton, "Finite element modeling of isotropic porous materials coupled with acoustical finite elements," J. Acoust. Soc. Am. **98**, 635 (1995).
[7] Y. Kang and J. Bolton, "A finite element model for sound transmission through foam-lined double-panel structures," J. Acoust. Soc. Am. **99**, 2755 (1996).
[8] R. Panneton and N. Atalla, "Numerical prediction of sound transmission through finite multilayer systems with poroelastic materials," J. Acoust. Soc. Am. **100**, 346 (1996).
[9] 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 (1997).
[10] J. Coyette and H. Wynendaele, "A finite element model for predicting the acoustic transmission characteristics of layered structures," *INTERNOISE 95* (1995), pp. 1279–1282.

[11] N. Atalla, R. Panneton, and P. Debergue, ''A mixed displacement-pressure formulation for poroelastic materials,'' J. Acoust. Soc. Am. **104**, 1444 (1998).
[12] 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 (1999).
[13] N. Dauchez, S. Sahraoui, and N. Atalla, ''Convergence of poroelastic finite elements based on biot displacement formulation,'' J. Acoust. Soc. Am. **109**, 33 (2001).
[14] P. Debergue, ''Développement d'une formulation mixte déplacement-pression pour les matériaux poroélastiques (Development of a mixed displacement-pressure formulation for poroelastic materials),'' Master thesis, University of Sherbrooke, 1997.
[15] F. Sgard, N. Atalla, and R. Panneton, ''A modal reduction technique for the finite element formulation of biot's poroelasticity equations in acoustics,'' J. Acoust. Soc. Am. **102**, 3112 (1997).
[16] F. Sgard, N. Atalla, and R. Panneton, ''A modal reduction technique for the finite element formulation of biot's poroelasticity equations in acoustics applied to multilayered structures,'' J. Acoust. Soc. Am. **103**, 2882 (1998).
[17] F. Sgard, N. Atalla, and R. Panneton, ''A mixed wave-finite element approach for solving biot's poroelasticity equations in acoustics,'' J. Acoust. Soc. Am. **105**, 1195 (1999).
[18] N. Dauchez, S. Sahraoui, and N. Atalla, ''Dissipation mechanisms in a porous layer bonded onto a plate,'' J. Sound Vib. (to be published).
[19] S. Rigobert, F. Sgard, and N. Atalla, ''Investigation on the dissipative effects in a poroelastic plate,'' J. Acoust. Soc. Am. **105**, 1295 (1999).
[20] I. Babuška and B. Szabò, ''The p and hp version of the finite element method, basic principle and properties,'' SIAM Rev. **36**, 578 (1994).
[21] I. Babuška and B. Szabò, ''On the rate of convergence of the finite element method,'' Int. J. Numer. Methods Eng. **18**, 323 (1982).
[22] B. Szabò and I. Babuška, *Finite Element Analysis* (Wiley, New York, 1991).
[23] B. Szabò, ''Hierarchic plate and shell models based on p-extension,'' Analytical and Computational Models of Shells presented at the Winter Annual Meeting of ASME Computers and Engineering division (10–15 December 1982).
[24] A. Côté, ''Modélisation vibroacoustique dans le domaine des moyennes fréquences par éléments finis de type p (Modeling vibroacoustic problems in the medium frequency range using p elements),'' Ph.D. thesis, University of Sherbrooke, 1998.
[25] H. Hinnant, ''Fast method of numerical quadrature for p-version finite element matrice,'' Int. J. Numer. Methods Eng. **37**, 3723 (1994).

Elastic constants of layers in isotropic laminates

Paul R. Heyliger
Department of Civil Engineering, Colorado State University, Fort Collins, Colorado 80523

Hassel Ledbetter
Los Alamos National Laboratory (E536), Los Alamos, New Mexico 87545

Sudook Kim
Materials Science and Engineering Laboratory, National Institute of Standards and Technology, Boulder, Colorado 80305

Ivar Reimanis
Metallurgical and Materials Engineering Department, Colorado School of Mines, Golden, Colorado 80401

(Received 21 June 2002; revised 25 November 2002; accepted 25 August 2003)

The individual laminae elastic constants in multilayer laminates composed of dissimilar isotropic layers were determined using ultrasonic-resonance spectroscopy and the linear theory of elasticity. Ultrasonic resonance allows one to measure the free-vibration response spectrum of a traction-free solid under periodic vibration. These frequencies depend on pointwise density, laminate dimensions, layer thickness, and layer elastic constants. Given a material with known mass but unknown constitution, this method allows one to extract the elastic constants and density of the constituent layers. This is accomplished by measuring the frequencies and then minimizing the differences between these and those calculated using the theory of elasticity for layered media to select the constants that best replicate the frequency-response spectrum. This approach is applied to a three-layer, unsymmetric laminate of W_pCu, and very good agreement is found with the elastic constants of the two constituent materials. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1618754]

Pages: 2618–2625

PACS numbers: 43.20.Ks, 43.40.Dx [YHB]

I. INTRODUCTION

Determining the elastic constants of materials with unknown constitution is a much-studied problem in science and engineering, and there are numerous methods available for solving it. One method that is particularly robust is ultrasonic-resonance spectroscopy (URS), in which the free-vibration response spectrum of the solid is used in combination with the theoretical values of frequencies for objects with known density, geometry, and elastic constants.[1–6] Numerous applications of this method, along with descriptions of the experimental system, measurements, and theoretical analysis, are given in Migliori and Sarrao.[6] The theoretical predictions are based on the Ritz method, in which the solution of the weak form of the equations of periodic motion is sought as given in Hamilton's principle, where the displacements are given in a finite series in terms of the spatial coordinates of the specimen geometry. Excellent accuracy can be obtained using this approach.

To date, most applications of URS for the determination of elastic constants have been for homogeneous media. Consistent with these applications, the computational algorithms developed for this purpose have without exception used basis functions for the displacement components that have been continuous with continuous derivatives (such as the Legendre polynomials used for parallelepipeds by Demarest[2] and power series as used in the more general method developed by Visscher and colleagues[7]). This is a valid and useful approach for homogeneous media, but special care must be taken when considering dissimilar media. At an interface between two materials that differ in elastic properties, the transverse shear stresses are continuous across the interface but there is a jump in shear modulus. Hence, there is a discontinuity in the through-thickness displacement gradient that cannot be represented using functions with C^1 continuity (i.e., functions such as power series, which possess continuous first derivatives over the specimen domain).

Recently, we developed a method to account for the through-thickness behavior of laminated elastic and piezoelectric media.[8] This method is based on a discrete-layer approximation to the weak form of the equations of motion, in which we split the through-thickness and in-plane dependence of the approximation functions in the Ritz method. Similar approaches have been used in related problems by Pauley and Dong[9] for wave propagation in laminated piezoelectric media, and a generalized discrete-layer approach for elastic laminates by Reddy.[10] With this model in hand, our present objective is to evaluate the elastic constants and densities of layered parallelepipeds where only the edge dimensions, layer thicknesses, total mass, and free-vibration-response spectra are given. To our knowledge, this is the first application of URS to dissimilar composites of this type, and it results in a method where, rather than separating the individual constituents of the composite, we can consider the component as a whole.

II. MATERIAL

Three-layer W_p/Cu laminates were fabricated by a powder-metallurgy approach in which powders of specific compositions were mixed, layered, and then densified in a hot press. High-purity (99.9) 1–5-μm-diameter copper

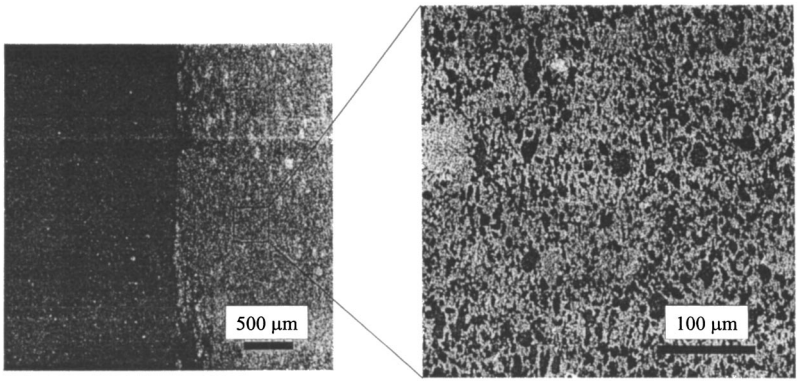


FIG. 1. Microstructure of the trilayer composite interface and the copper/tungsten composite.

powder[11] was mixed with high-purity (99.95) W powder with an average particle diameter of 1 μm.[12] Cu–W powders were made by mixing amounts of pure W and Cu that would lead to a 20 percent composite when fully dense, and milling in a polycarbonate bottle with four to five cleaned alumina balls for 24 h. The laminate was made by pouring first the Cu–W powder, then the pure Cu powder, and finally more Cu–W powder into a 25.4-mm-diameter graphite die. The inner walls of the die were coated with BN to prevent chemical reaction with carbon. After each powder type was poured, a clean steel punch was used for leveling. This resulted in a powder stack consisting of 20%W/0%W/20%W. The stacked powders were then sintered in a hot press under a vacuum of 1.33×10^{-2} Pa, a load of 40 MPa at 985 °C for 15 min. Cooling was accomplished with the load still applied at a rate of about 20 °C/min, until about 500 °C, at which point the rate slowed substantially. A dense trilayer sample consisting of 20%W/0%W/20%W resulted, with layer thicknesses of 1.50 mm/1.88 mm/1.66 mm, respectively.

The composite microstructure and the nature of the interface between layers is apparent in the scanning electron micrograph shown in Fig. 1. The light and dark contrasts represent the W and Cu, respectively. Consistent with the fact that Cu and W do not react and are mutually insoluble, observations in the scanning electron microscope (SEM) indicated that the interface between the 20% W and 0% W layers represents a discrete transition between these compositions, without the presence of a discernible interphase. (The SEM's resolution limit is approximately 0.1 μm). The interface plane does exhibit some roughness. Typically, the interface position varies by less than 100 μm over 2 mm.

Single-composition composites (20% W and pure Cu) were produced as described above to measure the elastic properties of the material in the outer layers. The hot-pressed trilayer composites were subsequently cut by diamond saw into specimens of dimensions $7.01\times7.84\times5.04$ mm, with the layers in the third dimension. Single-composition specimens were cut by electrodischarge machining to dimensions $10.0\times9.0\times8.0$ mm, with no layers.

III. THEORY

A. Geometry and boundary conditions

The rectangular parallelepiped described in the previous section is a special case of the general layered block assumed in our theoretical calculations. The parallelepiped is in general composed of an arbitrary number of elastic and isotropic layers. The parallelepiped has dimensions L_x, L_y, and L_z, and the z direction is perpendicular to each dissimilar-material interface. Each layer has constitutive equations that can be expressed as[13]

$$\sigma_{ij}=\lambda\,\delta_{ij}e_{kk}+2\mu e_{ij}\,. \tag{1}$$

Here, λ and μ are the Lamé parameters, δ_{ij} is the Kronecker delta, σ_{ij} are the components of stress, and e_{ij} are the components of infinitesimal strain. We also use the alternate form of the constitutive relation

$$\sigma_{ij}=C_{ijkl}e_{kl}\,, \tag{2}$$

where the C_{ijkl} are the components of the elastic-stiffness tensor, which can be expressed in terms of the two Lamé parameters. The strain-displacement relations are given by

$$e_{ij}=\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right). \tag{3}$$

Here, u_i represent the displacement components.

Hamilton's principle forms the basis for the weak form of the equations of motion[14]

$$\delta\int_{t_0}^{t}dt\int_V\left[\frac{1}{2}\rho\dot{u}_j\dot{u}_j-U_o(e_{kl})\right]dV+\int_{t_0}^{t}dt\int_S\overline{T}_k\delta u_k\,dS=0. \tag{4}$$

Here, t is time, V and S are the volume and surface occupied by and bounding the solid, $\overline{T}$ are the specified surface tractions, δ is the variational operator, the overdot superscript represents differentiation with respect to time, and U_o represents the strain-energy density, given for a linear elastic material as

$$U=\tfrac{1}{2}C_{ijkl}e_{ij}e_{kl}\,. \tag{5}$$

The weak form of the governing equations, as well as the governing differential equations themselves, can be found by substituting the above relations into Hamilton's principle. Here, we use the usual contracted notation for the elastic stiffnesses C_{ijkl} by compressing the ij and kl indices into a single index ranging from 1 to 6, and maintaining the range of m from 1 to 3. For example, C_{1122} becomes C_{12}, and so on. In rectangular Cartesian coordinates, we set $x_1=x$, $x_2=y$, and $x_3=z$, with the displacements as $u_1=u(x,y,z)$, $u_2=v(x,y,z)$, and $u_3=w(x,y,z)$. The weak form can be expressed using this nomenclature as

$$\delta\int_{t_0}^{t} dt \int_V \Big[\rho(\dot{u}\,\delta\dot{u}+\dot{v}\,\delta\dot{v}+\dot{w}\,\delta\dot{w}) - \Big[C_{11}\frac{\partial u}{\partial x}\frac{\partial \delta u}{\partial x} + C_{12}\frac{\partial u}{\partial x}\frac{\partial \delta v}{\partial y} + C_{13}\frac{\partial u}{\partial x}\frac{\partial \delta w}{\partial z} + C_{12}\frac{\partial v}{\partial y}\frac{\partial \delta u}{\partial x} + C_{22}\frac{\partial v}{\partial y}\frac{\partial \delta v}{\partial y}$$
$$+ C_{23}\frac{\partial v}{\partial y}\frac{\partial \delta w}{\partial z} + C_{13}\frac{\partial w}{\partial z}\frac{\partial \delta u}{\partial x} + C_{23}\frac{\partial w}{\partial z}\frac{\partial \delta v}{\partial y} + C_{33}\frac{\partial w}{\partial z}\frac{\partial \delta w}{\partial z} C_{44}\left(\frac{\partial v}{\partial z}+\frac{\partial w}{\partial y}\right)\left(\frac{\partial \delta v}{\partial z}+\frac{\partial \delta w}{\partial y}\right)$$
$$+ C_{55}\left(\frac{\partial u}{\partial z}+\frac{\partial w}{\partial x}\right)\left(\frac{\partial \delta u}{\partial z}+\frac{\partial \delta w}{\partial x}\right) + C_{66}\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)\left(\frac{\partial \delta u}{\partial y}+\frac{\partial \delta v}{\partial x}\right)\Big]\Big] dV + \int_{t_0}^{t} dt \int_S \overline{T}_k \delta u_k \, dS = 0. \tag{6}$$

B. Discrete-layer approximation

In many past studies, approximations to the three displacements are generated in terms of the global (x,y,z) coordinates. In this study, the dependence of the displacements on the z coordinate is separated from the functions in x and y. This allows for global functions in x and y that result in a subsequent reduction of the size of the computational problem. Hence, approximations for the three displacement components are sought in the form[10]

$$u(x,y,z,t) = \sum_{i=1}^{m}\sum_{j=1}^{n} U_{ji}(t)\Psi_i^u(x,y)\overline{\Psi}_j^u(z),$$
$$v(x,y,z,t) = \sum_{i=1}^{m}\sum_{j=1}^{n} V_{ji}(t)\Psi_i^v(x,y)\overline{\Psi}_j^v(z), \tag{7}$$
$$w(x,y,z,t) = \sum_{i=1}^{m}\sum_{j=1}^{n} W_{ji}(t)\Psi_i^w(x,y)\overline{\Psi}_j^w(z).$$

Here, m and n are the respective number of in-plane and through-thickness terms used to approximate each variable. The approximations for each of the three field quantities are constructed in such a way as to separate the dependence in the planar coordinate variables from that in the coordinate variable perpendicular to the interface. The reason for this is that the change in the material properties forces a break in the gradients of the displacements across an interface. This can be seen easily by considering the case of shear stress at a dissimilar-material interface. Since the stress must be continuous across an interface but the shear modulus is different for two layers, the shear strain must be different. This implies that the slope of the displacement variables across the interface must be different, thus eliminating functions such as the commonly used power series[7] or Legendre polynomials.[3]

In the thickness direction, one-dimensional Lagrangian interpolation polynomials are used for $\overline{\Psi}_j(z)$. For the in-plane approximations, different types of approximations can be used for the two-dimensional functions $\Psi_j(x,y)$. We use power series for the problem of traction-free vibration. For a parallelepiped with n layers $(n-1)$ is the number of subdivisions through the parallelepiped thickness (typically taken equal to or greater than the number of layers in the parallelepiped), and Γ_{ji} is the value of component Γ at height j corresponding to the ith in-plane approximation function.[10]

Substituting these approximations into the weak form in Eq. (6), introducing the assumption of periodic motion, collecting the coefficients of the variations of the displacements, and placing the results in matrix form, we obtain the result

$$\rho\omega^2\begin{bmatrix}[M^{11}] & [0] & [0]\\ [0] & [M^{22}] & [0]\\ [0] & [0] & [M^{33}]\end{bmatrix}\begin{Bmatrix}\{u\}\\ \{v\}\\ \{w\}\end{Bmatrix}$$
$$-\begin{bmatrix}[K^{11}] & [K^{12}] & [K^{13}]\\ [K^{21}] & [K^{22}] & [K^{23}]\\ [K^{31}] & [K^{32}] & [K^{33}]\end{bmatrix}\begin{Bmatrix}\{u\}\\ \{v\}\\ \{w\}\end{Bmatrix} = \begin{Bmatrix}\{0\}\\ \{0\}\\ \{0\}\end{Bmatrix}. \tag{8}$$

The elements of these submatrices are themselves submatrices whose elements are determined by evaluating the preintegrated elastic stiffnesses through the thickness multiplied by the various shape functions or their derivatives as determined by the variational statement. If these submatrices, each of order $(n+1)$, are defined by the subscripts α and β,

TABLE I. Groupings of approximation functions.

Group	Displacement	x	y	z	Group	Displacement	x	y	z
OD	u	O	E	E	OX	u	O	O	O
	v	E	O	E		v	E	E	O
	w	E	E	O		w	E	O	E
EY	u	O	O	E	EZ	u	O	E	O
	v	E	E	E		v	E	O	O
	w	E	O	O		w	E	E	E
EX	u	E	E	E	EV	u	E	O	O
	v	O	O	E		v	O	E	O
	w	O	E	O		w	O	O	E
OY	u	E	E	O	OZ	u	E	O	E
	v	O	O	O		v	O	E	E
	w	O	E	E		w	O	O	O

the corresponding elements can be expressed in a fairly compact form. These are given in the Appendix.

The matrix equations are general and can accommodate approximating functions in (x,y) that are either global and analytic (such as Fourier or power series) or local and exactly evaluated (such as finite-element polynomials). The dependence on the z coordinate has been eliminated by preintegrating, which manifests itself in the matrix equations above. Because of the nature of the approximating functions themselves, the derivatives of the displacement are continuous over only a specific sublayer. If the material is homogeneous, this is still an acceptable approximation even for a subdivided layer because the behavior trends toward a continuous derivative as the number of layers increases. For the case of dissimilar media, the functions allow a break in the slope, which matches physical reality much more accurately than does a global approximation.

For a homogeneous isotropic parallelepiped, Ohno[3] showed that the eigenvalue problem in Eq. (8) can be split into eight smaller problems using symmetry arguments of the displacement-field components and matching these with the appropriate series terms in the approximation functions. These are denoted by the eight groups listed in Table I,[3] where the letters O and E, respectively, denote functions that are odd or even with respect to the appropriate spatial coordinate. For example, power series can be used for each of the displacement functions (such as in the powerful algorithm of Visscher and colleagues[7]). Including six terms in the Pascal-triangle visualization of the approximation functions (i.e., terms up to $x^5y^5z^5$) means a general eigenvalue problem for three unknowns with 6^3 terms in each, or an eigenvalue problem of dimension 648. If symmetry is used, this problem can be split into eight problems of dimension 81, greatly increasing the speed of this computation. This calculation must be completed many times when the elastic constants are being computed, and hence the splitting of the original problem possesses much appeal.

For the layered bimaterial, however, there is no material symmetry about the $x-y$ plane, and hence the splitting operation for this characteristic must be removed. This results not in eight groups, but four, which are the more generalized groups defined in Table I before exploiting symmetry about the $x-y$ plane. We define these four groups and their labeling in Table II.

TABLE II. Group structure for layered isotropic laminate.

Group	Displacement	x	y
1	u	E	E
	v	O	O
	w	O	E
2	u	E	O
	v	O	E
	w	O	O
3	u	O	O
	v	E	E
	w	E	O
4	u	O	E
	v	E	O
	w	E	E

IV. MEASUREMENTS

We measured three materials: nearly texture-free copper, 0.2W/Cu composite, and a three layer 0.2W/Cu–Cu-0.2W/Cu laminate. Using an optical microscope, we measured layer thicknesses of 0.150–0.188-0.166 cm for the laminate. Mass densities were determined using Archimedes' method with distilled water as a standard. The copper contained a significant volume fraction of voids that lowered the expected mass density by about 8%. The composite specimen contained about 1% voids. For the elastic-constants determination, we used URS as described in the previous section and by Migliori and Sarrao[6] for homogeneous materials.

The sending and receiving transducers were poled polycrystalline lead zirconate titanate (PZT) that hold the specimens by their corners diagonally. One transducer transmits continuous sinusoidal waves to the specimen, and the other transducer detects the specimen's displacement response.

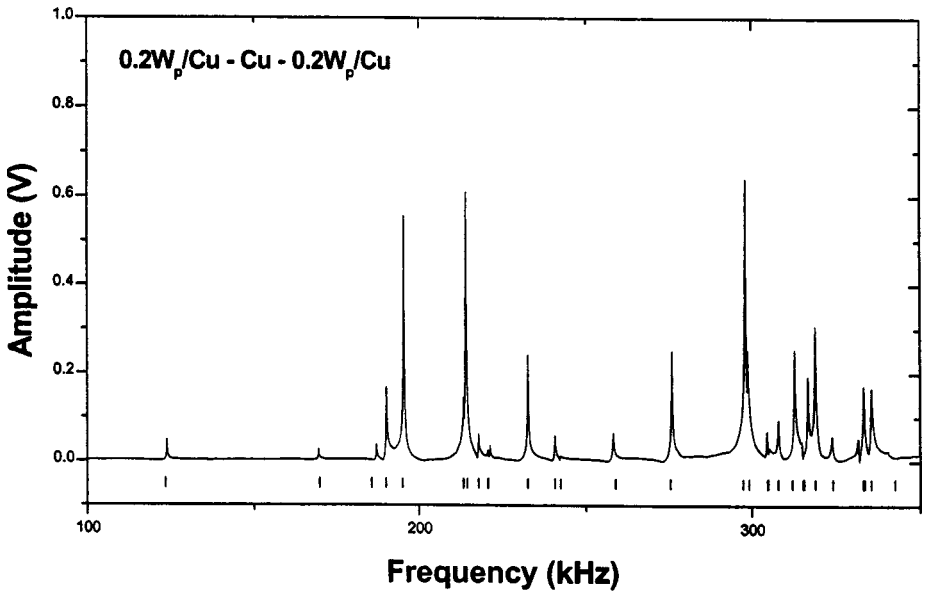


FIG. 2. Measured resonance spectrum of trilayer laminate. Resonant frequencies yield the C_{ij}. The bars at the bottom of the figure indicate resonance frequencies calculated from the deduced C_{ij}.

Their center frequencies were 10 MHz. Their broadband character permits using them far from their resonance frequencies, thus minimizing the piezoelectric-crystal contributions to the systems macroscopic frequencies and damping. Figure 2 shows the frequency-response spectrum for the layered composite. The two separate constituent materials also have a response spectrum, but their individual behavior is used only to compare with the results from our layered model, and hence these are not shown. As discussed above, the resonance frequencies depend on specimen shape, size, mass (or mass density), elastic stiffnesses, and layer thicknesses. Although not reported, internal frictions can be calculated from the half-power width of associated resonance peaks.

V. RESULTS AND DISCUSSION

The general procedure that forms the foundation of using URS to calculate the elastic constants is to iteratively select the elastic constants that give the best fit to the measured frequencies. This is accomplished by minimizing the differences between the measured and experimental frequencies using the Levenburg–Marquardt algorithm.[13] Details of this procedure are described by Migliori and Sarrao.[6]

Two types of models are used to determine the elastic constants of the constituent materials used in this study. The model of most interest is the discrete-layer model described above, which is necessary to represent the through-thickness behavior of the layered material. However, we also use a simpler method to calculate the elastic constants of each of the two constituent materials when these materials appear in a homogeneous block. In this case, global basis functions with C^1 continuity can be used since the displacement gradients are continuous within the deformed (vibrating) solid. Our weak form and matrix equations now remain the same, but in this case our z-direction approximations retain the form used for the in-plane approximation functions. This methodology is well developed (see, for example, Refs. 6 and 5) and we will not discuss it here other than to note that we used the same fitting procedure for the experimental frequencies of the homogeneous blocks as we did for the layered material, with a resulting frequency rms error of 0.11 percent for the Cu and 0.08 percent for the W_p/Cu composite. We note that for the homogeneous block, only two elastic parameters are fit to the measured frequencies; for the layered block, there are four.

It is usually necessary to use a number of frequencies at least five times larger than the number of unknown constants to be determined.[6] In this application, we use the lowest 29 frequencies in our inversion scheme for the layered material. Although it is not uncommon to have missing modes in a response spectrum, that is not the case here and our theoretical frequencies show a very strong relationship with the experimental frequencies. We iteratively solve the eigenvalue problem given in Eq. (8) until the differences between the measured and computed frequencies reach a minimum. The initial guess for the elastic constants has no effect on the final values provided they are within a reasonable range (5–10 percent) of the final elastic constants. This point is discussed in more detail by Migliori and Sarrao.[6]

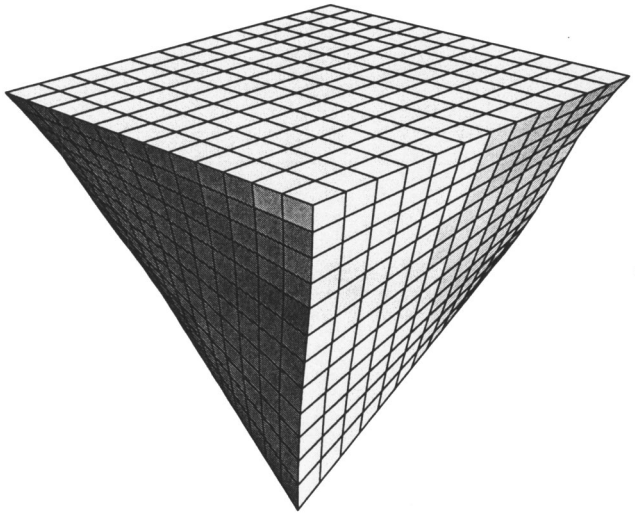


FIG. 3. Modal shape for the lowest mode in group 2 (and lowest overall frequency) for layered bimaterial. Physically, this corresponds to shear in the $x-y$ plane. This is the first nonzero frequency in group 2. This modal group has one rigid-body mode with zero frequency.

The 29 frequencies used in the minimization procedure can be split into four groups as described above. Each of these frequencies has a corresponding modal displacement pattern of the deformation pattern the specimen undergoes as it oscillates at this frequency. In Figs. 3–6 we show the lowest nonzero modal pattern for each of the four groups. For the type of assumed displacement field we have used in our solution methodology, there are six rigid-body modes (three translational and three rotational) that result in a displacement with zero strain energy. These each yield a frequency of 0, and are not included in our results other than to note that groups 1 and 3 possess one each of these modes, with groups 2 and 4 possessing two each. In several of these figures

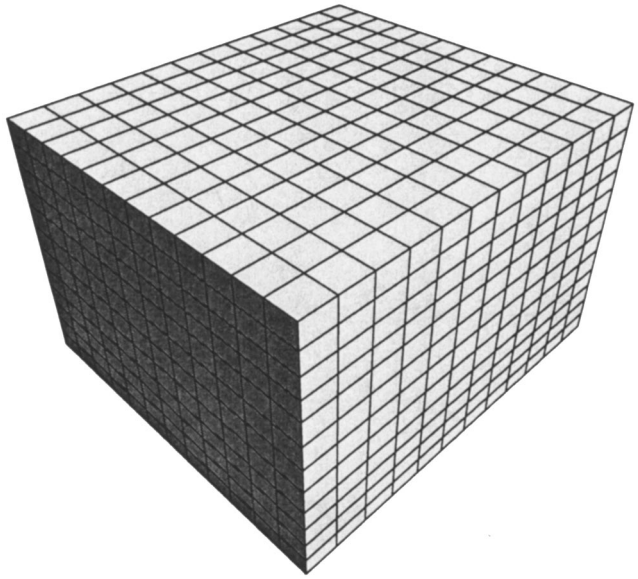


FIG. 4. Modal shape for the lowest mode in group 4 for layered bimaterial. This mode corresponds to the breathing mode associated with the flat faces remaining nearly plane but undergoing uniform deformation across each of the six faces. This is the first nonzero frequency in group 4 which, as in group 2, also has a single rigid-body mode.

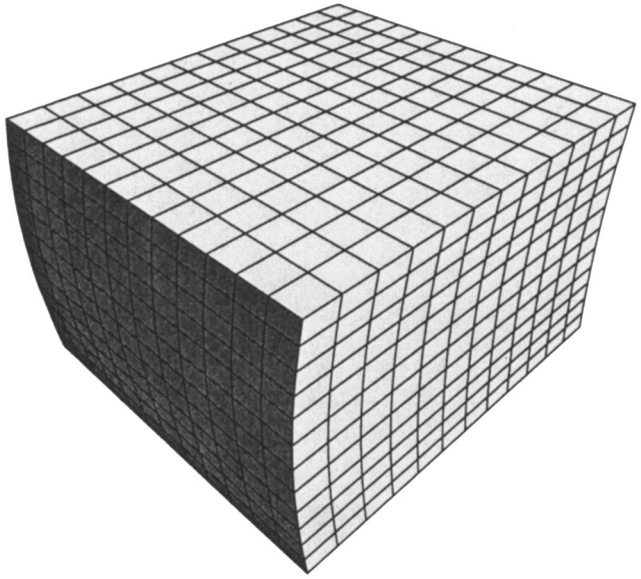


FIG. 5. Modal shape for the lowest mode in group 1 for layered bimaterial. This group has two rigid-body modes.

(where four divisions have been used for each physical layer of the specimen), the modal patterns show the break in the shear strain at a dissimilar material interface, indicating the need for appropriate approximating functions through the thickness.

The final values for the elastic constants of the two constituent materials are given in Table III, and were calculated with a final rms error in frequency of 0.26 percent. In this table, the elastic constants are compared with values calculated using methods described above. Our theoretical model for the layered material assumes that each of the layers is isotropic, and hence we estimate a single C_{11} and C_{44} for each of the two layers. The good fit to the measured frequencies is further demonstrated in Table IV, where we show the

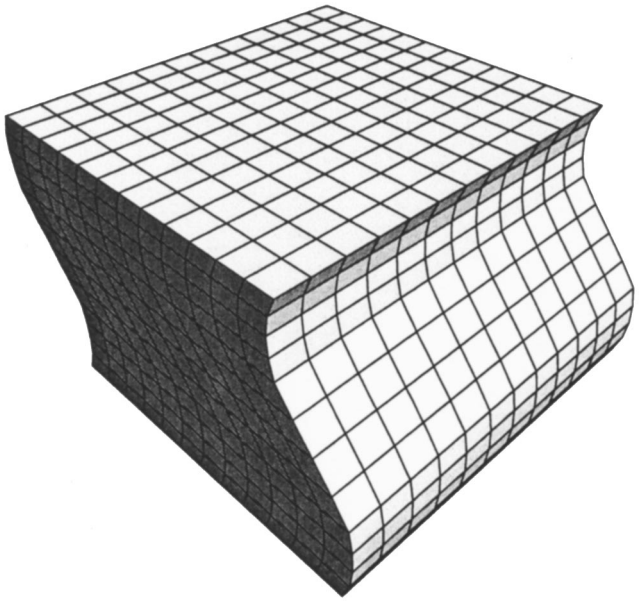


FIG. 6. Modal shape for the lowest mode in group 3, which also has two zero rigid-body frequencies and corresponding modal shapes.

TABLE III. Calculated material properties for Cu and 0.2 W_p/Cu composite. The subscript H denotes that a homogeneous specimen was used to calculate the elastic properties, whereas the subscript L denotes calculation using the layered specimen and discrete-layer theoretical model that is the primary focus of this study. The parameters C_L, B, G and E, are the longitudinal, bulk, shear, and Young moduli, and ν and c are the Poisson ratio and void volume fraction, respectively. All constants are in GPa except ρ, ν, and c.

	Cu (URS_H)	Cu (URS_L)	0.2 W_p/ Cu (URS_H)	0.2 W_p/ Cu (URS_L)
ρ (g/cm^3)	8.2348		10.8954	
C_{11}	155.71		239.74	
C_{22}	157.79	154.88	240.30	242.18
C_{33}	155.31		233.23	
C_{44}	40.001		60.015	
C_{55}	39.953	40.25	59.985	60.31
C_{66}	40.175		60.845	
C_{12}	76.518		118.51	
C_{13}	75.733	74.37	116.83	121.56
C_{23}	76.557		116.95	
C_L	156.30		237.80	
B	102.93		157.52	
G	40.025		60.213	
E	106.30		160.22	
ν	0.3279		0.3305	
c	0.078		0.009	

measured frequencies compared with the theoretical frequencies as calculated using the final elastic constants for the two layers. Also shown in this table are the frequencies that would result if the parallelepiped were a homogeneous block of either of the two materials, with block H1 denoting a homogeneous block of 20 percent tungsten and H2 a homogeneous block of pure copper. Lack of significant anisotropy in the elastic constants shown in Table IV reveals nearly texture-free materials, as expected for powder-metallurgy preparation.

Our results for C_{11} lie within 0.9% of those of the average equivalent moduli for the Cu and 1.9% of the W_p/Cu composite. Our values for C_{44} are within 0.6% and 0.05% for the same two materials, respectively. It is quite possible that for a block with such a low aspect ratio, modes involving shear deformation are more dominant, leading to better agreement for the shear moduli than that of C_{11}. Other natural sources of error include the precise nature of the dissimilar-layer bond and our treating each of the materials as isotropic in our layered model.

Our ability to extract the elastic constants of individual layers within a laminate could prove to be significant for certain types of materials for which a homogeneous specimen may be difficult to procure. Thin films on substrates or natural multilayered solids could potentially be studied using this approach without having to separate the materials for individual examination. Though not without limitations, we have shown for the first time that ultrasonic-resonance spectroscopy methods can be applied to layered systems with reasonably good results. Applications to more complex systems await future study.

TABLE IV. Modal groups and frequencies (in Hz) for laminated and homogeneous blocks.

Mode	ω_{exp}	ω_{lay}	G_{lay}	ω_{H1}	G_{H1}	ω_{H2}	G_{H2}
1	123 942	123 427	2	129 045	2	105 420	2
2	169 557	169 834	4	175 576	4	143 242	4
3	187 150	185 726	2	203 567	1	165 928	1
4	190 199	189 992	1	207 530	2	169 501	2
5	195 462	195 012	2	211 099	2	172 448	2
6	213 480	213 429	3	225 387	4	183 066	4
7	214 272	214 545	4	232 775	4	190 131	4
8	218 005	217 919	4	237 165	1	193 545	1
9	220 732	220 782	1	237 460	3	193 662	3
10	221 446	220 885	3	238 123	3	194 436	3
11	232 966	232 829	1	258 947	3	211 037	3
12	241 005	240 977	4	260 030	1	212 193	1
13	242 515	242 622	3	262 221	4	213 874	4
14	258 394	259 043	4	287 736	4	234 620	4
15	276 146	275 605	1	297 445	1	242 835	1
16	298 201	297 474	2	323 122	1	263 872	1
17	298 964	299 177	1	324 685	3	265 165	3
18	304 732	304 759	3	326 833	3	266 848	3
19	305 277	305 047	3	327 795	2	267 450	2
20	308 197	308 117	2	330 751	4	269 744	4
21	313 024	312 235	4	350 621	1	285 526	1
22	315 030	315 415	3	351 051	3	285 749	4
23	316 972	315 811	4	351 535	4	285 912	3
24	319 186	319 121	1	351 595	2	286 695	2
25	324 133	324 330	4	359 377	4	292 252	4
26	331 772	333 170	1	363 043	4	296 226	4
27	333 451	333 674	4	370 136	2	302 195	2
28	335 782	335 629	3	372 100	4	303 725	4
29	340 605	342 694	4	381 167	1	310 452	1

VI. CONCLUSIONS

(1) Discrete-layer models that account for kinks in displacement gradients across a dissimilar-material interface are crucial in obtaining accurate theoretical frequency predictions and subsequent estimates of elastic constants.
(2) Ultrasonic-resonance spectroscopy can be applied to layered isotropic laminates with good accuracy. For a trilayer composite, we found matching of the first 29 frequencies yielding an rms error of 0.26 percent between theoretical and measured frequencies. There were no missing modes.
(3) Layer elastic constants of the two dissimilar materials agree with elastic constants computed using homogeneous specimens within 0.9 percent for the Cu and 1.9 percent for the W_p/Cu composite for C_{11}, and within 0.6 and 0.05 percent, respectively, for C_{44}.

ACKNOWLEDGMENTS

P.R.H. acknowledges support of the Alexander von Humboldt Research Foundation in Germany. I.R. acknowledges support from the U.S. Army Research Office, Grant No. DAAD19-01-1-0590.

APPENDIX:

The entries in the element coefficient matrices can be expressed as

$$[K^{11}]_{\alpha\beta}=\int_A\left[[A^{11}]\frac{\partial\Psi^u_\alpha}{\partial x}\frac{\partial\Psi^u_\beta}{\partial x}+[D^{55}]\Psi^u_\alpha\Psi^u_\beta+[A^{66}]\frac{\partial\Psi^u_\alpha}{\partial y}\frac{\partial\Psi^u_\beta}{\partial y}\right]dx\,dy, \tag{A1}$$

$$[K^{12}]_{\alpha\beta}=\int_A\left[[A^{12}]\frac{\partial\Psi^u_\alpha}{\partial x}\frac{\partial\Psi^v_\beta}{\partial y}+[A^{66}]\frac{\partial\Psi^u_\alpha}{\partial y}\frac{\partial\Psi^v_\beta}{\partial x}\right]dx\,dy, \tag{A2}$$

$$[K^{13}]_{\alpha\beta}=\int_A\left[[B^{13}]\frac{\partial\Psi^u_\alpha}{\partial x}\Psi^w_\beta+[\bar{B}^{55}]\Psi^u_\alpha\frac{\partial\Psi^w_\alpha}{\partial x}\right]dx\,dy, \tag{A3}$$

$$[K^{22}]_{\alpha\beta}=\int_A\left[[A^{22}]\frac{\partial\Psi^v_\alpha}{\partial y}\frac{\partial\Psi^v_\beta}{\partial y}+[D^{44}]\Psi^v_\alpha\Psi^v_\beta+[B^{24}]^T\Psi^v_\alpha\frac{\partial\Psi^v_\beta}{\partial y}+[A^{66}]\frac{\partial\Psi^v_\alpha}{\partial x}\frac{\partial\Psi^v_\beta}{\partial x}\right]dx\,dy, \tag{A4}$$

$$[K^{23}]_{\alpha\beta}=\int_A\left[[B^{23}]\frac{\partial\Psi^v_\alpha}{\partial y}\Psi^w_\beta+[\bar{B}^{44}]\Psi^v_\alpha\frac{\partial\Psi^w_\alpha}{\partial y}\right]dx\,dy, \tag{A5}$$

$$[K^{33}]_{\alpha\beta} = \int_A \left[[A^{44}] \frac{\partial \Psi_\alpha^w}{\partial y} \frac{\partial \Psi_\beta^w}{\partial y} + [D^{33}] \Psi_\alpha^w \Psi_\beta^w + [A^{55}] \frac{\partial \Psi_\alpha^w}{\partial x} \frac{\partial \Psi_\beta^w}{\partial x} \right] dx\, dy. \tag{A6}$$

The submatrices listed above are computed by preintegrating the functions in z. For any general approximation through the thickness, these can be expressed as

$$A_{ij}^{km} = \sum_{l=1}^{N} \int_{z_l}^{z_{l+1}} C_{km} \bar{\Psi}_i(z) \bar{\Psi}_j(z) dz, \tag{A7}$$

$$B_{ij}^{km} = \sum_{l=1}^{N} \int_{z_l}^{z_{l+1}} C_{km} \bar{\Psi}_i(z) \frac{d\bar{\Psi}_j(z)}{dz} dz, \tag{A8}$$

$$\bar{B}_{ij}^{km} = \sum_{l=1}^{N} \int_{z_l}^{z_{l+1}} C_{km} \bar{\Psi}_i(z) \bar{\Psi}_j(z) dz, \tag{A9}$$

$$D_{ij}^{km} = \sum_{l=1}^{N} \int_{z_l}^{z_{l+1}} C_{km} \frac{d\bar{\Psi}_i(z)}{dz} \frac{d\bar{\Psi}_j(z)}{dz} dz. \tag{A10}$$

[1] E. P. Eer Nisse, "Variational method for electroelastic vibration analysis," IEEE Trans. Sonics Ultrason. **SU-14**, 153–160 (1967).

[2] H. H. Demarest, Jr., "Cube resonance method to determine the elastic constants of solids," J. Acoust. Soc. Am. **49**, 768–775 (1971).

[3] I. Ohno, "Free vibration of a rectangular parallelepiped crystal and its application to determination of elastic constants of orthorhombic crystals," J. Phys. Earth **24**, 355–379 (1976).

[4] P. R. Heyliger, A. Jilani, H. Ledbetter, R. Leisure, and C.-K. Wang, "Elastic constants of isotropic cylinders using resonant ultrasound," J. Acoust. Soc. Am. **94**, 1482–1487 (1995).

[5] H. Ledbetter, C. Fortunko, and P. Heyliger, "Orthotropic elastic constants of a boron–aluminum fiber-reinforced composite: An acoustic-resonance-spectroscopy study," J. Appl. Phys. **78**, 1542–1546 (1995).

[6] A. Migliori and J. L. Sarrao, *Resonant Ultrasound Spectroscopy: Applications to Physics, Materials Measurements and Non-Destructive Evaluation* (Wiley, New York, 1997).

[7] W. M. Visscher, A. Migliori, T. M. Bell, and R. A. Reinert, "On the normal modes of free vibration of inhomogeneous and anisotropic elastic objects," J. Acoust. Soc. Am. **90**, 2154–2162 (1991).

[8] P. Heyliger, "Traction-free vibration of layered elastic and piezoelectric parallelepipeds," J. Acoust. Soc. Am. **107**, 1235–1245 (2000).

[9] K. E. Pauley and S. B. and Dong, "Analysis of Plane Waves in Laminated Piezoelectric Plates," Wave Electron. **1**, 265–285 (1976).

[10] J. N. Reddy, "A generalization of displacement-based laminate theories," Commun. Appl. Numer. Methods **3**, 173–181 (1987).

[11] Atlantic Equipment Engineering, Bergenfield, NJ. Products or companies named here are cited only in the interest of complete scientific description, and neither constitute nor imply endorsement by NIST or by the US government. Other products may be found to serve just as well.

[12] Goodfellow Cambridge Ltd., Huntington, England.

[13] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, 1986), Chap. 15.

[14] H. F. Tiersten, *Linear Piezoelectric Plate Vibrations* (Plenum, New York, 1969).

Basic types of discontinuity in circular acoustic wave guide

Nick Solokhin[a)]
Ultrasonic S-Lab, 1164 Ramer Court, Concord, California 94520

(Received 30 March 2003; accepted for publication 8 September 2003)

It is considered the interaction of acoustic waves with two basic discontinuities—''diameter step'' and ''diaphragm with coaxial opening''—in a cylindrical acoustic waveguide with a rigid wall. The problem is considered and solved in rigorous formulation for the case of axial symmetry and below first radial resonance (f_{r1}) in the waveguide. Obtained data demonstrate that a one-dimensional (1-D) model provides good results for the ''diameter step'' if the frequency does not exceed 18%–50% of f_{r1}. Calculated data for the ''diameter step'' demonstrate that this discontinuity follows the reciprocity principle (below f_{r1})—the parameter of nonreciprocity does not exceed calculation errors. Data for the ''diaphragm'' demonstrate good agreement with the Lamb formula for a rigid screen with parallel slots—a 2-D problem in the Cartesian coordinates. Calculation errors for the both discontinuities were evaluated with the energy conservation law. The difference between sum of power fluxes in reflected and transmitted waves and the power flux in the incident wave does not exceed 1%. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1621862]

PACS numbers: 43.20.Mv, 43.20.Bi [MO] Pages: 2626–2632

I. INTRODUCTION

Two types of discontinuities in a circular acoustic waveguide with a rigid wall—will be considered—see Fig. 1. These discontinuities have important applications in acoustics and attract attention as objects of research. The first such object is the ''interaction of acoustic waves with periodic rigid grating.''[1] At normal incidence this problem may be transformed to a corresponding discontinuity in a waveguide with rigid walls. Another interesting object is an ''artificial matching layer''[2]—a periodic grating with final thickness that may be placed before a radiating transducer to improve radiation into a medium with low acoustic impedance (gas). One more example: a diameter step in a circular waveguide as an acoustic transformer—see Ref. 3, Chap. 17. Such an acoustic transformer was used in old speakers (in 1930 years) to improve radiation into air.

Any correct consideration of the acoustic wave problem in a waveguide is based on acoustic modes of such a waveguide. Fortunately, this problem has already been considered in rigorous formulation for a cylindrical waveguide with a rigid and soft wall—see, for example, Refs. 3 and 4. Some details of waveguide modes that are needed to proceed further will be briefly considered below.

A discontinuity ''diameter step'' can be considered in the frame 1-D model, and such an approximate approach (because a 1-D model is fundamentally approximate for such a problem) provides simple and well-known expressions—see, for example, Refs. 3 and 4—for reflection (R_0) and transmission (T_0) coefficients,

$$R_0=\frac{S_1-S_2}{S_1+S_2}, \quad T_0=\frac{2S_1}{S_1+S_2}, \tag{1}$$

where S_1 and S_2 are areas of a cross section in a region with the incident wave (the left section at $z<0$) and in the region with the transmitted wave (the right section at $z>0$), respectively. We may expect a high accuracy of these expressions at frequency $f\to 0$ and a worsening of accuracy as f approaches the first radial resonance. Previous works made significant contributions to the problem, but they applied some assumptions to simplify the calculation procedure. Let us consider examples of such assumptions in basic previous works: ''the gradient of the pressure has only axial component''[6] introduction of an ''equivalent'' electric circuit with lumped elements to simulate a ''diameter step,''[6–8] the introduction of ''discontinuity inductance.''[7,8] Assumptions can reduce calculation problems, but create other problems: difficulties with the evaluation of errors. The current situation remains as follows: ''there are no tables allowing the engineer to compute easily the effect of these discontinuities.''[8]

A discontinuity ''rigid diaphragm with opening'' cannot be considered in the frame 1-D model at all. First a good result has been obtained by Lamb[1] in the low-frequency case. Actually, Lamb considered the interaction of acoustic waves with periodic grating of slots in a thin rigid screen at normal incidence. This problem may be correctly transformed into the problem for a waveguide with a rigid diaphragm—see Fig. 2 and some recent works.[2,5] Lamb considered this 2-D problem in the Cartesian coordinates x, y and without dependence on coordinate z (the wave travels along the x axis). The period of the grating was $2a$; the width of the slots was $2b$. Hence, the width of an equivalent waveguide should be $2a$ and the width of the diaphragm opening should be $2b$. The Lamb formula[1] for a reflection coefficient is

$$R=\frac{ik\Delta}{1+ik\Delta}, \quad \Delta=\frac{2a}{\pi}\ln\left[1\Big/\sin\left(\frac{\pi b}{2a}\right)\right], \tag{2}$$

where k is the wave number in the acoustic medium. The above formula may be applied only for the first traveling mode in the waveguide (below first resonance along width

a)Electronic mail: solokhin@comcast.net

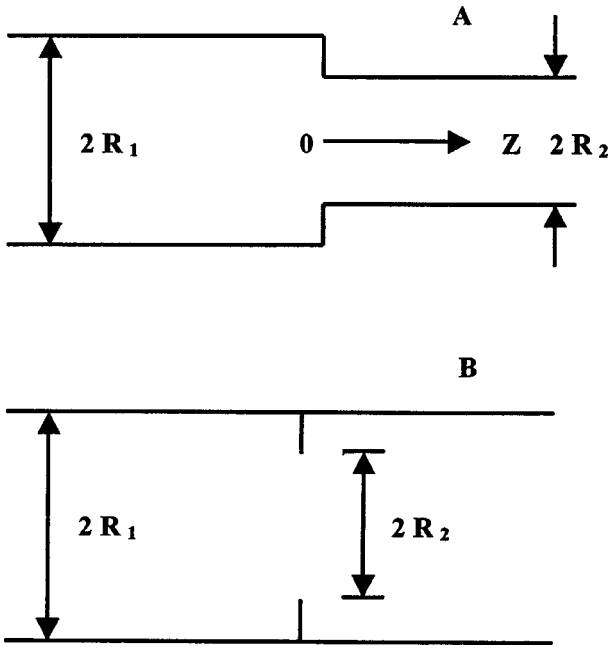


FIG. 1. Two basic types of discontinuity in an acoustic waveguide with a rigid wall: diameter step (a) and rigid diaphragm with a coaxial opening (b).

$2a$). The expression in form (2) provides the correct complex value for R in a system with positive time—$\exp(i\omega t)$—see Note 1 for clarification.

It needs to consider some details about modes of a circular waveguide with a rigid wall to proceed further with the analysis of discontinuities. A traditional way is to start with

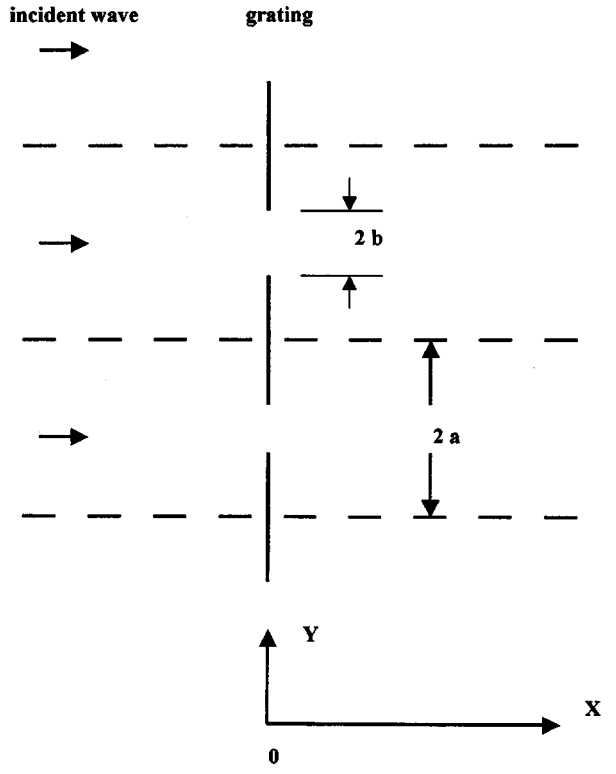


FIG. 2. The 2-D grating of slots in a rigid screen at normal incidence (the Z axis is perpendicular to the sketch plane). Rigid walls, that do not provide any disturbance to the acoustic field, are shown by dotted lines. Such walls create a system of equivalent waveguides.

the wave equation (in a cylindrical coordinate system) for the velocity potential Φ. After the separation of variables [as it was suggested by Fourier: $\Phi(r,z)=R(r)Z(z)$] we come to different equations for two coordinate functions, $R(r)$ and $Z(z)$. Solutions of these equations are well known,

$$R(r)=AJ_0(\alpha r), \quad Z(z)=C\exp(-i\beta z)+D\exp(i\beta z). \tag{3}$$

Radial and axial wave numbers have the relation: $\alpha^2+\beta^2=k^2$; $k=\omega/c$. Wave velocity in an unbounded acoustic medium is c. We consider a waveguide with a rigid wall at $r=R_1$, hence, there should be a boundary condition: a radial component of velocity must be zero on this surface. Velocity V in the acoustic medium,

$$V=-\mathrm{grad}(\Phi)=-e_r\frac{\partial\Phi}{\partial r}-e_z\frac{\partial\Phi}{\partial z}, \tag{4}$$

where e_r and e_z are the unit vectors in the radial and axial direction, respectively. A mentioned boundary condition at $r=R_1$ defines an infinite sequence of radial wave numbers,

$$\alpha R_1=\gamma_m \quad (m=0,1,2,...), \quad \gamma_0=0, \quad \gamma_1=3.832, \quad \gamma_2=7.016,..., \tag{5}$$

that leads to a corresponding sequence for axial wave numbers β_m,

$$\beta_m=\sqrt{k^2-\alpha_m^2}; \quad \alpha_m=\frac{\gamma_m}{R_1}. \tag{6}$$

So, the mode can exist as a traveling mode if the difference under the root in (6) is positive (a zero value corresponds to the radial resonance). Hence, the mode with $\alpha=\alpha_0=0$ can exist always and it exists as a single traveling mode in the frequency range from zero to the first radial resonance,

$$0<f<\frac{\gamma_1 c}{2\pi R_1}=\frac{3.832c}{2\pi R_1}. \tag{7}$$

So, we are ready to consider both types of discontinuity and do it as the boundary-value problems and without any simplifying assumption—see below.

II. DISCONTINUITY "DIAMETER STEP"

The geometry of this structure is shown in Fig. 1(a). Such a structure may exist in two versions: "transmission into smaller tube" and "transmission into larger tube." An analysis of these two versions is not absolutely identical and we start here with a case "transmission into smaller tube." The structure has axial symmetry and the axis z is directed along the axis of symmetry. The diameter step is located at $z=0$. There are two different regions in the structure: region 1 at $z\leqslant 0$ with outer radius R_1 and region 2 at $z\geqslant 0$ with outer radius R_2. Let us consider (briefly) the calculation procedure we are going to use. Solutions have to be built for each region. The solution in the first region (Φ_1) contains an incident wave and a reflected wave. The solution in the second region (Φ_2) contains only a transmitted wave. Hence, reflected and transmitted waves follow the "Sommerfeld radiation condition." These two solutions have to be "sewed" at junction ($z=0$) to get continuity for the potential and its

first axial derivative. Sewing creates a system of linear equations. This procedure is used for both types of discontinuity and without any simplifying assumption. The same procedure may be used for more complex cases with several traveling modes.

So, ''transmission into smaller tube.'' The problem is considered in frequency range (7): below first radial resonance in region 1. The incident wave (the single traveling mode) has unit amplitude and goes from left to right in region 1. Discontinuity generates a reflected wave (with amplitude R) in region 1 and a transmitted wave (with amplitude T) in region 2. So, an expression for the potential in the region 1 can be written—see (3)–(6)—in the form

$$\Phi_1(r,z)=e^{-ikz}+\mathrm{Re}^{ikz}+\sum_m A_m e^{\beta_m z}J_0(\alpha_m r),$$

$$\alpha_m=\frac{\gamma_m}{R_1},\quad \beta_m=\sqrt{\alpha_m^2-k^2},\quad m=1,2,\ldots. \tag{8}$$

The dependence on time has been chosen in the form $\exp(i\omega t)$—with positive time. β_m is the axial wave number (infinite sequence). An infinite sum presents local inhomogeneous waves near the junction.

Note 1: It needs to be noted that the wave equation allows us to use negative time $[\exp(-i\omega t)]$ as well as positive $[\exp(i\omega t)]$. Hence, there are two different possible systems to build a solution: with positive time or with negative time. There is a very important point: reflection and transmission coefficients for the same structure demonstrate complex conjugated values in these systems. Therefore, it needs to be very accurate to compare calculated data from different or not identified systems. For example, the cited Lamb formula (2) has been presented here in the system with positive time, but the reader can find it in conjugated form[5] in the system with negative time.

The velocity potential in the region 2 can be written—see (3)–(6)—in the form

$$\Phi_2(r,z)=Te^{-ikz}+\sum_n B_n e^{-\delta_n z}J_0(\epsilon_n r),$$

$$\epsilon_n=\frac{\gamma_n}{R_2},\quad \delta_n=\sqrt{\epsilon_n^2-k^2},\quad n=1,2,\ldots. \tag{9}$$

To proceed further we need expressions for axial velocity $[\nu(r)]$ and velocity potential $[f(r)]$ in both regions at a junction (at $z=0$). So, in the region 1 (at $z=-0$) we have

$$f_1(r)=1+R+\sum_m A_m J_0(\alpha_m r),$$

$$\nu_1(r)=ik(1-R)-\sum_m A_m\beta_m J_0(\alpha_m r). \tag{10}$$

In the region 2 (at $z=+0$),

$$f_2(r)=T+\sum_n B_n J_0(\epsilon_n r),$$

$$\nu_2(r)=ikT+\sum_n B_n\delta_n J_0(\epsilon_n r). \tag{11}$$

The boundary conditions in the region 1 (at the junction) have a mixed form: zero axial velocity on the rigid wall and continuity of the potential and axial velocity in the pass,

$$(R_2<r<R_1),\quad \nu_1(r)=0, \tag{12}$$

$$(0<r<R_2),\quad f_1(r)=f_2(r),$$

$$\nu_1(r)=\nu_2(r). \tag{13}$$

Condition (12) is very simple and may be satisfied in different ways, for example, with the collocation method,

$$\sum_m A_m\beta_m J_0(\alpha_m r_j)=ik(1-R)\quad (R_2<r_j<R_1). \tag{14}$$

Conditions (13) are not so simple and may be satisfied in different ways as well. Actually, expressions (13) represent physical properties of hydrodynamic flow at the junction. Let us consider the possible mathematical formulation of such conditions. Functions $J_0(\epsilon_n r)$ form an orthogonal system in the pass—over the interval $(0,R_2)$. Hence, coefficients B_n can be obtained by direct decomposition of f_1 and ν_1 in the Fourier–Bessel series. This routine mathematical operation provides continuity ''automatically.'' But there is one very important point here: we can obtain coefficients B_n in two different ways—from the continuity of the potential (B_n^f) and from the continuity of axial velocity (B_n^ν). So, we came to conditions of compatibility in the form

$$B_n^f=B_n^\nu,\quad n=1,2,\ldots, \tag{15}$$

$$T^f=T^\nu. \tag{16}$$

Conditions (15)–(16) may be considered as a possible mathematical form of conditions (13). Let us consider it in detail. Continuity of the potential provides

$$1+R+\sum_{m=1}^{\infty} A_m J_0(\alpha_m r)=T+\sum_{n=1}^{\infty} B_n^f J_0(\epsilon_n r). \tag{17}$$

Decomposition with the Hankel transform over $0<r<R_2$,

$$B_n^f=\frac{1}{\int_0^{R_2}J_0^2(\epsilon_n r)r\,dr}\sum_{m=1}^{\infty} A_m\int_0^{R_2}J_0(\alpha_m r)J_0(\epsilon_n r)r\,dr. \tag{18}$$

Continuity of the axial velocity provides

$$ik(1-R)-\sum_{m=1}^{\infty} A_m\beta_m J_0(\alpha_m r)$$

$$=ikT+\sum_{n=1}^{\infty} B_n^\nu\delta_n J_0(\epsilon_n r). \tag{19}$$

Decomposition over $0<r<R_2$ provides

$$B_n^\nu=\frac{-1}{\delta_n\int_0^{R_2}J_0^2(\epsilon_n r)r\,dr}\sum_{m=1}^{\infty} A_m\beta_m\int_0^{R_2}J_0(\alpha_m r)$$

$$\times J_0(\epsilon_n r)r\,dr. \tag{20}$$

Substituting (18) and (20) into (15), we come to a condition of compatibility in the form

$$\sum_m A_m\left(1+\frac{\beta_m}{\delta_n}\right)\int_0^{R_2} J_0(\alpha_m r)J_0(\epsilon_n r)r\,dr=0$$

$$(n=1,2,...). \quad (21)$$

Expression (21) is equivalent of condition (15) and now may be directly used in calculation practice.

Note 2. Conditions of compatibility: If a solution is built with variable parameter R, both conditions (15)–(16) cannot be imposed for any arbitrary R, because it can create contradiction with the energy conservation law. The energy conservation law for this case is: power flux in the incident wave equal to the sum of power fluxes in the transmitted and reflected waves. Mathematically it may be written in the form

$$|R|^2+\frac{R_2^2}{R_1^2}|T|^2-1=0. \quad (22)$$

The energy conservation law will not be violated if we impose condition (15) only. Condition (16) can be used as a criterion to check choice for the variable parameter R: the better choice for R the better accuracy of (16). Condition (16) can be realized with following expressions:

$$T^f=\frac{2}{R_2^2}\int_0^{R_2} f_1(r)r\,dr, \quad T^\nu=\frac{2}{ikR_2^2}\int_0^{R_2}\nu_1(r)r\,dr. \quad (23)$$

So, the calculation procedure ''diameter step—transmission into smaller tube''—is built in the following way. The system of linear equations to calculate coefficients A_m contains Eqs. (14) and (21) with variable parameter R. Coefficients A_m are calculated for any R value, but only a correct R values meets with the criterion (16). If we obtain an R value that meets with criterion (16), it is accompanied by a satisfaction of energy conservation law in the form (22). Small incompliance in (22) may be considered as an evaluation of an existing calculation error.

Let us consider another case of discontinuity diameter step: ''transmission into larger tube'' $(R_1<R_2)$. The condition for frequency should be in the form

$$0<f<\frac{\gamma_1 c}{2\pi R_2}=\frac{3.832c}{2\pi R_2}, \quad (24)$$

because we assume the existence of only one traveling mode in both regions. Expressions for the potential in both regions are the same with the previous case—see (8) and (9), but boundary conditions at the junction must be changed to the form

$$(R_1<r<R_2), \quad \nu_2(r)=0, \quad (25)$$

$$(0<r<R_1), \quad f_1(r)=f_2(r), \quad \nu_1(r)=\nu_2(r). \quad (26)$$

The boundary conditions (25)–(26) have a mixed form in the region 2. We need to build a system of linear equations (to calculate unknown coefficients B_n) in the region 2. The variable parameter is T. So, condition (25) provides the expressions

TABLE I. Diameter step: transmission into a smaller tube. $R_2=0.71$ mm ($R_0=0.330$, $T_0=1.330$).

f, kHz	f/f_{r1}	R	T	Error
20	0.0964	$0.330+0.015i$	$1.329-0.031i$	0.0002
40	0.1929	$0.331+0.031i$	$1.327-0.062i$	0.0004
60	0.2893	$0.333+0.047i$	$1.323-0.094i$	0.0002
80	0.3857	$0.336+0.064i$	$1.317-0.127i$	0.0003
100	0.4822	$0.340+0.083i$	$1.309-0.165i$	0.0005
120	0.5786	$0.346+0.104i$	$1.297-0.206i$	0.0007
140	0.6750	$0.355+0.128i$	$1.280-0.254i$	0.0004
160	0.7715	$0.369+0.158i$	$1.251-0.313i$	0.0008
180	0.8679	$0.395+0.199i$	$1.199-0.395i$	0.0006
190	0.9161	$0.420+0.229i$	$1.150-0.454i$	0.0006

$$\sum_n B_n\delta_n J_0(\epsilon_n r_j)=-ikT \quad (R_1<r_j<R_2). \quad (27)$$

We employ the collocation method, like it has been done before—see (14). Using the same integral transformation [see (17)–(21)], we can obtain a condition of compatibility for the considering case: $A_m^f-A_m^\nu=0$,

$$\sum_n B_n\left(1+\frac{\delta_n}{\beta_m}\right)\int_0^{R_1} J_0(\alpha_m r)J_0(\epsilon_n r)r\,dr=0$$

$$(m=1,2,...). \quad (28)$$

The criterion of accuracy for this case has the form

$$R^f-R^\nu=0,$$

$$R^f=-1+\frac{2}{R_1^2}\int_0^{R_1} f_2(r)r\,dr, \quad (29)$$

$$R^\nu=1-\frac{2}{ikR_1^2}\int_0^{R_1}\nu_2(r)r\,dr.$$

The system of linear equations to calculate unknown coefficients B_n is built with expressions (27) and (28). The criterion (29) directs calculations to correct the T value. The energy conservation law in the form (22) is used to check the calculation error.

Calculated data for the ''diameter step—transmission into smaller tube''—are in Tables I–III. $R_1=1.0$ mm, wave velocity $c=340$ m/s (the air at room conditions). First radial resonance in the region 1 is at $f_{r1}=207.4$ kHz. Parameter ''error'' in the tables is the value on the right side of expres-

TABLE II. Diameter step: transmission into a smaller tube. $R_2=0.50$ mm ($R_0=0.600$, $T_0=1.600$).

f, kHz	f/f_{r1}	R	T	Error
20	0.0964	$0.601+0.018i$	$1.597-0.073i$	0.0006
40	0.1929	$0.603+0.037i$	$1.587-0.149i$	0.0009
60	0.2893	$0.608+0.056i$	$1.568-0.224i$	0.0003
80	0.3857	$0.615+0.076i$	$1.540-0.304i$	0.0005
100	0.4822	$0.625+0.097i$	$1.500-0.388i$	0.0002
120	0.5786	$0.639+0.119i$	$1.443-0.475i$	0.0004
140	0.6750	$0.660+0.143i$	$1.361-0.571i$	0.0002
160	0.7715	$0.691+0.168i$	$1.235-0.671i$	0.0005
180	0.8679	$0.745+0.192i$	$1.021-0.769i$	0.0008
190	0.9161	$0.790+0.200i$	$0.840-0.799i$	0.0005

TABLE III. Diameter step: transmission into a smaller tube. $R_2=0.40$ mm ($R_0=0.724$, $T_0=1.724$).

f, kHz	f/f_{r1}	R	T	Error
20	0.0964	$0.725+0.014i$	$1.720-0.090i$	0.0003
40	0.1929	$0.727+0.029i$	$1.705-0.182i$	0.0005
60	0.2893	$0.731+0.044i$	$1.680-0.276i$	0.0007
80	0.3857	$0.738+0.059i$	$1.640-0.369i$	0.0010
100	0.4822	$0.746+0.075i$	$1.587-0.468i$	0.0006
120	0.5786	$0.758+0.091i$	$1.511-0.568i$	0.0007
140	0.6750	$0.775+0.107i$	$1.405-0.670i$	0.0001
160	0.7715	$0.800+0.123i$	$1.250-0.770i$	0.0003
180	0.8679	$0.841+0.136i$	$0.995-0.850i$	0.0006
190	0.9161	$0.873+0.137i$	$0.795-0.860i$	0.0010

sion (22). R and T values, calculated with expressions (1), are denoted as R_0 and T_0 and they demonstrate perfect consistency with the calculated data in Tables I–III (at lowest frequency).

Calculated data for the ''diameter step—transmission into larger tube''—are below in Table IV. $R_2=1.0$ mm, wave velocity $c=340$ m/s (the air at room conditions). First, radial resonance in the region 2 is at $f_{r2}=207.4$ kHz. Parameter ''error'' in the table is a result of a check with the energy conservation law: a value on the right side of expression (22). R and T values, calculated with expressions (1), are denoted as R_0 and T_0, and they demonstrate perfect consistency with the calculated data at lowest frequency.

III. DISCONTINUITY "RIGID DIAPHRAGM WITH COAXIAL OPENING"

Geometry of this coaxial structure is shown in Fig. 1(b). The structure has two regions with the same outer radius R_1: region 1 at $z<0$ and region 2 at $z>0$. The radius of the opening in the diaphragm is R_2 ($R_2<R_1$). An incident wave with unit amplitude travels in the region 1 from left to right and generates a reflected wave (with amplitude R) in the region 1 and transmitted wave (with amplitude T) in the region 2. Expressions for the potential and axial velocity can be borrowed from (8)–(11). There are several important details here

(1) Both regions have the same outer radius that leads to

$$\alpha_m=\epsilon_m, \quad \beta_m=\delta_m, \quad m=n. \tag{30}$$

$$(2)\quad \nu_1(r)=\nu_2(r)\rightarrow T=1-R, \quad B_m=-A_m. \tag{31}$$

TABLE IV. Diameter step: transmission into a larger tube. $R_1=0.5$ mm, $R_2=1.00$ mm, $c=340$ m/s ($R_0=-0.600$, $T_0=0.400$).

f, kHz	f/f_{r1}	R	T	Error
20	0.0964	$-0.597+0.074i$	$0.399-0.019i$	0.001
40	0.1929	$-0.587+0.148i$	$0.397-0.037i$	0.003
60	0.2893	$-0.568+0.225i$	$0.392-0.056i$	0.0005
80	0.3857	$-0.540+0.304i$	$0.385-0.076i$	0.0005
100	0.4822	$-0.500+0.387i$	$0.375-0.097i$	0.006
120	0.5786	$-0.444+0.475i$	$0.361-0.119i$	0.001
140	0.6750	$-0.361+0.570i$	$0.340-0.143i$	0.003
160	0.7715	$-0.235+0.671i$	$0.309-0.168i$	0.002
180	0.8679	$-0.021+0.768i$	$0.255-0.192i$	0.003
190	0.9161	$+0.160+0.798i$	$0.210-0.200i$	0.003

Relations (31) have been obtained in Ref. 5. Now we can continue further with (31) and energy conservation law to get the relation between the real and imaginary parts of R,

$$\mathrm{Im}(R)=\sqrt{\mathrm{Re}(R)-(\mathrm{Re}(R))^2} \quad [0<\mathrm{Re}(R)<1]. \tag{32}$$

The above square root may be taken with a plus on a minus. One sign is for the system with positive time and second—for negative time (see Note 1).

Let us consider boundary conditions at the junction. The continuity of axial velocity has been already used and it generated relations (31). Hence, boundary conditions may be reduced to form

$$R_2<r<R_1, \quad \nu_1(r)=0, \tag{33}$$

$$0<r<R_2, \quad f_1(r)=f_2(r). \tag{34}$$

Condition (33) can be realized with the collocation method—see (14). Condition (34) provides

$$1+R+\sum_m A_m J_0(\alpha_m r)=1-R-\sum_m A_m J_0(\alpha_m r)$$

$$\rightarrow \sum_m A_m J_0(\alpha_m r)=-R. \tag{35}$$

So, expressions (14) and (35) provide a system of linear equations to calculate unknown coefficients A_m with variable parameter R. Relation (32) reduces variations of R. There is only one more undefined detail—criterion K to choose the correct R value. It could be done, for example, in such a way,

$$K=\frac{1}{R_2^2}\int_0^{R_2}|f_1(r)-f_2(r)|r\,dr. \tag{36}$$

The less K the better choice for R.

IV. CONCLUDING REMARKS ABOUT PROPERTIES OF DISCONTINUITIES

The ''diameter step'' demonstrates nonzero reflection at $f\rightarrow 0$, its magnitude is pure real, and it is in perfect agreement with formula (1). The R value grows with an increase of frequency mostly due to the fast growing of an imaginary part. Let us consider case $R_2/R_1=0.5$. If we accept an error of 5% for the module of T, we come to the following restriction of the frequency range in (1),

$$0<f<0.58f_{r1}. \tag{37}$$

If we accept that the phase of the transmission coefficient should not exceed 5°, we come to another restriction (for the same case $R_2/R_1=0.5$):

$$0<f<0.182f_{r1}. \tag{38}$$

The diaphragm, in contrary to the diameter step, demonstrates zero reflection at $f\rightarrow 0$ and the fast growing of real and imaginary parts of R with increasing frequency. Actually, reflection from the diaphragm grows with frequency so fast that it exceeds reflection from the diameter step (!) with the same ratio R_1/R_2 at $f>0.3f_{r1}$—compare data in Tables II and V.

It is interesting to compare obtained data for the diaphragm with the Lamb formula (2). Such a comparison can-

TABLE V. Reflection coefficient for a rigid diaphragm with an opening in the circular waveguide. $R_1=1.0$ mm, $R_2=0.5$ mm. Wave velocity $c=340$ m/s (the air at room conditions). First radial resonance is at 207.4 kHz. R represents exact data for the diaphragm in a circular waveguide, R_L—according to the Lamb formula (1)–(2).

f, kHz	f/f_{r1}	R	R_L at $a=R_1$	R_L at $a=0.819R_1$
20	0.0964	$0.058+0.234i$	$0.049+0.215i$	$0.033+0.179i$
40	0.1929	$0.203+0.403i$	$0.170+0.375i$	$0.121+0.326i$
60	0.2893	$0.373+0.484i$	$0.315+0.464i$	$0.236+0.424i$
80	0.3857	$0.529+0.499i$	$0.450+0.497i$	$0.354+0.478i$
100	0.4822	$0.658+0.477i$	$0.561+0.496i$	$0.461+0.499i$
120	0.5786	$0.756+0.429i$	$0.648+0.478i$	$0.552+0.497i$
140	0.6750	$0.832+0.374i$	$0.715+0.452i$	$0.627+0.484i$
160	0.7715	$0.890+0.313i$	$0.766+0.424i$	$0.687+0.464i$
180	0.8679	$0.938+0.241i$	$0.805+0.396i$	$0.735+0.441i$

not be done absolutely correct because formula (2) was built in the Cartesian coordinate system, but we considered the tube structure with axial symmetry. However, there is an almost obvious approximate way to do it if we can compare structures with the same ratio of opening area to the total area of the cross section. The period of the grating, probably, should be equal to the diameter of the tube: $2a=2R_1$. The word ''probably'' is used because there may be supposed, as well, the condition ''equal lateral resonant frequencies'' in the circular tube and between two parallel rigid planes at distance $2a$. First resonance (with corresponding symmetry) between two planes occurs at

$$k=\pi/a. \tag{39}$$

The condition ''equal resonant frequencies'' provides the expression

$$\frac{\pi}{a}=\frac{3.832}{R_1}\rightarrow a=\frac{\pi}{3.832}R_1=0.819R_1. \tag{40}$$

So, Table V contains data for the reflection coefficient on frequency for a coaxial diaphragm in a circular waveguide and the results of the Lamb formula for $a=R_1$ and $a=0.819R_1$. Such a comparison provides an indirect confirmation of good accuracy of the Lamb formula almost down to f_{r1}.

There is one more interesting property of the diameter step that now may be checked directly—does it follow the reciprocity principle or it does not. Let us consider it briefly. There is a well-known and almost trivial reciprocity theorem for the Green's function of a simple source of velocity in an unbounded homogeneous medium,

$$G(r_1r_2)=G(r_2r_1). \tag{41}$$

The corresponding theorem is usually applied in acoustics in such a form: ''... a simple source at any point with coordinates r_1 produces the same pressure amplitude in another point with coordinates r_2 as it would be produced at r_1 by the same source, located at r_2.'' The cited formulation is ascribed to Helmholtz.[9] But the noted triviality of the reciprocity theorem is valid only for the case ''unbounded medium in a rest.'' The reciprocity in the presence of homogeneous flow may be obtained only with the additional condition of ''reversed flow.''[10] Probably, in all complex

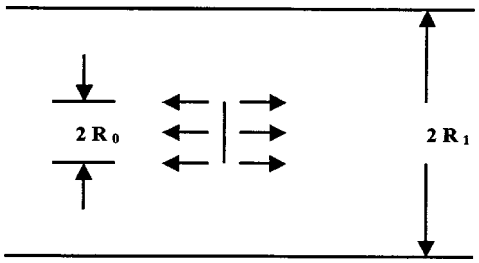


FIG. 3. A possible version of the monopole-type source of velocity with axial symmetry in a cylindrical waveguide. The source is located at $z=0$ and generates an even field of pressure over the z axis.

cases we should follow Ref. 10: ''if some characteristic of the field is invariant with respect to interchange locations source–receiver it is said that this characteristic follows to the reciprocity principle.''

It needs to clear the problem of a ''simple source'' in a cylindrical waveguide with a rigid wall to proceed further. This problem is absolutely clear in an infinite acoustic medium: it is a monopole-type point source of velocity (or very small in respect to one wavelength). A monopole-type source in a cylindrical waveguide may be done like it is shown in Fig. 3. Let us consider a double piston-type source of axial velocity V_0 with radius R_0 in the rigid tube with radius R_1 ($R_0<R_1$) at $z=0$. Such source forms not even the field of velocity over the z axis (and even the field of pressure). Hence, such a source may be considered as a double piston with not even displacements over the both radiating surfaces in a rigid wall: at $z=+0$ and at $z=-0$. Velocity at $z=+0$ is

$$\begin{aligned} &0<r<R_0, \quad V=V_0=\frac{1}{\pi R_0^2},\\ &R_0<r<R_1, \quad V=0. \end{aligned} \tag{42}$$

The sign of velocity at $z<0$ should be ''$-$.'' The velocity potential (Φ_s) of this source at $z>0$ and the condition that there exists only one traveling mode,

$$\begin{aligned} &\Phi_s=\Phi_0e^{-ikz}+\sum_m \Phi_m e^{-\beta_m z}J_0(\alpha_m r),\\ &k=\frac{\omega}{c}, \quad \alpha_m=\frac{\gamma_m}{R_1}, \quad \beta_m=\sqrt{\alpha_m^2-k^2}, \quad m=1,2,\ldots. \end{aligned} \tag{43}$$

TABLE VI. Parameter of nonreciprocity Δ on frequency for a diameter step with $R_1=1.0$ mm and $R_2=0.5$ mm. This parameter was calculated with data for T from Tables II and IV.

f, kHz	f/f_{r1}	Re (Δ)	Im (Δ)
20	0.0964	0.000 63	0.001 88
40	0.1929	−0.000 63	−0.000 63
60	0.2893	0.000 00	0.000 00
80	0.3857	0.000 00	0.000 00
100	0.4822	0.000 00	0.000 00
120	0.5786	−0.000 66	0.000 66
140	0.6750	0.000 68	0.000 68
160	0.7715	−0.000 71	0.000 71
180	0.8679	0.000 78	−0.000 78
190	0.9161	0.000 00	0.000 86

So, the source generates one traveling mode with amplitude Φ_0 and this amplitude does not depend on distance. Amplitude Φ_0 can be directly obtained with the Hankel transform over the interval $(0,R_1)$:

$$\Phi_0 = \frac{V_0 R_0^2}{ikR_1^2} = \frac{1}{ik\pi R_1^2}. \tag{44}$$

So, Φ_0 depends on the radius of the waveguide and does not depend on the source radius R_0 at a fixed ''source strength'' [expression (42) introduces the unit ''source strength''].

If we denote the transmission coefficient from the region with radius R_1 into the region with radius R_2 as T_{12} and inverse transmission as T_{21}, we can introduce a normalized parameter of nonreciprocity (ΔT) in the form

$$\Delta T = \frac{|T_{12}| - (R_1/R_2)^2 |T_{21}|}{|T_{12}|}. \tag{45}$$

It was figured out that ΔT is very small, probably, it does not exceed calculation errors—see Table VI. Hence, the discontinuity ''diameter step'' follows the reciprocity principle—at least below the first radial resonance and in some vicinity of ratio $R_2/R_1 = 0.5$.

[1] H. Lamb, *Hydrodynamics*, 6th ed. (Dover, New York, 1993), Article 306.
[2] N. Solokhin and M. Sumbatyan, ''Artificial layer,'' J. Res. Nondestructive Eval. **6**, 19–34 (1994).
[3] E. Skudrzyk, *Foundations of Acoustics Basic Mathematics and Basic Acoustics* (Springer-Verlag, Berlin, 1972).
[4] P. Morse and K. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1968).
[5] M. Sumbatyan, ''Low-frequency penetration of acoustic waves through periodic arbitrary-shaped grating: 3-D problem,'' Wave Motion **22**, 133–144 (1995).
[6] J. Miles, ''The analysis of plane discontinuity in cylindrical tubes. Part I and II,'' J. Acoust. Soc. Am. **17**, 259–284 (1946).
[7] F. Karal, ''The analogous acoustical impedance for discontinuities of circular cross section,'' J. Acoust. Soc. Am. **25**, 327–334 (1953).
[8] J. Kergomard and A. Garsia, ''Simple discontinuities at low frequencies: critical analysis and formulae,'' J. Sound Vib. **114**, 465–479 (1987).
[9] L. Kinsler and A. Frey, *Foundations of Acoustics* (Wiley, New York, 1950), p. 324.
[10] L. Brekhovskikh and O. Godin, *Acoustics of Layered Media II* (Springer-Verlag, Berlin, 1999), Chap. 4.2.

Unidimensional models of acoustic propagation in axisymmetric waveguides

Thomas Hélie[a)]
IRCAM, Centre G. Pompidou, CNRS UMR 9912, Équipe Analyse-Synthèse, 75004 Paris, France

(Received 8 November 2002; revised 25 June 2003; accepted 21 July 2003)

This paper presents a model of the linear acoustic propagation in axisymmetric waveguides, the pressure depending on a single space variable. The approach consists of writing the wave equation and the boundary conditions for a coordinate system rectifying the isobaric map at each time. The two-dimensional dependence of the problem is thus transferred from the pressure to the coefficients of the wave equation. From this result, an exclusively geometrical necessary condition is deduced for the admissibility of isobaric maps. However, the knowledge of the waveguide geometry is not sufficient to separate the pressure and the isobaric map solutions. In order to develop a unidimensional wave equation, a geometrical hypothesis is discussed. For lossless and motionless rigid waveguides, the deduced equation leads to exact results for tubes and cones. It may be interpreted as a Webster equation for a particular coordinate system so that the particular profiles for which analytical solutions of the pressure exist are redefined. The wave equation is also established for pipes with visco-thermal losses and, more generally, for mobile walls having a small admittance. The compatibility of the geometrical hypothesis with the exact model is specified for this general case. [DOI: 10.1121/1.1608962]

PACS numbers: 43.20.Mv, 43.20.Bi [MO] Pages: 2633–2647

I. INTRODUCTION

This work derives models accounting for acoustic propagation in axisymmetric waveguides and which depend on a single spatial variable. Such models are interesting because of their compactness of description which reveals the main effect involved in the variation of the cross section. For the same reason, they are also well adapted to the derivation of low-cost numerical simulations. The first establishment of such one-dimensional (1D) models is due to Bernoulli[1] and Lagrange,[2] even if the corresponding equation is commonly mentioned as the Webster equation.[3] This equation has been extensively investigated, as witnessed by the bibliography compiled by Eisner,[4] and the various geometrical hypotheses used for its derivation have been periodically discussed. Thus, planar wavefronts were contested by Lambert[5] and Weibel[6] who postulated spherical and curved ones, respectively. They reported the inadequacy of the first assumption, exhibiting the fact that wavefronts may be orthogonal to any curved rigid wall. The quasisphericity was experimentally confirmed for a horn profile in the low frequency range by Benade and Jansson.[7] Later, Putland[8] looked for necessary and sufficient conditions for a propagative acoustic mode to be spatially dependent on a single parameter. He pointed out that one-parameter acoustic fields obey a Webster equation and exhibited parallel planes, coaxial cylinders, and concentric spheres as the only possible corresponding potential surfaces. Nevertheless, even if finding a general 1D model is hopeless and leads to other approaches,[9–11] the simplicity of the Webster equation still stimulates the search for more accurate 1D models. Thus, Agulló, Barjau, and Keefe[12] recently assumed the time-invariance of equipotential surfaces and developed 1D models for both spherical and oblate ellipsoidal surfaces.

In this paper, a local geometrical hypothesis is proposed, namely, the quasi-sphericity of isobars near the wall. This hypothesis agrees with Benade and Jansson's experiment[7] but does not require the wavefronts to be fixed. The general method relies on a time-domain wave equation established for a coordinate system locally rectifying the isobaric map at each time. The rectification ensures that the pressure is reduced to a 1D function, but, concurrently, it transfers the two-dimensional (2D) dependence of the pressure to the coefficients of the isobaric wave equation. Establishing 1D models requires making an assumption. The quasisphericity hypothesis is chosen as a natural extension of the property satisfied by plane and spherical waves travelling in cylinders and cones, respectively. Moreover, it does not resort to integral equations or averaging operators as usual, and makes the rectification method able to treat the case of rigid immobile walls as well as that of moving walls with a small admittance.

The structure of the paper is as follows: Sec. II presents the problem. Section III establishes all the exact derivations. General geometrical definitions are presented. The wave equation is expressed in a coordinate system which locally rectifies the isobaric map at each time. An admissibility condition necessarily satisfied by any isobaric map is then deduced. The case of static isobaric maps is investigated. Section IV develops a 1D model for rigid motionless walls. The geometrical hypothesis is first detailed. The derived 1D model is proven to be equivalent to the Webster equation for which the longitudinal coordinate measures the arc length of the wall. The validity of both the hypothesis and the model is discussed. The particular profiles for which analytical solu-

[a)]Electronic mail: thomas.helie@ircam.fr; Université de Paris XI-Orsay, Paris, France.

TABLE I. List of symbols.

ρ_0	density of air
c	speed of sound in air
$\begin{cases} z=f(s,u,t) \\ r=g(s,u,t) \end{cases}$	time-varying change of spatial coordinates
$p(z,r,t)$, $\tilde{p}(s,u,t)$	acoustic pressure located at (z,r,t) and (s,u,t), respectively
$\mathbf{v}(z,r,t)$	particle acoustic velocity located at (z,r,t)
$\mathcal{I}_{s,t}$	isobars indexed by s at the time t
$\mathcal{T}_{u,t}$	axisymmetric surfaces indexed by u at the time t
$\mathcal{W}_t$	wall of the guide at the time t
$q\|_w(s,t)$	quantity $q(s,u,t)$ evaluated in $u=w$ *i.e.* on $\mathcal{T}_{w,t}$
$\mathbf{u_z}$, $\mathbf{u_r}$	unit vectors associated with the longitudinal (z) and the transverse (r) coordinates of the cylindrical basis
$\mathbf{u_s}(s,u,t)$, $\mathbf{u_u}(s,u,t)$	field of unit vectors tangent to $\mathcal{I}_{s,t}$ and $\mathcal{T}_{u,t}$, respectively
$\mathbf{w_s}(s,u,t)$, $\mathbf{w_u}(s,u,t)$	vectors $\mathbf{u_s}(s,u,t)$ and $\mathbf{u_u}(s,u,t)$ rotated from $+\pi/2$
$\sigma_s(s,u,t)$, $\sigma_u(s,u,t)$	norms of the vectors of the basis induced by the change (f,g) for the coordinates s and u, respectively
$\xi_s(s,u,t)$, $\xi_n(s,u,t)$	Mach numbers of a point located at (s,u) for the orthogonal directions $\mathbf{u_s}$ and $\mathbf{w_s}$
$\theta(s,u,t)$, $\phi(s,u,t)$, and $\delta(s,u,t)$	oriented angles $(\mathbf{u_z},\mathbf{u_s})$, $(\mathbf{u_r},\mathbf{u_u})$, and $(\mathbf{w_s},\mathbf{u_u})$, respectively
∂_z^k	partial derivative with respect to z of order k
grad	gradient operator
div	divergence operator
$\partial_{\mathbf{w_s}}=\mathbf{w_s}\cdot\mathbf{grad}$	partial derivative in the direction $\mathbf{w_s}$
$\Delta=\nabla^2$	Laplacian operator

tions of the pressure exist are then redefined. Finally, Sec. V extends the 1D model to the case of mobile walls having small admittances. The model is first derived, the compatibility of the geometrical hypothesis with this generalization is detailed, and the particular cases of mobile rigid walls and of visco-thermal boundary layers are investigated. (See Table I.)

II. PROBLEM STATEMENT

Throughout this paper, the problem considered is linear acoustic propagation in guides that are symmetrical with respect to the longitudinal axis (Oz). Only axisymmetric excitations are considered so that the whole problem is axisymmetric.

A. Basic equations

For adiabatic lossless media, the equation of mass conservation[13]

$$\rho_0 c^2 \,\mathrm{div}(\mathbf{v}) = -\partial_t p, \tag{1}$$

and the equation of momentum conservation

$$\rho_0 \partial_t \mathbf{v} = -\mathbf{grad}(p), \tag{2}$$

yield the wave equation

$$\left(\Delta - \frac{1}{c^2}\partial_t^2\right) p = 0, \tag{3}$$

where p denotes the acoustic pressure, $\mathbf{v}$ is the particle acoustic velocity, ρ_0 is the density of the air, and c is the speed of sound in air. The boundary conditions will be specified in the following, for each particular case studied.

Because of the symmetry of the problem, p and $\mathbf{v}$ only depend on the longitudinal and the transverse coordinates, noted z and r, respectively. For this 2D problem, the spatial operators are reduced to[13]

$$\mathrm{div}(\mathbf{v}) = \partial_z(\mathbf{v}\cdot\mathbf{u_z}) + \frac{\mathbf{v}\cdot\mathbf{u_r}}{r} + \partial_r(\mathbf{v}\cdot\mathbf{u_r}), \tag{4}$$

$$\mathbf{grad}(p) = \partial_z p\,\mathbf{u_z} + \partial_r p\,\mathbf{u_r}, \tag{5}$$

and

$$\Delta p = \partial_z^2 p + \frac{1}{r}\partial_r p + \partial_r^2 p, \tag{6}$$

where $\mathbf{u_z}$ and $\mathbf{u_r}$ are the normal unit vectors, respectively, associated with the coordinates z and r, and "$\cdot$" denotes the scalar product.

B. Prospective goals and description of the problem

Simulating acoustic propagation in an axisymmetric waveguide requires solving the 2D problem Eq. (3) with Eq. (6) for the boundary conditions on the wall, as well as at the input and the output of the guide. For instance, target applications may require wall conditions such as motionless rigid walls (possibly with visco-thermal boundary layers) for models of wind instruments, or controlled moving and vibrating walls for vocal tract models. The purpose of this paper is to propose a method to derive models of such waveguides, which do not require a resolution in the 2D-inside space, allowing low-cost simulations (e.g., real-time applications). The idea developed in the following consists of reducing the 2D problem to a 1D complexity such as the Webster equation does, making assumptions as weak as possible.

The first step of the method is to separate the geometrical information carried by the time-varying isobaric maps from that of the propagation of the pressure travelling on it. This is done considering a change of coordinates $(z,r) = (f(s,u,t),g(s,u,t))$ where s is chosen as an isobar index (Sec. III B), defining f and g in an implicit way [Eq. (15)].

Thus, z and r become (like the pressure) dependent variables of the partial differential equation which models the wave equation: they describe the isobaric map with respect to the independent variables (s,u,t). Writing that s indexes isobars [Eq. (15)] enables the derivation of the gradient of the pressure [Eq. (19)] and the wave equation [Eq. (20)] for

(s,u,t). This last equation exactly models the coupling of the propagation of the pressure with the isobaric map geometry, maps for which a necessary condition is straightforwardly deduced [Eq. (24)] and time-invariant specimens are exhibited (Sec. III D).

As the pressure does not depend on u, having explicit expressions of the coefficients of the wave equation for a given u suffices to furnish 1D models. Choosing to describe the wall for a constant $u=w$ leads to an explicit parametrization $f|_w, g|_w$ of the wall [Eq. (40)]. An additional hypothesis is needed to separate the propagation and the geometrical problems because partial derivatives of f and g with respect to u are involved.

The hypothesis of quasisphericity of isobars near the wall [Eq. (49)] makes the isolation of these problems possible. For lossless and motionless rigid walls, the deduced 1D model [Eq. (51)] can be written for a z description of the wall [Eq. (54)] or an ℓ-description [Eq. (55)] which leads to a Webster equation, ℓ measuring the arc length of the wall. For more general wall conditions, the 1D model [Eq. (69)] requires that isobars are nearly orthogonal to the wall (Sec. V B) to preserve the linearity of the propagation with respect to the pressure. This is fulfilled for mobile walls in many practical cases and for motionless rigid walls which induce visco-thermal losses, leading to the models in Eq. (78) and Eq. (82), respectively.

III. DERIVATION OF EXACT EQUATIONS

This section establishes the wave equation in any coordinate system rectifying the isobaric map for axisymmetric problems. The calculations are exact whenever isobaric maps can be locally described by a $\mathcal{C}^2$-regular diffeomorphism Φ; that is, Φ has an inverse Φ^{-1} and both have continuous derivatives up to the second order. Although topological aspects are not discussed here, a necessary geometrical condition is given for the map's admissibility. Before the method is described, the various geometrical quantities involved are precisely defined for an arbitrary change of coordinates. The definitions and the notations of this preliminary part constitute the reference list of the geometrical quantities used in this paper.

A. Geometrical definitions

Let O and $\mathcal{B}^0=(\mathbf{u_z},\mathbf{u_r})$ be the origin located on the axis of symmetry and the oriented canonical basis of the reference 2D frame (O,z,r), respectively. Let Φ be an arbitrary regular diffeomorphism defining a spatiotemporal change of coordinates $(s,u,t)=\Phi(z,r,t)$, which does not distort time. Then, the functions f and g defined by

$$z=f(s,u,t), \tag{7a}$$

$$r=g(s,u,t), \tag{7b}$$

exist and have the same regularity as Φ. Coordinates s and u are commonly called curvilinear because the variation of one of them does not make the corresponding point describe a straight line in the original space. The associated curves are noted, for each time t,

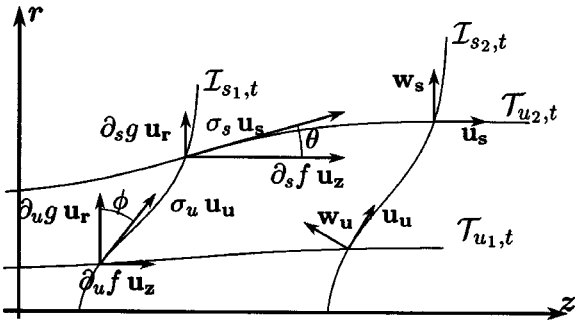


FIG. 1. Definition of the geometrical quantities related to the local basis $\mathcal{B}_t$. Note that although all represented vectors and angles coexist at every point located at (s,u), various quantities are described for several distinct points to improve the clarity and legibility of the figure.

$\mathcal{T}_{u,t}$ if s varies and u remains constant,
$\mathcal{I}_{s,t}$ if u varies and s remains constant.

If Φ has a $\mathcal{C}^1$ regularity, a spatial local basis $\mathcal{B}_t$ for the (s,u)-coordinate system may be defined for each time t by the vectors $\partial_s f\mathbf{u_z}+\partial_s g\mathbf{u_r}$ and $\partial_u f\mathbf{u_z}+\partial_u g\mathbf{u_r}$. Then, their respective norms σ_s and σ_u, their associated unit vectors $\mathbf{u_s}$ and $\mathbf{u_u}$, and the characteristic oriented angles $\theta=(\mathbf{u_z},\mathbf{u_s})$ and $\phi=(\mathbf{u_r},\mathbf{u_u})$ are defined. These geometrical quantities represented in Fig. 1 are given by the following expressions:

$$\sigma_s=\sqrt{(\partial_s f)^2+(\partial_s g)^2}, \tag{8a}$$

$$\sigma_u=\sqrt{(\partial_u f)^2+(\partial_u g)^2}, \tag{8b}$$

$$\cos\theta=\partial_s f/\sigma_s, \tag{9a}$$

$$\sin\theta=\partial_s g/\sigma_s, \tag{9b}$$

$$\cos\phi=\partial_u g/\sigma_u, \tag{9c}$$

$$\sin\phi=-\partial_u f/\sigma_u, \tag{9d}$$

$$\mathbf{u_s}=\cos\theta\,\mathbf{u_z}+\sin\theta\,\mathbf{u_r}, \tag{10a}$$

$$\mathbf{u_u}=-\sin\phi\,\mathbf{u_z}+\cos\phi\,\mathbf{u_r}. \tag{10b}$$

For convenience, the vectors $\mathbf{u_s}$ and $\mathbf{u_u}$ rotated counterclockwise through the angle $+\pi/2$ are also introduced, namely,

$$\mathbf{w_s}=-\sin\theta\,\mathbf{u_z}+\cos\theta\,\mathbf{u_r}, \tag{11a}$$

$$\mathbf{w_u}=-\cos\phi\,\mathbf{u_z}-\sin\phi\,\mathbf{u_r}. \tag{11b}$$

Then, the default of orthogonality of $\mathcal{B}_t$ can be exhibited by the angular deviation $\delta=(\mathbf{w_s},\mathbf{u_u})$ and more generally by its tangent ϵ given by

$$\delta=\phi-\theta, \tag{12a}$$

$$\epsilon=\tan\delta. \tag{12b}$$

Indeed, a local basis $\mathcal{B}_t$ is orthogonal if and only if $\epsilon=0$. This is also equivalent to $\mathbf{u_s}\cdot\mathbf{u_u}=0$ which yields the useful relation

$$\partial_s f\partial_u f+\partial_s g\partial_u g=0. \tag{13}$$

Finally, the Mach numbers $\xi_s=(1/c)\mathbf{V}\cdot\mathbf{u_s}$ and $\xi_n=(1/c)\mathbf{V}\cdot\mathbf{w_s}$ associated with the tangential and the normal components of the velocity $\mathbf{V}$ of a geometrical point (not that

of the particle) located at (s,u), respectively, are introduced. These dynamic quantities are given by

$$\xi_s=\frac{1}{c}[\partial_t f\cos\theta+\partial_t g\sin\theta], \tag{14a}$$

$$\xi_n=\frac{1}{c}[-\partial_t f\sin\theta+\partial_t g\cos\theta]. \tag{14b}$$

Note that a coordinate system compatible with the axial symmetry is such that for all u, t, there exists a u^* such that, for all s,

$$f(s,u^*,t)=f(s,u,t),$$

$$g(s,u^*,t)=-g(s,u,t).$$

This implies in particular that for each time, there exists u_0 such that $\mathcal{I}_{u_0,t}$ is the axis of symmetry. Generally, given an axisymmetric map $\{\mathcal{I}_{s,t},\mathcal{T}_{u,t}\}$, a natural choice for its description consists of using f and g which have an even and odd u-parity, respectively.

B. Rectification of the isobaric map

A change of coordinates Φ which rectifies the isobaric map is such that one ordinate indexes isobars while the other is not colinear to the first one to preserve Φ as a bijection. Conferring the indexation of isobars to s signifies that, for each time t, the curves $\mathcal{I}_{s,t}$ represent the isobars while the curves $\mathcal{T}_{u,t}$ are nowhere tangent to $\mathcal{I}_{s,t}$ (see Fig. 2).

Thus, for any given s, the pressure does not depend on u, so the local rectification on a domain Ω is obtained for the following condition: there exists a function $\tilde{p}$, such that for all (s,u,t) in Ω,

$$p(f(s,u,t),g(s,u,t),t)=\tilde{p}(s,t). \tag{15}$$

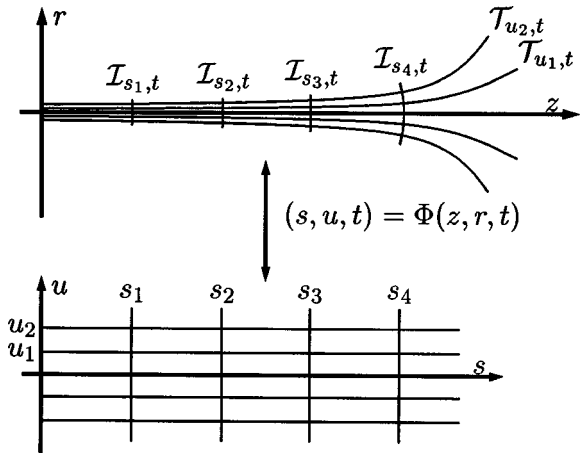


FIG. 2. This figure gives an example of a change of coordinates which rectifies the isobaric maps at a given time t, and for which u indexes the field lines. For the coordinate system (s,u), the pressure does not vary with u. In this illustration, the simple topology of isobars ($\mathcal{I}_{s,t}$) makes the definition of the diffeomorphism Φ possible over the whole map. When more complex topologies appear (closed curves, singular points, splitting curve, etc.), several changes of coordinates should be considered on local domains. The study of a global resolution for such topologies is not discussed here.

Assuming the $\mathcal{C}^2$ regularity for f, g, and $\tilde{p}$, this implicit relation on f and g makes the derivation of the wave equation which governs $\tilde{p}$ possible on Ω. The method consists of expressing the partial derivatives of $p(z,r,t)$ evaluated in $(z,r,t)=(f(s,u,t),g(s,u,t),t)$ from the partial derivatives of $\tilde{p}(s,t)$, as described below.

Applying the differential operators ∂_s, ∂_u, and ∂_t on Eq. (15) up to the second order leads to the system

$$\tilde{\mathcal{D}}(s,u,t)=\mathcal{M}(s,u,t)\mathcal{D}(f(s,u,t),g(s,u,t),t). \tag{16}$$

The corresponding explicit formulation is developed in Eq. (17), omitting the variables of evaluation for sake of compactness,

$$\underbrace{\begin{bmatrix}\partial_s\tilde{p}\\ \partial_u\tilde{p}\\ \partial_t\tilde{p}\\ \partial_s^2\tilde{p}\\ \partial_u^2\tilde{p}\\ \partial_t^2\tilde{p}\\ \partial_s\partial_u\tilde{p}\\ \partial_s\partial_t\tilde{p}\\ \partial_u\partial_t\tilde{p}\end{bmatrix}}_{\tilde{\mathcal{D}}}=\underbrace{\begin{bmatrix}\partial_s f & \partial_s g & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ \partial_u f & \partial_u g & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ \partial_t f & \partial_t g & 1 & 0 & 0 & 0 & 0 & 0 & 0\\ \partial_s^2 f & \partial_s^2 g & 0 & (\partial_s f)^2 & (\partial_s g)^2 & 0 & 2\partial_s f\partial_s g & 0 & 0\\ \partial_u^2 f & \partial_u^2 g & 0 & (\partial_u f)^2 & (\partial_u g)^2 & 0 & 2\partial_u f\partial_u g & 0 & 0\\ \partial_t^2 f & \partial_t^2 g & 0 & (\partial_t f)^2 & (\partial_t g)^2 & 1 & 2\partial_t f\partial_t g & 2\partial_t f & 2\partial_t g\\ \partial_s\partial_u f & \partial_s\partial_u g & 0 & \partial_s f\partial_u f & \partial_s g\partial_u g & 0 & \partial_s f\partial_u g+\partial_u f\partial_s g & 0 & 0\\ \partial_s\partial_t f & \partial_s\partial_t g & 0 & \partial_s f\partial_t f & \partial_s g\partial_t g & 0 & \partial_s f\partial_t g+\partial_t f\partial_s g & \partial_s f & \partial_s g\\ \partial_u\partial_t f & \partial_u\partial_t g & 0 & \partial_u f\partial_t f & \partial_u g\partial_t g & 0 & \partial_u f\partial_t g+\partial_t f\partial_u g & \partial_u f & \partial_u g\end{bmatrix}}_{\mathcal{M}}\underbrace{\begin{bmatrix}\partial_z p\\ \partial_r p\\ \partial_t p\\ \partial_z^2 p\\ \partial_r^2 p\\ \partial_t^2 p\\ \partial_z\partial_r p\\ \partial_z\partial_t p\\ \partial_r\partial_t p\end{bmatrix}}_{\mathcal{D}}. \tag{17}$$

Then, $\mathcal{D}$ is formally obtained from Eq. (16) by computing $\mathcal{M}^{-1}\tilde{\mathcal{D}}$ and remarking that

$$\tilde{\mathcal{D}}^T=[\partial_s\tilde{p},0,\partial_t\tilde{p},\partial_s^2\tilde{p},0,\partial_t^2\tilde{p},0,\partial_s\partial_t\tilde{p},0], \tag{18}$$

since $\partial_u\tilde{p}(s,t)=0$. Expressions expressed explicitly with the monospace dependent pressure $\tilde{p}(s,t)$ are straightforwardly deduced for the gradient from Eq. (5) and for the wave equation from Eq. (3) and Eq. (6) evaluated in $(z,r,t)=(f(s,u,t),g(s,u,t),t)$. After simplification and using the

notations defined in Sec. III A, these relations become

$$\mathbf{grad}(\tilde{p}) = \frac{\partial_s \tilde{p}}{\sigma_s}[\mathbf{u_s} + \boldsymbol{\epsilon}\mathbf{w_s}] \tag{19}$$

and

$$A_{s,s}\partial_s^2\tilde{p} + A_s\partial_s\tilde{p} + A_{s,t}\partial_s\partial_t\tilde{p} - \frac{1}{c^2}\partial_t^2\tilde{p} = 0, \tag{20}$$

where the coefficients given by

$$A_{s,s} = \frac{1+\epsilon^2-(\xi_s+\epsilon\xi_n)^2}{\sigma_s^2}, \tag{21a}$$

$$A_s = \frac{\sin\theta+\epsilon\cos\theta}{g\sigma_s} + \frac{1-\xi_n^2}{\sigma_s^2}\left[\partial_s(\epsilon^2) + (1+\epsilon^2)\left(\frac{\sigma_s\partial_u\phi}{\sigma_u\cos\delta} + \epsilon\partial_s\theta - \partial_s\ln\sigma_s\right)\right] - \frac{1}{2}\partial_s\left(\frac{\xi_s^2+2\epsilon\xi_s\xi_n-\xi_n^2}{\sigma_s^2}\right) - \frac{\xi_s\xi_n}{\sigma_s^2}\partial_s\epsilon + \frac{1}{c}\partial_t\left(\frac{\xi_s+\epsilon\xi_n}{\sigma_s}\right) + \frac{\xi_n}{c\sigma_s}\partial_t\epsilon, \tag{21b}$$

$$A_{s,t} = 2\frac{\xi_s+\epsilon\xi_n}{c\sigma_s}, \tag{21c}$$

are expressions of f and g exclusively and depend on (s,u,t). Although the tedious calculations are not detailed, note that this process cannot be reduced to writing the Laplacian operator for the change of coordinates Φ, as for usual static coordinates. Thus, the Mach numbers ξ_s and ξ_n involved in Eqs. (21) are precisely the contribution of the term $\partial_t^2 p$ in Eq. (3). They are the expression of the dynamic of the isobaric map.

Note that choosing u such that $\boldsymbol{\mathcal{T}}_{u,t}$ represent the field lines leads without loss of generality to simpler expressions for which $\epsilon=0$ because of the orthogonality of $\boldsymbol{\mathcal{I}}_{s,t}$ and $\boldsymbol{\mathcal{T}}_{u,t}$ at each point. This concise formulation is exploited in Sec. III D to exhibit static maps, and in Sec. IV to derive a monospace dependent wave equation for motionless rigid walls. However, the study dealing with mobile walls having small admittances presented in Sec. V requires being driven with nonorthogonal coordinates so that Eq. (20) is established in this more general context.

A remarkable property of these equations is that they are not modified by any bijective regular change of variables such that $s=\alpha(\hat{s},\hat{t})$, $u=\beta(\hat{s},\hat{u},\hat{t})$, and $t=\hat{t}$: taking $\hat{s}$, $\hat{u}$, $\hat{t}$, $\hat{f}(\hat{s},\hat{u},\hat{t})=f(\alpha(\hat{s},\hat{t}),\ \beta(\hat{s},\hat{u},\hat{t}),\ \hat{t})$, and $\hat{g}(\hat{s},\hat{u},\hat{t})=g(\alpha(\hat{s},\hat{t}),\beta(\hat{s},\hat{u},\hat{t}),\hat{t})$, and $\hat{p}(\hat{s},\hat{t})=\tilde{p}(\alpha(\hat{s},\hat{t}),\hat{t})$, in place of s, u, t, f, g, and $\tilde{p}$ keeps the formula identical. Moreover, the axial symmetry is also naturally supported by the isobaric wave equation. Indeed, for f and g having an even and odd u-parity, respectively, applying $u\mapsto -u$ in Eq. (20) leaves this equation invariant.

C. Regular isobaric maps and admissibility

A necessary condition for isobaric maps, which is independent of the pressure, can be deduced from Eq. (20) on (f,g). This criterion of admissibility furnishes a property inherent in the set of regular isobaric maps. It proves in particular that any arbitrary regular map does not always match with a wave propagation phenomenon.

The derivation of this condition relies on the u-independence of $\tilde{p}$. The method consists of applying operators ∂_u^k in Eq. (20). As the factors represented by $\boldsymbol{\mathcal{P}}=[\partial_s^2\tilde{p}\ \partial_s\partial_s\tilde{p}\ \partial_s\partial_t\tilde{p}\ \partial_t^2\tilde{p}]^T$, do not depend on u, they may be eliminated by making linear combinations of four distinct relations proceeding from four distinct integers k. In fact, only three of these relations suffice as the proof below describes.

If the $\mathcal{C}^2$ functions f and g have in addition a $\mathcal{C}^5$-regularity for the variable u, the operators ∂_u^k $(k=0,1,2,3)$ applied on Eq. (20) yield the system

$$\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}} = \mathbf{0}, \tag{22}$$

where

$$\boldsymbol{\mathcal{A}} = \begin{bmatrix} A_{s,s} & A_s & A_{s,t} & -\frac{1}{c^2} \\ \partial_u A_{s,s} & \partial_u A_s & \partial_u A_{s,t} & 0 \\ \partial_u^2 A_{s,s} & \partial_u^2 A_s & \partial_u^2 A_{s,t} & 0 \\ \partial_u^3 A_{s,s} & \partial_u^3 A_s & \partial_u^3 A_{s,t} & 0 \end{bmatrix}. \tag{23}$$

For a propagative phenomenon $\boldsymbol{\mathcal{P}}\neq\mathbf{0}$. $\boldsymbol{\mathcal{P}}$ is then an eigenvector of $\boldsymbol{\mathcal{A}}$ associated with the eigenvalue 0. This proves that $\det(\boldsymbol{\mathcal{A}})=0$. By developing the determinant with respect to the last column, a simpler equivalent formulation is obtained for

$$\det\begin{bmatrix} \partial_u A_{s,s} & \partial_u A_s & \partial_u A_{s,t} \\ \partial_u^2 A_{s,s} & \partial_u^2 A_s & \partial_u^2 A_{s,t} \\ \partial_u^3 A_{s,s} & \partial_u^3 A_s & \partial_u^3 A_{s,t} \end{bmatrix} = 0. \tag{24}$$

This necessary condition only involves the equations derived for $k=1,2,3$. Equation (24) closes the proof since $A_{s,s}$, A_s, and $A_{s,t}$ depend on f and g only, and thus, on the dynamic geometry.

The criterion Eq. (24) can be straightforwardly computed for any given maps to check their admissibility. Except for a few trivial dynamic and geometrical starting conditions, an analytic derivation of admissible maps from this criterion becomes very tricky.

D. Static isobaric maps

This section presents an investigation into the restricted case of static maps on which propagative waves may travel. In order to exhibit the physical propagative solutions only, the pressure is assumed to be such that the functions $\partial_s^2\tilde{p}$, $\partial_s\tilde{p}$, and $\partial_t^2\tilde{p}$ are nonzero and real. The investigation is run directly from the isobaric wave equation rather than the admissibility criterion which only gives a necessary condition.

A static map may be described by time-independent functions $f(s,u)$ and $g(s,u)$. Choosing $\boldsymbol{\mathcal{T}}_u$ as the field lines yields the constraint $\epsilon=0$. Thus, the solution map is represented without loss of generality by (f,g) satisfying Eq. (13) and

$$\partial_s^2\tilde{p}+\partial_s\mathcal{L}\partial_s\tilde{p}-\frac{\sigma_s^2}{c^2}\partial_t^2\tilde{p}=0, \tag{25}$$

calculated from Eq. (20) multiplied by σ_s^2. $\mathcal{L}$ is given by

$$\mathcal{L}=\ln\left|g\frac{\sigma_u}{\sigma_s}\right|, \tag{26}$$

remarking that Eq. (13) implies $\partial_u\phi/\cos^3\delta=\partial_u\theta=\partial_s\sigma_u/\sigma_s$. Then, two cases are examined separately: σ_s depends on s only, or on (s,u).

1. σ_s does not depend on u

In this case, $\partial_s\mathcal{L}$ does not depend on u either [see Eq. (25)]. Parallel planes orthogonal to (Oz) and coaxial cylinders are obtained for $\partial_s g=0$ and $\partial_u g=0$, respectively. Except for these pathological cases, Eq. (13) ensures that $\partial_s f$, $\partial_u f$, $\partial_s g$, and $\partial_u g$ are nonzero.

Then, Eq. (13) makes the elimination of $\partial_s f$ possible in $\partial_u(\sigma_s^2)=0$ calculated from the definition Eq. (8a). The result $\partial_s\ln|\partial_u g|=\partial_s\ln|\partial_u f|$ equivalent to $\partial_s\ln|\tan\phi|=0$ [see Eq. (9c) and Eq. (9d)] proves that ϕ depends only on u. As $\epsilon=\tan(\phi-\theta)$ is zero, θ is an exclusive function of u as well. Equations (9a) and (9b) show that

$$f(s,u)=R(s)\cos\theta(u)+F(u), \tag{27a}$$

$$g(s,u)=R(s)\sin\theta(u)+G(u), \tag{27b}$$

where $R'=\sigma_s$, and F and G are arbitrary.

Now, from Eq. (9c), $\sigma_u=\partial_u g/\cos\phi$ so that $\partial_s\partial_u\mathcal{L}=0$ can be written $\partial_u\partial_s\ln|g\partial_u g|=0$ which, with Eq. (27b), leads to

$$G(u)=R_0\sin\theta(u),$$

R_0 being an arbitrary constant. Finally, the calculation of Eq. (13) using Eqs. (27) with this function G yields

$$F(u)=R_0\cos\theta(u)+F_0,$$

where F_0 is arbitrary. The corresponding isobars are concentric spheres. This conclusion completes the proof that parallel planes, coaxial cylinders, and concentric spheres are the only static isobaric maps for which σ_s does not depend on u.

2. σ_s depends on u

In this case, $\partial_s\mathcal{L}$ also depends on u. Applying ∂_u to Eq. (25) leads to

$$\partial_u\partial_s\mathcal{L}\partial_s\tilde{p}-\frac{\partial_u(\sigma_s^2)}{c^2}\partial_t^2\tilde{p}=0.$$

Since $\partial_u\sigma_s$ and $\partial_s\tilde{p}$ are assumed nonzero, this relation proves that $\partial_t^2\tilde{p}/\partial_s\tilde{p}$ does not depend on t. Then, Eq. (25) proves that $\partial_s^2\tilde{p}/\partial_s\tilde{p}=\partial_s\ln|\tilde{p}|$ does not depend on t either, so that a general solution is $\tilde{p}(s,t)=A(s)b(t)+C(t)$ where A, b, and C are real functions. Writing Eq. (25) for this solution and applying the operator $X\mapsto\partial_t(X/b(t))$ on it show that both $c^{-2}b''(t)/b(t)=-k_0^2$ and $C''(t)/b(t)=C_0$ are real constants so that $C(t)=C_0b(t)+C_1t+C_2$ with $(C_1,C_2)\in\mathbb{R}^2$. Defining $a(s)=A(s)+C_0$, this proves that the general solution is

$$\tilde{p}(s,t)=a(s)b(t)+C_1t+C_2,$$

with

$$b(t)=B_1e^{ik_0ct}+B_2e^{-ik_0ct},$$

where B_1, B_2 are complex arbitrary constants such that $b(t)$ is real. Although it does not affect the following, note that physical cases require that $C_1=C_2=0$ since $\tilde{p}$ is the acoustic pressure.

Using Eq. (26), Eq. (25) is then reduced to

$$\partial_s\ln|a|\,\partial_s\ln\left|\frac{a'}{\sigma_s}g\sigma_u\right|+(k_0\sigma_s)^2=0.$$

Moreover, if the spatial dependence a for the wave number k_0 is locally bijective, posing $\hat{s}=a(s)$, and then defining $\hat{f}(\hat{s},u,t)=f(s,u,t)$ and $\hat{g}(\hat{s},u,t)=g(s,u,t)$ yields

$$\partial_{\hat{s}}\ln\left|\frac{\hat{g}\hat{\sigma}_{\hat{s}}}{\hat{\sigma}_u}\right|+(k_0\sigma_s)^2\hat{s}=0.$$

This equation shows that every static map such that σ_s depends on u is associated with a time dependency corresponding to a pure frequency (k_0 real) or a pure exponential (k_0 imaginary). More precisely, once given the geometry of the wall parameterized with s, u and expressing the boundary conditions, this last equation and Eq. (13) give the isobaric map associated with the wave number k_0.

3. Conclusion: Theorem

(i) Parallel planes, coaxial cylinders, and concentric spheres are the only static axisymmetric isobaric geometries on which non-necessarily sinusoidal or exponential waves can propagate.

(ii) Other static geometries are associated with a wave number $k_0\in\mathbb{R}^+\cup i\mathbb{R}^+$ which characterizes the geometric invariant given by

$$\frac{\partial_{\hat{s}}\ln\left|\frac{\hat{g}\hat{\sigma}_{\hat{s}}}{\hat{\sigma}_u}\right|}{\hat{s}\hat{\sigma}_{\hat{s}}^2}=-k_0^2, \tag{28}$$

where $\hat{s}$ denotes the level of the spatial dependence of the pressure and u is orthogonal to $\hat{s}$.

The first part of this theorem (i) corroborates the result of Putland[8] who exhibited parallel planes, coaxial cylinders, and concentric spheres as the only maps making the separation of variables possible. Nevertheless, other maps exist even if they are associated with particular waveforms (sinusoidal or exponential). This last case would correspond to having the Wronskian defined by Putland being zero, a case that he discarded [see Eqs. (19) and (20) in his paper].

4. Example of a static map

As only a few of them can be performed analytically, knowing that all the modal static maps satisfy Eq. (28) can be very useful. However, it is interesting to illustrate this second part of the theorem through an example for which the spatial dependence takes an analytical expression.

The example, detailed below, concerns the field of pressure inside a paraboloid of revolution. Using the separation principle of the Helmholtz equation for the orthogonal parabolic coordinates[14] $(\mu,\nu)\in\mathbb{R}^+\times\mathbb{R}$ defined by

$$z=(1/2)(\mu^2-\nu^2), \tag{29a}$$

$$r=\mu\nu, \tag{29b}$$

a particular solution of the wave equation is proven to be

$$P_0 J_0\left(\frac{k_0}{2}\mu^2\right) J_0\left(\frac{k_0}{2}\nu^2\right)\cos(ck_0(t-t_0)), \tag{30}$$

where P_0 and t_0 are arbitrary real constants, and J_0 is the 0th order Bessel function of the first kind. This solution can be checked straightforwardly from Eq. (3) and Eq. (6) using the reciprocal change of coordinates $(1/2)\mu=\sqrt{z+\sqrt{z^2+r^2}}$ and $\nu=r/\mu$, and recalling that

$$J_0''(X)=\frac{J_0'(X)}{X}-J_0(X). \tag{31}$$

Now, for the wave number k_0, the isobaric map is such that

$$\hat{s}=J_0\left(\frac{k_0}{2}\mu(\hat{s},u)^2\right) J_0\left(\frac{k_0}{2}\nu(\hat{s},u)^2\right), \tag{32}$$

where $\hat{s}$ defines the level of the spatial dependence of the pressure. It is represented in Fig. 3.

As expected, the second part (ii) of the theorem is satisfied. However, because the definition of the $\hat{s}$-level curves is implicit [Eq. (32)], the proof is not straightforward as detailed below.

From $\hat{f}(\hat{s},u)=(1/2)(\mu(\hat{s},u)^2-\nu(\hat{s},u)^2)$ and $\hat{g}(\hat{s},u)=\mu(\hat{s},u)\nu(\hat{s},u)$, it follows that $\hat{\sigma}_{\hat{s}}^2=(\mu^2+\nu^2)[(\partial_{\hat{s}}\mu)^2+(\partial_{\hat{s}}\nu)^2]$, $\hat{\sigma}_u^2=(\mu^2+\nu^2)[(\partial_u\mu)^2+(\partial_u\nu)^2]$, and that the orthogonality Eq. (13) of $(\hat{s},u)$ is also given by

$$\partial_{\hat{s}}\mu\partial_u\mu+\partial_{\hat{s}}\nu\partial_u\nu=0. \tag{33}$$

These equations yield $\partial_{\hat{s}}\hat{\mathcal{L}}=\partial_{\hat{s}}\ln|\mu\nu\partial_u\nu/\partial_{\hat{s}}\mu|$.

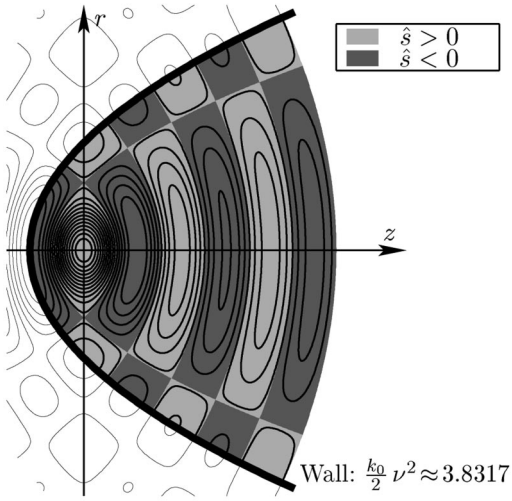


FIG. 3. The $\hat{s}$-level curves given by Eq. (32) define the static isobaric map for an oscillating mode of wave number k_0. This map can be defined over the whole space, but it also corresponds to that inside paraboloids described by a constant ν-coordinate, such that $J_0'[(k_0/2)\nu^2]=0$. Indeed, the boundary condition of rigid walls is fulfilled on such paraboloids (the gradient of the pressure has no component orthogonal to the wall). This figure presents the wall associated with the first zero of J_0', namely, $(k_0/2)\nu^2\approx 3.8317$.

Applying $\partial_{\hat{s}}$ and ∂_u to Eq. (32) leads to

$$\partial_{\hat{s}}\mu=\frac{1-k_0\nu J_0'\left(\frac{k_0}{2}\nu^2\right)J_0\left(\frac{k_0}{2}\mu^2\right)\partial_{\hat{s}}\nu}{k_0\mu J_0'\left(\frac{k_0}{2}\mu^2\right)J_0\left(\frac{k_0}{2}\nu^2\right)}, \tag{34a}$$

$$\partial_u\mu=-\frac{\nu J_0'\left(\frac{k_0}{2}\nu^2\right)J_0\left(\frac{k_0}{2}\mu^2\right)}{\mu J_0'\left(\frac{k_0}{2}\mu^2\right)J_0\left(\frac{k_0}{2}\nu^2\right)}\partial_u\nu. \tag{34b}$$

Substituting $\partial_{\hat{s}}\mu$ and $\partial_u\mu$ in Eq. (33) yields (for $\partial_u\nu\neq 0$)

$$\partial_{\hat{s}}\nu=\frac{\nu J_0'\left(\frac{k_0}{2}\nu^2\right)J_0\left(\frac{k_0}{2}\mu^2\right)}{k_0\left[\left(\mu J_0'\left(\frac{k_0}{2}\mu^2\right)J_0\left(\frac{k_0}{2}\nu^2\right)\right)^2+\left(\nu J_0'\left(\frac{k_0}{2}\nu^2\right)J_0\left(\frac{k_0}{2}\mu^2\right)\right)^2\right]}, \tag{35}$$

and then from Eq. (34a)

$$\partial_{\hat{s}}\mu=\frac{\mu J_0'\left(\frac{k_0}{2}\mu^2\right)J_0\left(\frac{k_0}{2}\nu^2\right)}{k_0\left[\left(\mu J_0'\left(\frac{k_0}{2}\mu^2\right)J_0\left(\frac{k_0}{2}\nu^2\right)\right)^2+\left(\nu J_0'\left(\frac{k_0}{2}\nu^2\right)J_0\left(\frac{k_0}{2}\mu^2\right)\right)^2\right]}. \tag{36}$$

These last equations give expressions of $\partial_{\hat{s}}\mu$ and $\partial_{\hat{s}}\nu$ which do not involve derivatives of μ and ν. They specifically make the computation of $\partial_{\hat{s}}\hat{\mathcal{L}}$ with this same property possible, overcoming the difficulty due to the implicit definition for the proof. The expressions of $\partial_{\hat{s}}\ln|\mu|$ and $\partial_{\hat{s}}\ln|\nu|$ are straightforward. That of $\partial_{\hat{s}}\ln|\partial_u\nu|$ is computed applying ∂_u to Eq. (35), substituting occurrences of $\partial_u\mu$ by $\partial_u\nu$ according to Eq. (34b), and dividing by the common factor $\partial_u\nu$. That of $\partial_{\hat{s}}\ln|\partial_{\hat{s}}\mu|$ is computed applying $\partial_{\hat{s}}$ to Eq. (36), dividing by $\partial_{\hat{s}}\mu$, and substituting occurrences of $\partial_{\hat{s}}\mu$ and $\partial_{\hat{s}}\nu$ according to Eq. (36) and Eq. (35). The two first terms involve μ, ν, J_0, and J_0'. The two last also involve J_0''.

On the one hand, the sum of these terms yields

$$\partial_{\hat{s}}\hat{\mathcal{L}} = -\frac{(\mu^2+\nu^2)J_0\left(\frac{k_0}{2}\mu^2\right)J_0\left(\frac{k_0}{2}\nu^2\right)}{\left(\mu J_0'\left(\frac{k_0}{2}\mu^2\right)J_0\left(\frac{k_0}{2}\nu^2\right)\right)^2+\left(\nu J_0'\left(\frac{k_0}{2}\nu^2\right)J_0\left(\frac{k_0}{2}\mu^2\right)\right)^2}, \tag{37}$$

after the substitution of J_0'' according to Eq. (31). On the other hand, computing $\hat{\sigma}_{\hat{s}}^2$ from Eq. (36) and Eq. (35), and using Eq. (32) to compute $\hat{s}$, the computation of $-k_0^2\hat{s}\hat{\sigma}_{\hat{s}}^2$ is straightforward. It exactly yields the right-hand side of Eq. (37), proving Eq. (28).

IV. MONO-SPACE-DEPENDENT WAVE EQUATION FOR LOSSLESS AND MOTIONLESS RIGID WALLS

A rigorous derivation of a mono-space model of a waveguide proves to be unworkable for arbitrary geometries. Even if the exact equation (20) governs a mono-space-dependent pressure, expressing $A_{s,s}$, A_s, and $A_{s,t}$ from the shape of the waveguide does not succeed in such a 1D model, as described below.

A. Problem posed by the derivation of a 1D model

Let Φ now be a particular local change of coordinates such that the wall noted $\mathcal{W}_t$ is simply described for a fixed u noted w so that

$$\mathcal{T}_{w,t} = \mathcal{W}_t. \tag{38}$$

This can be achieved if $\mathcal{W}_t$ is not tangent to isobars, which is discussed below. Adopting the notation

$$q|_w(s,t) = q(s,w,t), \tag{39}$$

to indicate that a quantity $q(s,u,t)$ is evaluated on the wall $\mathcal{W}_t$, the profile of $\mathcal{W}_t$ is parametrized by

$$z = f|_w(s,t), \tag{40a}$$

$$r = g|_w(s,t). \tag{40b}$$

Then, deriving a 1D model could consist of evaluating the coefficients $A_{s,s}$, A_s, and $A_{s,t}$ on $\mathcal{W}_t$, namely imposing $u=w$, and then, expressing them from the known parametrization $(f|_w, g|_w)$. Unfortunately, this cannot be achieved in general. Nevertheless, executing this process reveals the difficulty and, simultaneously, the reason why decoupling the waveform resolution from that of the geometry of isobars is impossible.

This section exhibits the problem for walls assumed ideally rigid, lossless, and motionless: $\mathcal{T}_{w,t} = \mathcal{W}_t = \mathcal{W}$ so that $f|_w$ and $g|_w$ can be chosen time independent and $\xi_s|_w = \xi_n|_w = 0$ [see Eq. (14)]. In this case, the corresponding boundary condition is that the particle velocity $\mathbf{v}$, and so from Eq. (2), the gradient of the pressure $\mathbf{grad}(p)$, have no component normal to the wall. For locally nondegenerate cases [$\mathbf{grad}(p) \neq 0$], isobars are necessarily orthogonal to the wall $\mathcal{W}$. Considering a domain Ω on which this condition is satisfied, this proves that $\mathcal{W}$ belongs to the field lines so that $\epsilon|_w = 0$ [see Eq. (12b)]. Note that if the gradient is zero on a set of isolated points, local solutions of the same 1D model can be concatenated to form a maximal solution under the $\mathcal{C}^2$ regularity assumption. If this set has a nonzero measure, the pressure computed from a mono-space partial differential equation of finite order is necessarily locally constant for both the time and the space variables. On this set, the partial derivatives of the pressure are zero and will make the 1D linear model locally trivial ($0=0$), independently of the involved coefficients. Practically, these properties make it possible to proceed considering only the case $\epsilon|_w = 0$, without loss of generality.

Now, $\epsilon|_w = 0$ and $\xi_s|_w = \xi_n|_w = 0$ yields, for $u=w$, the simplified coefficients $A_{s,s} = 1/\sigma_s^2$, $A_{s,t} = 0$, and using Eq. (9b), $A_s = \partial_s \ln(g/\sigma_s)/\sigma_s^2 + \partial_u\phi/(\sigma_s\sigma_u\cos\delta)$.

Since $f|_w$ and $g|_w$ can be differentiated with respect to s, then σ_s and $\partial_s \ln(g/\sigma_s)$ can be evaluated for $u=w$ using Eq. (8a). On the other hand, $f|_w$ and $g|_w$ do not give any information about $\partial_u f|_w$ and $\partial_u g|_w$, and the local orthogonality ($\epsilon|_w = 0$) only yields $|\cos\delta|_w = 1$. Thus, starting with $f|_w$, $g|_w$, and the relation $\epsilon|_w = 0$, one cannot evaluate $\partial_u\phi/(\sigma_u\cos\delta)$.

To cope with this difficulty, a local geometrical hypothesis is presented, which gives a natural extension of the exact models of propagation in tubes and cones, and agrees with the isobars mobility.

B. Geometrical hypothesis

Let $M(s)$ be a point of $\mathcal{W}$ indexed by s. Let $\mathcal{S}_s$ be the sphere tangent to $\mathcal{T}_{s,t}$ at $M(s)$ and centered at $O_s \in (Oz)$ (see Fig. 4). Then, the z-ordinate $z_{O_s}(s)$ of O_s, the radius

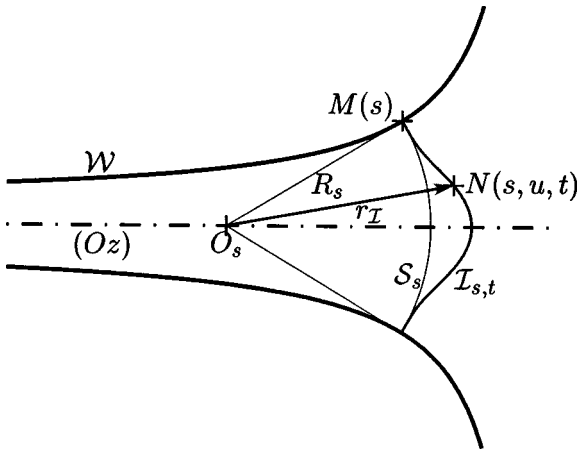


FIG. 4. $O_s(s) \in (Oz)$ and $R_s(s)$ are the center and the radius of the sphere $\mathcal{S}_s$ tangent to $\mathcal{T}_{s,t}$ at $M(s) \in \mathcal{W}$. $r_{\mathcal{I}}(s,u,t)$ is the distance between O_s and the point $N \in \mathcal{T}_{s,t}$ located at (s,u,t). It is assumed that $\mathcal{T}_{s,t}$ and $\mathcal{S}_s$ are close at the second order at $M(s)$. Namely, for each (s,t), $u \mapsto r_{\mathcal{I}}(s,u,t)$ is approximated by the constant $R(s)$ at the second order for u in the vicinity of w.

$R_s(s)$ of the sphere $\mathcal{S}_s$, and for any point $N(s,u,t)$, the distance $r_{\mathcal{I}}(s,u,t)$ between O_s and N are defined.

By definition, the radius $\mathbf{O_sM}$ is orthogonal to the sphere $\mathcal{S}_s$ and so to $\mathbf{u_u}$ at $M(s)$. Writing that the scalar product of $\mathbf{O_sM}$ and $\mathbf{u_u}$ is zero, $z_{O_s}(s)$ is proven to be such that

$$g|_w = \tan\phi|_w (f|_w - z_{O_s}). \tag{41}$$

The positive radii $R_s(s)$ and $r_{\mathcal{I}}(s,u,t)$ are given by

$$R_s^2 = \left.\frac{g^2}{\sin^2\theta}\right|_w = \left.\frac{g^2}{\sin^2\phi}\right|_w, \tag{42a}$$

$$r_{\mathcal{I}}^2 = g^2 + (f - z_{O_s})^2. \tag{42b}$$

The relative divergence $\varsigma(s,u,t)$ is then introduced by

$$\varsigma = \frac{r_{\mathcal{I}}}{R_s} - 1. \tag{43}$$

ς may be interpreted as an indicator of the deformation engendering $\mathcal{I}_{s,t}$ from $\mathcal{S}_s$. Note that if the wall $\mathcal{W}$ is locally parallel to (Oz), ς may keep a meaning by continuation, although R_s becomes infinite. In particular, the study below proves that $\varsigma|_w$ is zero at the first order at least.

By definition, $r_{\mathcal{I}}(s,w,t) = R_s(s)$ so that

$$\varsigma|_w = 0. \tag{44}$$

Now, $\partial_u \varsigma = \partial_u(r_{\mathcal{I}}^2)/(2 r_{\mathcal{I}} R_s)$. From Eq. (42b), Eq. (9c), and Eq. (9d),

$$\tfrac{1}{2}\partial_u(r_{\mathcal{I}}^2) = \sigma_u[g\cos\phi - (f - z_{O_s})]. \tag{45}$$

The definition of z_{O_s} shows that Eq. (45) is zero for $u = w$ so that

$$(\partial_u \varsigma)|_w = 0. \tag{46}$$

Note that, in fact, this last equation is related to the tangency of $\mathcal{S}_s$ and $\mathcal{I}_{s,t}$ at $M(s)$. Finally,

$$\partial_u^2 \varsigma = \frac{\partial_u^2(r_{\mathcal{I}}^2)}{2 r_{\mathcal{I}} R_s} - \frac{(\partial_u(r_{\mathcal{I}}^2))^2}{2 r_{\mathcal{I}}^2 R_s},$$

so that the second term is zero in $u = w$ [see Eq. (45)], and

$$\tfrac{1}{2}\partial_u^2(r_{\mathcal{I}}^2) = \partial_u \sigma_u[\cos\phi - (f - z_{O_s})\sin\phi] + \sigma_u^2 - \sigma_u \partial_u \phi[g\sin\phi + (f - z_{O_s})\cos\phi], \tag{47}$$

for which the first term is zero in $u = w$. Then, $(\partial_u^2 \varsigma)|_w$ is given by

$$(\partial_u^2 \varsigma)|_w = \left.\left[\frac{\sigma_u}{R_s^2}\left(\sigma_u - \frac{g\partial_u\phi}{\sin\phi}\right)\right]\right|_w. \tag{48}$$

The geometrical hypothesis proposed in this section consists of assuming that near the wall $\mathcal{W}$, isobars $\mathcal{I}_{s,t}$ deviate from $\mathcal{S}_s$ "slower than a parabola," namely,

$$(\partial_u^2 \varsigma)|_w = 0. \tag{49}$$

As $\sigma_u / R_s^2 \neq 0$, this hypothesis yields

$$\left.\frac{\partial_u \phi}{\sigma_u}\right|_{s,w,t} = \left.\frac{\sin\phi}{g}\right|_{s,w,t}, \tag{50}$$

which will make $A_s|_w$, and so Eq. (20) for $u = w$, possibly be expressed explicitly with $f|_w$ and $g|_w$. Indeed, the right-hand side of Eq. (50) is known from the shape of $\mathcal{W}$.

C. Mono-space wave equation

As $\epsilon|_w = 0$, $\sin\phi/\cos\delta = \sin\theta = \partial_s g/\sigma_s$ for $u = w$. The geometrical hypothesis and this last equation make the evaluation of $A_s|_w$ possible. Finally, recalling that $\xi_s|_w = \xi_n|_w = 0$, this leads to the significantly simplified equation

$$A_{s,s}|_w \partial_s^2 \tilde{p} + A_s|_w \partial_s \tilde{p} - \frac{1}{c^2}\partial_t^2 \tilde{p} = 0, \tag{51}$$

where $A_{s,s}|_w$ and $A_s|_w$ are given by

$$A_{s,s}|_w = \left.\frac{1}{\sigma_s^2}\right|_w, \tag{52a}$$

$$A_s|_w = \left.\frac{2\partial_s \ln g}{\sigma_s^2}\right|_w. \tag{52b}$$

Furthermore, because $\epsilon|_w = 0$, Eq. (19) leads to

$$\mathbf{grad}(\tilde{p}) = \frac{\partial_s \tilde{p}}{\sigma_s}\mathbf{u_s}, \tag{53}$$

on $\mathcal{W}$.

The obtained 1D model is available for any choice made for s, all the deduced models being equivalent. This model is now established for two specific and natural parametrizations. The first one consists of choosing $s = z$ so that $g|_w(s) = R(s)$ represents the radius $R(z)$ of the section of the waveguide located by the z-ordinate $z = f|_w(s) = s$. In this case, $\sigma_s(z,w,t) = \sqrt{1 + R'(z)^2}$ leading to

$$\partial_z^2 \tilde{p}(z,t) + 2\frac{R'(z)}{R(z)}\partial_z \tilde{p}(z,t) - \frac{1 + R'(z)^2}{c^2}\partial_t^2 \tilde{p}(z,t) = 0. \tag{54}$$

Note that this equation does not model the propagation of the lowest mode since it differs from that established by Pagneux[10] [Eq. (45)] when higher-order modes are ignored. The second choice is to take $s = \ell$ where ℓ measures the arc length of the wall $\mathcal{W}$ so that $\sigma_s|_w$ simplifies into the constant 1. Noting $\mathcal{R}(\ell)$ the radius of the section of the waveguide located by ℓ, the obtained wave equation is that of Webster:

$$\partial_\ell^2 \tilde{p}(\ell,t) + 2\frac{\mathcal{R}'(\ell)}{\mathcal{R}(\ell)}\partial_\ell \tilde{p}(\ell,t) - \frac{1}{c^2}\partial_t^2 \tilde{p}(\ell,t) = 0. \tag{55}$$

Note that if $\ell = z = 0$ locates a referent point $M_0 \in \mathcal{W}$, the bijective relation between ℓ and z is given by

$$\ell = L(z) = \int_0^z \sqrt{1 + R'(z)^2}\,dz. \tag{56}$$

The radius $\mathcal{R}(\ell)$ is straightforwardly deduced from $R(z)$ (at least numerically) through the computation of $L(z)$ since

$$\mathcal{R}(\ell)=R(L^{-1}(\ell)). \tag{57}$$

This last choice makes it possible to attribute all the known properties of the Webster equation to the general model Eq. (51), such as the conservation of acoustic energy, or the existence of analytical solutions for particular profiles. For this reason, Eq. (55) is used in the following rather than Eq. (51) or Eq. (54).

D. Observations and discussion

As a first result, the hypothesis of quasisphericity Eq. (50) makes the arc length ℓ appear as the natural longitudinal ordinate for the 1D model. This exactly coincides with the ordinate for which Putland establishes the Webster equation. However, the coefficients ahead of $\partial_\ell \tilde{p}$ are distinct for both models. That of Putland is $\partial_\ell S(\ell,t)/S(\ell,t)$ where S denotes the unknown area of the isobaric (or the "ℓ-iso") surfaces[8] versus $2\mathcal{R}'/\mathcal{R}$ for Eq. (55) which is known and time-invariant. Considering $2\mathcal{R}'/\mathcal{R}$ as a ratio $S'(\ell)/S(\ell)$, where S would represent the plane section of the waveguide drawn for ℓ but not z, yields a similar formulation but remains physically distinct.

Moreover, this model holds various notable properties. First of all, it gives exact results for both straight and conical pipes. Thus, for a straight pipe defined by $R(z)=R_0$, and a conical pipe defined by $R(z)=z\tan\theta_0$ or $\mathcal{R}(\ell)=\ell\sin\theta_0$ [see Eq. (56)] with $\theta_0\in]0,\pi/2[$, the 1D model leads to

$$\partial_z^2\tilde{p}(z,t)-\frac{1}{c^2}\partial_t^2\tilde{p}(z,t)=0, \tag{58a}$$

$$\partial_\ell^2\tilde{p}(\ell,t)+\frac{2}{\ell}\partial_\ell\tilde{p}(\ell,t)-\frac{1}{c^2}\partial_t^2\tilde{p}(\ell,t)=0, \tag{58b}$$

respectively. As expected, Eq. (58a) and Eq. (58b) yield the exact models for plane waves and spherical waves.

Compared to the Webster equations established assuming plane, spherical, or oblate spheroidal wavefronts, appreciable improvements of the presented model may be highlighted. Not only is the orthogonality between isobars and the wall respected, but the unavoidable mobility of isobars (see Sec. III D) as well. As a matter of fact, this is a consequence of the locality as well as the minimal order of the hypothesis which only requires $\partial_u^2\mathfrak{s}|_w=0$ while $\mathfrak{s}|_w$ and $\partial_u\mathfrak{s}|_w$ are both naturally zero.

Unfortunately, most of the limitations known for the Webster equation remain. The validity of the geometrical hypothesis is restricted to a low frequency range (no transverse mode, e.g., Fig. 3) and smooth geometries so that high mode coupling will not occur.[10] Walls must have a small curvature as well as isobars.

In particular, the model makes the control of the geometry of isobars possible neither at the input nor at the output of the pipe: spheres appear as the best appropriate shapes for the quasisphericity hypothesis. Rigorously, these input–output geometries should be exhibited by solving Eq. (20) for (f,g) and knowing $(f|_w,g|_w)$, once the pressure $\tilde{p}(s,t)$ has been computed for the 1D model and the 1D boundary conditions. Note that, practically, the geometrical resolution could be driven with numerical methods and for the simplifying assumption $\epsilon=0$, leading to a description of the isobaric maps by time-varying orthogonal coordinate systems. But since 1D models are generally not exact, the pressure established for any of them should be linked to "aberrant maps" (unsolvable or not axisymmetric). The prospect of deriving a 1D model from Eq. (20) for new more relaxed geometrical hypotheses and making control possible on the boundary isobars will be discussed in the conclusion.

E. Particular profiles associated with analytical waves

Analytical solutions of the Webster equation are known for particular shapes defined by $(R_0>0)$

(i) $\mathcal{R}(\ell)=R_0\exp(\alpha\ell)$ (exponential),
(ii) $\mathcal{R}(\ell)=R_0\cosh(\alpha\ell)$ (catenoidal),
(iii) $\mathcal{R}(\ell)=R_0\sin(\alpha\ell)$ (sinusoidal),
(iv) $\mathcal{R}(\ell)=R_0\ell^\alpha$ (Bessel).

This section is dedicated to determining the physical profiles $R(z)$ corresponding to these four cases.

1. Analytical solutions

Noting $\psi(\ell,\omega)$ the Fourier transform of $\mathcal{R}(\ell)\tilde{p}(\ell,t)$ where ω denotes the angular frequency, the wave equation corresponding to the cases (i–iii) is obtained for

$$\partial_\ell^2\psi(\ell,\omega)-\left(\Upsilon-\frac{\omega^2}{c^2}\right)\psi(\ell,\omega)=0, \tag{59}$$

where $\Upsilon=\mathcal{R}''(\ell)/\mathcal{R}(\ell)$ is constant. The solutions in the Fourier domain are then

$$\psi(\ell,\omega)=\psi_0(\omega)e^{r(\omega)\ell}+\psi_1(\omega)e^{-r(\omega)\ell}, \tag{60}$$

where $r(\omega)$ is a square root of $\Upsilon-(\omega/c)^2$ and ψ_0, ψ_1 are arbitrary. When $\Upsilon>0$, a cutoff pulsation $\omega_c=c\sqrt{\Upsilon}$ may be defined: $r(\omega)$ is imaginary and the corresponding wave is propagative only for $\omega\geqslant\omega_c$. Note that although this equation is usually written in this form, Berners[15] has shown that the travelling modes which constitute the Fourier basis set do not furnish a complete set for convex profiles ($\Upsilon<0$). In this case, the equation Eq. (59) must be considered in the Laplace domain rather than that of Fourier. Indeed, he has shown that the set of eigenfunctions may include some so-called "trapped modes." Considering the associated Sturm–Liouville problem, he has developed a method which leads to well-posed numerical solutions.

For the Bessel horns (iv), noting $P(\ell,\omega)$ the Fourier transform of $\tilde{p}(\ell,t)$, the corresponding wave equation is

$$\partial_\ell^2P(\ell,\omega)+\frac{2\alpha}{\ell}\partial_\ell P(\ell,\omega)+\frac{\omega^2}{c^2}P(\ell,\omega)=0, \tag{61}$$

so that solutions take the form

$$P(\ell,\omega)=P_0(\omega)\ell^{1/2-\alpha}J_{\alpha-1/2}\left(\frac{\ell\omega}{c}\right)+P_1(\omega)\ell^{1/2-\alpha}Y_{\alpha-1/2}\left(\frac{\ell\omega}{c}\right), \tag{62}$$

TABLE II. This table sums up for the cases (i)–(iv) the expression of F used to compute $\mathcal{R}'$ from $\mathcal{R}$, and the minimum [(iii),(iv) $0<\alpha<1$] or maximum [(i),(ii),(iv) $\alpha\notin[0,1]$] radius R^* for which an infinite slope is reached on the physical shape.

Case	$F(R)$	R^*
(i) $\alpha>0$	αR	α^{-1}
(ii) $\alpha>0$	$\alpha\sqrt{R^2-R_0^2}$	$\sqrt{R_0^2+\alpha^{-2}}$
(iii) $\alpha>0$	$\alpha\sqrt{R_0^2-R^2}$	$\sqrt{R_0^2-\alpha^{-2}}$ if $\lvert\alpha R_0\rvert\geqslant 1$
(iv) $\alpha\neq 0$ $\alpha\neq 1$	$\alpha R_0^{(1/\alpha)}R^{(\alpha-1)/\alpha}$	$R_0^{1/(1-\alpha)}\lvert\alpha\rvert^{\alpha/(1-\alpha)}$

where J_ν and Y_ν are the Bessel functions of the first and the second kind[16] (Chap. 9), respectively, and P_0, P_1 are arbitrary.

Nevertheless, such profiles are known when the Webster equation is written for the variable z rather than ℓ. Thus, although analytical results are unchanged for the pressure resolution, that is not the case for the real physical shapes. In particular, the new computed profiles have an unusual characteristic: for some of them, their radius or their length is required to be bounded.

2. Properties of physical shapes

Let $z\mapsto L(z)$ be the length of the wall $\mathcal{W}$ from $z_0=0$ to z, defined by Eq. (56). Differentiating the expression $R(z)=\mathcal{R}(L(z))$ gives

$$\mathcal{R}'(L(z))=\frac{R'(z)}{\sqrt{1+R'(z)^2}}. \tag{63}$$

This implies that $\lvert\mathcal{R}'\rvert\leqslant 1$, the limit case $\lvert\mathcal{R}'\rvert=1$ corresponding to an infinite slope for the physical shape. Except for the sinusoidal profile when $\lvert\alpha R_0\rvert<1$, the formula $\mathcal{R}(\ell)$ remains physically meaningful only on lower- or upper-bounded intervals so that a maximum or minimum radius R^* associated with a length ℓ^* may be defined (see Table II).

Note that if a $\mathcal{C}^1$-regular profile ends with a slope $\mathcal{R}'=1$ at $\ell=\ell^*$, prolonging the profile by the cone defined by $\mathcal{R}'(\ell)=1$ for $\ell\geqslant\ell^*$ corresponds to a baffled pipe. This profile still satisfies the $\mathcal{C}^1$-regularity. In this case, Eq. (55) implies that, for $\ell\geqslant\ell^*$, the pressure propagates as spherical waves.

3. Computation of physical shapes

The computation of $R(z)$ from $\mathcal{R}(\ell)$ can be achieved in an implicit way if there exists an F such that $\mathcal{R}'=F(\mathcal{R})$. In this case, Eq. (63) yields

$$\frac{R'(z)}{\sqrt{1+R'(z)^2}}=F(\mathcal{R}(L(z)))=F(R(z))$$

which is proven equivalent to

$$\frac{F(R(z))}{\sqrt{1-F(R(z))^2}}R'(z)=1. \tag{64}$$

If $z\mapsto F(R(z))/\sqrt{1-F(R(z))^2}$ is bijective, the integration of Eq. (64) from z_0 to z may be written

$$\int_{R(z_0)}^{R(z)}\frac{F(r)}{\sqrt{1-F(r)^2}}\mathrm{d}r=z, \tag{65}$$

which solves z as a function of the radius. The functions F associated with the shapes (i)–(iv) are specified in Table II. The geometrical differences between these profiles drawn for both z and ℓ may be assessed in Fig. 5. Note that this figure exhibits the maximum and minimum radii.

V. GENERALIZATION FOR SMALL WALL ADMITTANCES AND MOBILE WALLS

When the wall $\mathcal{W}_t$ is mobile or is not ideally rigid and lossless, the normal component of the acoustic velocity is no longer zero. Nevertheless, the wall $\mathcal{W}_t$ may still be described by $\mathcal{T}_{w,t}$ for a constant $u=w$, and the general notations Eq. (38) to Eqs. (40) remain valid. The only differences are that $f|_w$ and $g|_w$ may be time-varying and that $\epsilon|_w\neq 0$ since the wall condition ($\tilde{\mathbf{v}}\cdot\mathbf{w_s}|_w\neq 0$) prevents isobars from being orthogonal to $\mathcal{W}_t$. In this case, the boundary conditions on $\mathcal{W}_t$ may be described by a relation linking $\tilde{p}$ and $\tilde{\mathbf{v}}\cdot\mathbf{w_s}|_w$. Such a relation is usually specified in the Fourier domain using a *wall admittance* Y.

As an example, if a quasiplanar wall is vibrating, a standard boundary condition is described for $u=w$ by[17] (p. 47)

$$\partial_{\mathbf{w_s}}\hat{p}+\frac{i\omega}{c}Y(\omega)\hat{p}=-i\omega\rho_0\hat{\mathbf{V}}\cdot\mathbf{w_s}, \tag{66}$$

where $\partial_{\mathbf{w_s}}=\mathbf{w_s}\cdot\mathbf{grad}$ is the derivative in the direction $\mathbf{w_s}$, $Q\mapsto\hat{Q}$ gives the Fourier transform defined for the angular frequency ω, and $\mathbf{V}$ is the local velocity of $\mathcal{W}_t$. $Y=\rho_0 c[\mathbf{w_s}\cdot(\hat{\mathbf{v}}-\hat{\mathbf{V}})]|_w/\hat{p}$ is the specific admittance of the material constituting the wall. Note that Y is ordinarily given such that $\mathbf{w_s}$ is outwardly directed, so that particular attention must be paid to the conventions used. This direction is achieved choosing $g|_w\geqslant 0$ and $f|_w$ such that the inside of the guide is at the right-hand side of $\mathbf{u_s}$.

This section establishes a mono-space wave equation in this general context for cases such that $\epsilon|_w\ll 1$, making $\epsilon^2|_w$ and $\epsilon^3|_w$ negligible. This restriction still encompasses many physical cases for which the wall admittance and thus $\partial_{\mathbf{w_s}}\tilde{p}|_w$ are small, as clarified below by Eq. (67). The reason for this restriction is also detailed.

A. Derivation of the 1D model

Establishing a 1D model from Eq. (20) requires coping with two problems. The first one consists of making the boundary condition usable, showing a relation between $\epsilon|_w$ and $\partial_{\mathbf{w_s}}\tilde{p}$. This stage stands in for the simplification run in Sec. IV considering that $\epsilon=0$. Projecting Eq. (19) on $\mathbf{w_s}$ leads to this requisite identification given by

$$\epsilon=\sigma_s\frac{\partial_{\mathbf{w_s}}\tilde{p}}{\partial_s\tilde{p}}. \tag{67}$$

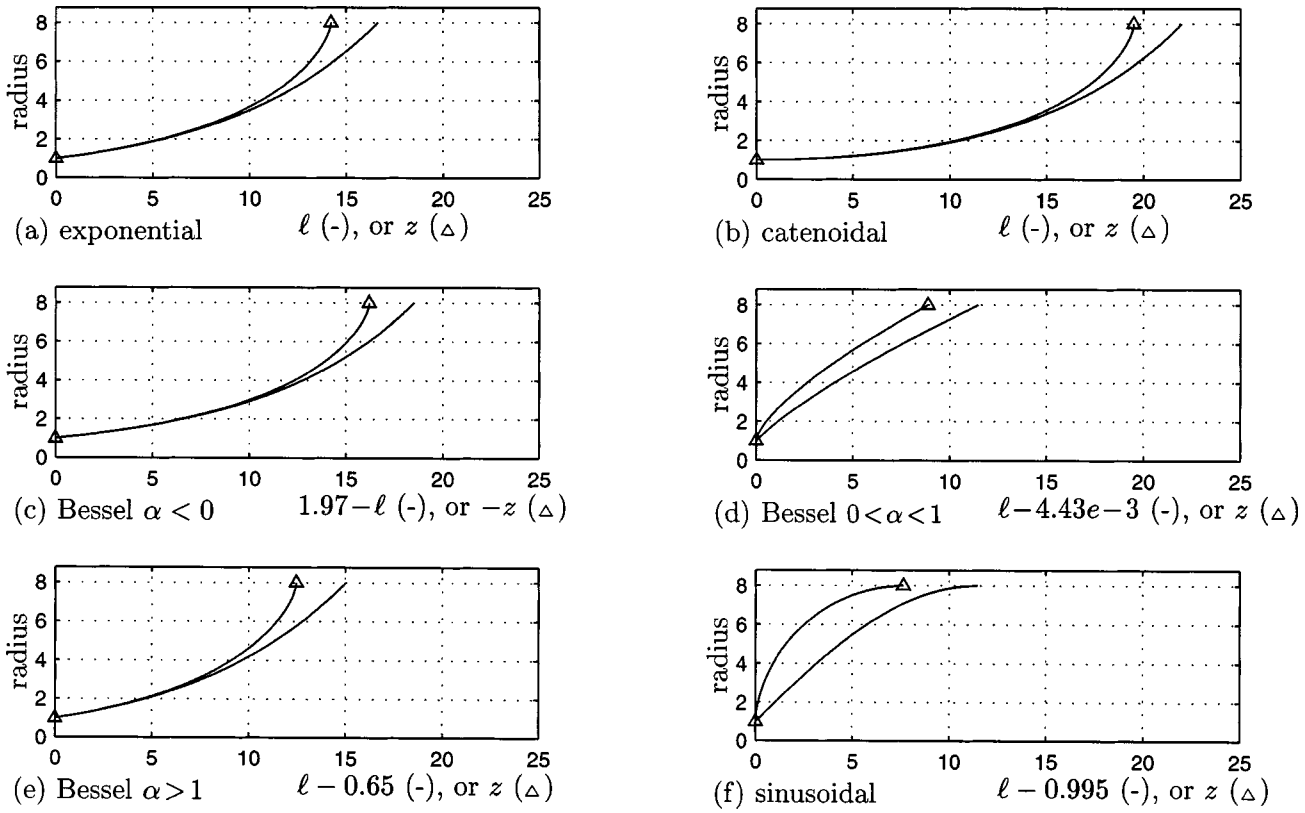


FIG. 5. The profiles corresponding to the cases (i)–(iv) are drawn for both the ℓ-ordinate (—), and the z-ordinate (△). Parameters are computed so that R^* is reached, $R_{\min}=1$ unit, and $R_{\max}=8$ units: (a) $R_0=1$, $\alpha=1/8$, (b) $R_0=1$, $\alpha=1/\sqrt{63}$, (c) $R_0\approx 8.6e-3$, $\alpha=-10$, (d) $R_0\approx 0.381$, $\alpha\approx 0.743$ [computed to have the same length for (d) and (f)], (e) $R_0\approx 0.745$, $\alpha=10$, (f) $R_0=8$, $\alpha=1/\sqrt{63}$. Appropriate translations in ℓ [and a symmetry for (c)] are used to make the profiles increasing and starting at $\ell=z=0$.

The main remaining problem is the evaluation of $\partial_u\phi/(\sigma_u\cos\delta)$ on $\boldsymbol{\mathcal{W}}_t$. Once again, the local quasisphericity of isobars near the wall makes the solution of the propagation problem separate from that of the geometry of isobars. Actually, this hypothesis is obtained as in Sec. IV B. The only noteworthy differences are that a point $M(s,t)$ of $\boldsymbol{\mathcal{W}}_t$ indexed by s can be nonstationary, and that $\epsilon|_w$ is not required to be zero. Nevertheless, all the definitions and the calculations [Eqs. (41)–(50)] remain exact, adapting the notations as illustrated in Fig. 6, and considering $O_{s,t}$, $z_{O_{s,t}}$, $R_{s,t}$ in place of O_s, z_{O_s}, and R_s in the previous formula.

Then, recalling that $\delta=\phi-\theta$, Eq. (50) yields

$$\left.\frac{\partial_u\phi}{\sigma_u\cos\delta}\right|_w=\left.\frac{\cos\theta+\epsilon\sin\theta}{g}\right|_w. \tag{68}$$

Using Eq. (67) and neglecting ϵ^2 and ϵ^3 in $A_{s,s}$ and A_s, the hypothesis of quasisphericity Eq. (68) leads to

$$B_{s,s}\partial_s^2\tilde{p}+B_s\partial_s\tilde{p}+B_{s,t}\partial_s\partial_t\tilde{p}-\frac{1}{c^2}\partial_t^2\tilde{p}+B_{\mathbf{w}_s}\partial_{\mathbf{w}_s}\tilde{p}$$
$$+B_{\mathbf{w}_s,s}\partial_s\partial_{\mathbf{w}_s}\tilde{p}+B_{\mathbf{w}_s,t}\partial_t\partial_{\mathbf{w}_s}\tilde{p}=0, \tag{69}$$

where

$$B_{s,s}=(1-\xi_s^2)/\sigma_s^2, \tag{70a}$$

$$B_s=\frac{2-\xi_n^2}{\sigma_s^2}\partial_s\ln g-\frac{1-\xi_n^2}{\sigma_s^2}\partial_s\ln\sigma_s$$
$$-\frac{1}{2}\partial_s\left(\frac{\xi_s^2-\xi_n^2}{\sigma_s^2}\right)+\frac{1}{c}\partial_t\left(\frac{\xi_s}{\sigma_s}\right), \tag{70b}$$

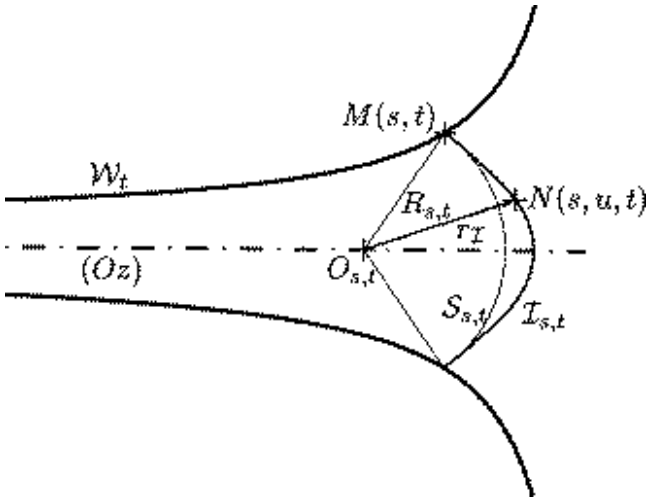


FIG. 6. When the wall is not rigid, motionless, or lossless, isobars are not necessarily orthogonal to the wall. The hypothesis of their quasisphericity near the wall can still be considered in this more general case. Note that the spheres $\boldsymbol{\mathcal{S}}_{s,t}$ tangent to isobars $\boldsymbol{\mathcal{I}}_{s,t}$ now move with respect to the angle $\phi(s,w,t)$ which is no longer a right angle. Nevertheless, the formula of definition Eq. (41) to Eq. (50) are not modified: the only differences are that O_s, z_{O_s}, and R_s may now depend on the time.

$$B_{s,t}=2\frac{\xi_s}{c\sigma_s}, \tag{70c}$$

$$B_{\mathbf{w_s}}=\frac{2-\xi_n^2}{g}\cos\theta+\frac{1-2\xi_n^2}{\sigma_s}\partial_s\theta-\frac{\xi_s}{\sigma_s}\partial_s\xi_n+\frac{1}{c}\partial_t\xi_n, \tag{70d}$$

$$B_{\mathbf{w_s},s}=-2\xi_s\xi_n/\sigma_s, \tag{70e}$$

$$B_{\mathbf{w_s},t}=2\xi_n/c, \tag{70f}$$

are all evaluated for $u=w$. Finally, Eq. (69) and boundary conditions such as Eq. (66) furnish a 1D model of the waveguide. As analyzed in Sec. IV D, the validity of this 1D model is still limited to walls having a small curvature. This assumption holds only if $\partial_s\theta|_w$ is negligible. Practically, $B_{\mathbf{w_s}}$ is well approximated by

$$B_{\mathbf{w_s}}=\frac{2-\xi_n^2}{g}\cos\theta-\frac{\xi_s}{\sigma_s}\partial_s\xi_n+\frac{1}{c}\partial_t\xi_n. \tag{71}$$

Particular derivations for visco-thermal losses on the wall or for mobile walls are described below. Beforehand, the requirement $\epsilon|_w\ll 1$ is clarified.

B. Geometrical hypothesis and compatibility with the problem

Let $\mathcal{P}$ describe the problem associated with the propagation in a given waveguide for which the wall conditions are linear with respect to the pressure. The linearity of the wave equation Eq. (3) and of the wall conditions ensures that of the acoustic propagation in the guide. Consequently, for such conditions, the exactness of the establishment of Eq. (20) and Eq. (67) guarantees the preservation of this property. However, if $\tilde{p}_1$ and $\tilde{p}_2$ are two solutions of these equations, $\tilde{p}=\tilde{p}_1+\lambda\tilde{p}_2$ does not appear as an obvious solution. Only $\lambda\tilde{p}_1$ or $\lambda\tilde{p}_2$ appears obvious when the motion of $\mathcal{W}_t$ is imposed. These particular properties are now described and investigated.

Under the linearity assumption of the wall conditions, the operator $\tilde{p}\mapsto\lambda\tilde{p}$ keeps ϵ invariant [see Eq. (67)]. It then acts on the first member of Eq. (20) as a multiplication by λ (first-order homogeneity) for imposed ξ_s and ξ_n. This proves that, for any solution $\tilde{p}$ of a problem $\mathcal{P}$, $\lambda\tilde{p}$ is also a solution. On the contrary, making the operator $(\tilde{p}_1,\tilde{p}_2)\mapsto\tilde{p}_1+\lambda\tilde{p}_2$ act on Eq. (20) and Eq. (67) does not lead to such relations. This is not a paradox but is simply due to the writing of $\mathcal{P}$ in the isobaric coordinate system. This indicates that the inherent linearity of $\mathcal{P}$ explicitly involves the coupling between the propagation and the geometry of isobars: multiplying a solution pressure by λ does not change the isobaric map, but another change of solution does. More precisely, it gives information about $\partial_u\phi/\sigma_u$ for which the compatibility with the quasisphericity hypothesis requires study.

Starting from Eq. (20) and Eq. (67), the exact wave equation for the problem $\mathcal{P}$ may be written as the nullity of the sum of three terms $T_{ne}[\tilde{p}]$, $T_{nl}[\tilde{p}]$, and $T_l[\tilde{p}]$ defined below. $T_{ne}[\tilde{p}]$ corresponds to the term of Eq. (20) which is not evaluable for $u=w$ from the single geometry of the wall. It corresponds to

$$T_{ne}[\tilde{p}]=(1-\xi_n^2)\left[1+\left(\frac{\sigma_s\partial_{\mathbf{w_s}}\tilde{p}}{\partial_s\tilde{p}}\right)^2\right]\frac{\partial_u\phi}{\sigma_u\cos\delta}\frac{\partial_s\tilde{p}}{\sigma_s}, \tag{72}$$

for which it is recalled that $\partial_u\phi/\sigma_u$ is the only quantity nonevaluable on $\mathcal{W}_t$, and where $\cos\delta$ is not developed using $\partial_{\mathbf{w_s}}\tilde{p}$ and $\partial_{\mathbf{u_s}}\tilde{p}$ to keep the formula compact. $T_{nl}[\tilde{p}]$ contains all other terms which are nonlinear in ϵ, and is given by

$$T_{nl}[\tilde{p}]=(1-\xi_n^2)\left[\left(\frac{\partial_{\mathbf{w_s}}\tilde{p}}{\partial_s\tilde{p}}\right)^2(\sigma_s\partial_s\theta\partial_{\mathbf{w_s}}\tilde{p}+\partial_s\ln\sigma_s\partial_s\tilde{p}-\partial_s^2\tilde{p})+2\frac{\partial_{\mathbf{w_s}}\tilde{p}}{\partial_s\tilde{p}}\partial_s\partial_{\mathbf{w_s}}\tilde{p}\right]. \tag{73}$$

The remaining term $T_l[\tilde{p}]$ is constituted only of linear terms evaluable for $u=w$: all the first-order terms such as $-2\xi_n\xi_s\epsilon\partial_s^2\tilde{p}/\sigma_s^2$ in $A_{s,s}$, or $2\xi_n\epsilon\partial_s\partial_t\tilde{p}/(c\sigma_s)$ in $A_{s,t}$, vanish with those of A_s when expressing ϵ thanks to Eq. (67). $T_l[\tilde{p}]$ would correspond exactly to the first member of Eq. (69) omitting the contribution in B_s and $B_{\mathbf{w_s}}$ of T_{ne} after using the quasisphericity hypothesis.

Defining $T[\tilde{p}]=T_{ne}[\tilde{p}]+T_{nl}[\tilde{p}]+T_l[\tilde{p}]$, the wave equation is obtained writing $T[\tilde{p}]=0$. The condition of linearity is then obtained writing

$$T[\tilde{p}_1+\lambda\tilde{p}_2]=T[\tilde{p}_1]+\lambda T[\tilde{p}_2], \tag{74}$$

for all $\tilde{p}_1$ and $\tilde{p}_2$ solutions, and for all λ. When ξ_s and ξ_n are unchanged for distinct solutions (for instance, rigid but controlled mobile walls), T_l is linear so that this condition is reduced to Eq. (74) taking $T[\tilde{p}]=(T_{ne}[\tilde{p}]+T_{nl}[\tilde{p}])/(1-\xi_n^2)$. Equation (74) exhibits the nonlinear relation between the three geometrical coefficients $(\partial_u\phi/\sigma_u)_{\mathcal{P}}$ associated with the problem $\mathcal{P}$ for the pressures $\tilde{p}=\tilde{p}_1$, $\tilde{p}_2$, $\tilde{p}_1+\lambda\tilde{p}_2$, and these three pressures.

The quasisphericity hypothesis does not mostly satisfy this condition since it makes $(\partial_u\phi/\sigma_u)|_w$ depend only on the wall geometry but not on the pressure. As a consequence, the linearity of the 1D model for this hypothesis requires $\epsilon|_w\ll 1$, so that the induced nonlinear terms may be neglected, leading to Eq. (69).

C. Influence of mobile walls

Equation (69) can be straightforwardly used to establish a 1D model for a controlled mobile wall.

From Eq. (10a) and Eq. (11a), the time derivatives of $\mathbf{u_s}$ and $\mathbf{w_s}$ are obtained for

$$\partial_t\mathbf{u_s}=\partial_t\theta\,\mathbf{w_s}, \tag{75a}$$

$$\partial_t\mathbf{w_s}=-\partial_t\theta\,\mathbf{u_s}. \tag{75b}$$

Since from Eqs. (14), $\mathbf{V}/c=\xi_s\mathbf{u_s}+\xi_n\mathbf{w_s}$, the acceleration $\partial_t\mathbf{V}$ is such that

$$\partial_t\left(\frac{\mathbf{V}}{c}\right)=(\partial_t\xi_s-\xi_n\partial_t\theta)\mathbf{u_s}+(\partial_t\xi_n+\xi_s\partial_t\theta)\mathbf{w_s}. \tag{76}$$

As $\partial_{\mathbf{w_s}}\tilde{p}=\mathbf{w_s}\cdot\mathbf{grad}(\tilde{p})$, Eq. (2) and Eq. (76) yields

$$\partial_{\mathbf{w}_s}\tilde{p} = -\rho_0 c(\partial_t \xi_n + \xi_s \partial_t \theta). \tag{77}$$

Equation (77), evaluated on the wall, and Eq. (69) furnish the 1D model.

Note that the hypotheses involved to linearize the Navier–Stokes equations and used to derive Eq. (3) require that $\xi_s \ll 1$ and $\xi_n \ll 1$, so that terms of high order in ξ_s or ξ_n may be neglected. Describing $\mathcal{W}_t$ with $\mathcal{R}(\ell,t) = g|_w(\ell,t)$, ℓ being the curvilinear ordinate such that $\sigma_s|_w(\ell,t) = 1$, the model takes a simpler form

$$\partial_\ell^2 \tilde{p} + \left(2\frac{\partial_\ell \mathcal{R}}{\mathcal{R}} + \frac{\partial_t \xi_s}{c}\right)\partial_\ell \tilde{p} + 2\frac{\xi_s}{c}\partial_t \partial_\ell \tilde{p} - \frac{1}{c^2}\partial_t^2 \tilde{p}$$
$$= \rho_0 c\left[\left(\frac{2\cos\theta}{\mathcal{R}} + \frac{\partial_t \xi_n}{c}\right) + 2\frac{\xi_n}{c}\partial_t\right](\partial_t \xi_n + \xi_s \partial_t \theta), \tag{78}$$

where the geometrical quantities are still evaluated for $u = w$ and the second member may be interpreted as sources induced by the motion of $\mathcal{W}_t$.

For many practical cases, the controlled motion or the vibrations of the wall are sufficiently small to consider much stronger approximations so that Eq. (78) can be reduced to

$$\partial_\ell^2 \tilde{p} + 2\frac{\partial_\ell \mathcal{R}}{\mathcal{R}}\partial_\ell \tilde{p} - \frac{1}{c^2}\partial_t^2 \tilde{p} = \rho_0 c \frac{2\cos\theta}{\mathcal{R}}(\partial_t \xi_n + \xi_s \partial_t \theta), \tag{79}$$

for $u = w$. For vibrating walls, this last acoustic equation may be coupled with the model of the wall vibrations.

D. Influence of visco-thermal losses

Equation (69) also enables treating the case of a pipe with visco-thermal losses due to the wall, now assumed motionless. The radius of the pipe is assumed to be large in relation to the boundary-layer thickness, but small compared to the wavelength.

Let significant parameters specifying the properties and the nature of the fluid at rest be defined: the coefficient of shear viscosity μ, the coefficient of thermal conductivity λ, the heat coefficients at constant pressure and constant volume per unit of mass C_P and C_V, the specific heat ratio $\gamma = C_P/C_V$, and finally the characteristic lengths $l_v' = \mu/(\rho_0 c)$ and $l_h = \lambda/(\rho_0 c C_P)$. The effect of the visco-thermal losses on the acoustics may be described for travelling waves by an equivalent specific wall admittance Y given by[17] (pp. 112–115)

$$Y(\omega,s) = \left(\frac{i\omega}{c}\right)^{1/2}[\kappa(s)\sqrt{l_v'} + (\gamma-1)\sqrt{l_h}], \tag{80}$$

where κ is linked to the angle of incidence of the wavefronts on $\mathcal{W}_t$ as described below. Note that to be physically meaningful, Y must have an Hermitian symmetry so that the complex $(i\omega)^{1/2}$ needs to be specified: this quantity may be understood as $\sqrt{|\omega|}\exp(i\pi/4)$ for $\omega \geqslant 0$, and the conjugate $\sqrt{|\omega|}\exp(-i\pi/4)$ for $\omega < 0$. More precisely, this definition makes it correspond to the time operator $\partial_t^{1/2}$ for causal functions.[18]

The validity of Eq. (80) relies on the fact that the thickness of the boundary layer must be very small with regard to both the radius and the radius of curvature of $\mathcal{W}$. For a pulsation ω, the thickness is given by $\sqrt{2cl_v'/\omega}$ for the viscous effects, and $\sqrt{2cl_h/\omega}$ for the thermal effects. For usual conditions, l_v' and l_h are about 4×10^{-8} m and 6×10^{-8} m so that the thermal effects are the most restrictive. The thickness decreases with the frequency f as $f^{-1/2}$ and is about 2.5 mm at $f = 1$ Hz for the thermal effects. This condition is then fulfilled for many practical cases.

Now, $\kappa(s)$ corresponds to the square of the sine of the angle[17] (p. 155) $[\mathbf{u_s}, \mathbf{grad}(\tilde{p})]$, or identically, to $\cos^2\delta = 1/(1-\epsilon^2)$. As ϵ^2 is neglected above, κ may be approximated by 1. Then, if $\mathbf{w_s}$ is outwardly directed, the boundary condition Eq. (66) (for $\mathbf{V} = \mathbf{0}$) is written in the time domain by

$$\partial_{\mathbf{w}_s}\tilde{p} + \frac{\sqrt{l_v'} + (\gamma-1)\sqrt{l_h}}{c^{3/2}}\partial_t^{3/2}\tilde{p} = 0, \tag{81}$$

where for usual conditions $\sqrt{l_v'} + (\gamma-1)\sqrt{l_h}$ is about 3×10^{-4} m$^{1/2}$. This order of magnitude confirms that the effect due to the boundary layer is small so that the assumption $\epsilon \ll 1$ is well founded in this case, and the quasisphericity hypothesis is compatible with the problem.

Finally, the dominating practical requirement is the slowness of the variation of the cross section of the pipe. The 1D model obtained for these pipes from Eq. (69) and Eq. (81) with $\xi_s|_w = \xi_n|_w = 0$ is given for $u = w$ by

$$\frac{1}{\sigma_s^2}\partial_s^2 \tilde{p} + \frac{2\partial_s \ln g}{\sigma_s^2}\partial_s \tilde{p} - \frac{1}{c^2}\partial_t^2 \tilde{p} - \frac{2\cos\theta}{g}\frac{\Gamma}{c^{3/2}}\partial_t^{3/2}\tilde{p} = 0, \tag{82}$$

where $\Gamma = \sqrt{l_v'} + (\gamma-1)\sqrt{l_h}$. For the ordinate ℓ for which $\sigma_s|_w = 1$, this equation appears as a Webster equation perturbed by the low fractional differential term $2(\cos\theta/(c^{3/2}g))\Gamma\partial_t^{3/2}\tilde{p}$. Note that for the profile $g|_w(s) = R_0$, this equation exactly yields the well-known equation of plane waves guided in cylindrical tubes with visco-thermal losses[17] (p. 145).

VI. CONCLUSION

A rigorous derivation of the linear acoustic wave equation in any local isobaric coordinate system has been presented for axisymmetric problems. As the main theoretical result of this work, it formally demonstrates the exact coupling between the geometry of the isobars and the propagation of the pressure. Straightforward derivations have shown that any regular isobaric map may satisfy a purely geometrical criterion of admissibility. The general static isobaric maps have been proven to be parallel planes, coaxial cylinders, and concentric spheres. Other static maps exist but, in this case, only a pure sinusoid or a pure exponential can travel on each of them. Furthermore, the isobaric wave equation shows that separating the resolution of the isobaric map from that of the 1D pressure is usually impossible. Assuming the quasisphericity near the wall for a minimal order is proven to be sufficient to dispose of this problem. The 1D models derived for motionless rigid walls or mobile ones with small admittances constitute the second main result of

this work. They make the arc length of the wall appear as the natural distance travelled by a wave with the speed c.

The interests are to draw benefit from the low computational cost of 1D models, and to improve the quality, thanks to a prior, quite complex differential calculus and a geometrical hypothesis which is weaker than usual. A spatial resolution for the 1D models having motionless walls yields a representation with input–output systems that are not expensive to simulate[19] and make real-time applications possible.

Even if the geometrical hypothesis does not require fixed wavefronts as usual, the limitations mostly remain those of the classical Webster equation, namely, smooth wall contours with low curvature, and sufficiently large wavelengths ensuring that transverse modes are not excited. To quantify the quality of the 1D models, a numerical validation will be run comparing the pressure deduced for them to that computed near the wall with the PAK algorithm,[10,11] but also with finite-element methods. In particular, the horns presented in Sec. IV E could be tested for an excitation $\tilde{p}(0,t)$ at the input $s=0$ and for a given load admittance at the output $s=L$ such as that of divergent spherical waves. For the 2D algorithms, the same boundary conditions may be taken on the spheres orthogonal to the wall in $s=0$ and $s=L$, the quasisphericity hypothesis being appropriate to this geometry. Other meaningful comparisons with 2D models may be done for boundary conditions which are compatible with the quasisphericity hypothesis and adaptable to 1D models [e.g., radiating portion of a sphere[17] (p. 246)].

Nevertheless, whatever the success of a numerical validation, this work can be expanded to other models having a mono-spatial dependence, which would exceed the above-mentioned limitations. Indeed, starting from the general rigorous Eq. (20), geometrical hypotheses more relaxed than that of quasisphericity could be used. Mainly, choosing hypotheses of higher orders is an interesting prospect. A particularly interesting one is an order of regularity and of flexibility imposed on the wavefront geometry by $\partial_{\mathbf{u}_\mathbf{u}}^K \phi|_w = 0$ [with $\partial_{\mathbf{u}_\mathbf{u}} = (1/\sigma_u)\partial_u$] for a given $K \geqslant 2$ ($K=0$ would impose the wrong angle $\phi|_w = 0$, and $K=1$ would impose quasiplanar wavefronts near the wall rather than the more appropriate quasispherical ones). But such alternatives will involve extensive investigation. Indeed, establishing the associated 1D models requires solving the system of equations obtained by applying $X \mapsto \partial_{\mathbf{u}_\mathbf{u}}^k X|_w$ on Eq. (20) for $k=0,1,...,K-1$, the hypothesis being used in the last equation. A careful study of fundamental properties such as the linearity or the compatibility with the symmetry of the problem must be performed since they are not guaranteed *a priori*. In addition to the relaxation of the above-mentioned constraints induced by the quasisphericity hypothesis, the main interest of this extension is potentially having control of the input–output isobar geometry through the integrating constants linked to $\partial_{\mathbf{u}_\mathbf{u}}^k \phi|_w$. In this case, the corresponding 1D models account for the propagation of the pressure as well as that of the geometrical information represented.

ACKNOWLEDGMENTS

The research reported here is based on a doctoral dissertation completed by the author at the University of Paris XI-Orsay. The author thanks X. Rodet, the Analysis–Synthesis team at IRCAM, and the Ministry of Research for supporting this research. The author is very grateful to J. Kergomard and C. Vergez for their critical reading of this paper, as well as S. Mac Adams and M. Lo Cascio for proofreading this paper before submission to the Journal of the Acoustical Society of America.

[1] D. Bernoulli, ''Sur le son et sur les tons des tuyaux d'orgues différemment construits'' (Physical, Mechanical and Analytical Researches on Sound and on the Tones of Differently Constructed Organ Pipes), Mém. Acad. Sci. (Paris), 1764 [in French, reference extracted from the paper of Eisner (Ref. 4)].

[2] J. L. Lagrange, ''Nouvelles recherches sur la nature et la propagation du son'' (New Researches on the Nature and Propagation of Sound), Misc. Taurinensia (Mélanges Phil. Math., Soc. Roy. Turin), 1760–1761 [in French, reference extracted from the paper of Eisner (Ref. 4)].

[3] A. G. Webster, ''Acoustical impedance, and the theory of horns and of the phonograph,'' Proc. Natl. Acad. Sci. U.S.A. **5**, 275–282 (1919); **6**, 320(E) (1920).

[4] E. Eisner, ''Complete solutions of the ''Webster'' horn equation,'' J. Acoust. Soc. Am. **41**, 1126–1146 (1967).

[5] R. F. Lambert, ''Acoustical studies of the tractrix horn. I,'' J. Acoust. Soc. Am. **26**, 1024–1028 (1954).

[6] E. S. Weibel, ''On Webster's horn equation,'' J. Acoust. Soc. Am. **27**, 726–727 (1955).

[7] A. H. Benade and E. V. Jansson, ''On plane and spherical waves in horns with nonuniform flare. I. Theory of radiation, resonance frequencies, and mode conversion,'' Acustica **31**, 79–98 (1974).

[8] G. R. Putland, ''Every one-parameter acoustic field obeys Webster's horn equation,'' J. Audio Eng. Soc. **6**, 435–451 (1993).

[9] A. F. Stevenson, ''Exact and approximate equations for wave propagation in acoustic horns,'' J. Appl. Phys. **22**(12), 1461–1463 (1951).

[10] V. Pagneux, N. Amir, and J. Kergomard, ''A study of wave propagation in varying coss-section waveguides by modal decomposition. Part I. Theory and validation,'' J. Acoust. Soc. Am. **100**, 2034–2048 (1996).

[11] J. A. Kemp, ''Theoretical and experimental study of wave propagation in brass musical instruments,'' Ph.D. thesis, Edinburgh University, 2002.

[12] J. Agulló, A. Barjau, and D. H. Keefe, ''Acoustic propagation in flaring, axisymmetric horns: I. A new family of unidimensional solutions,'' Acustica **85**, 278–284 (1999).

[13] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).

[14] *Field Theory Handbook, Including Coordinate Systems, Differential Equations and their Solutions*, 2nd ed. (1971) corrected 3rd printing (1988) (Springer-Verlag, New York, Heidelberg, Berlin, 1988), pp. 34–36.

[15] D. P. Berners, ''Acoustics and signal processing techniques for physical modeling of brass instruments,'' Ph.D. thesis, Standford University, 1999.

[16] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1970).

[17] M. Bruneau, *Manuel d'acoustique fondamentale (Manual of fundamental acoustics), Etudes en mécanique des matériaux et des structures (Studies in the mechanics of materials and structures)* (Hermès, Éditions Hermès, Paris, France, 1998).

[18] D. Matignon, ''Stability properties for generalized fractional differential systems,'' ESAIM Proc. **5**, 145–158 (1998) (available on the World Wide Web).

[19] T. Hélie, ''Modélisation physique des instruments de musique en systèmes dynamiques et inversion (Physical modeling of musical instruments with dynamic systems and inversion processes),'' Ph.D. thesis, Université de Paris XI-Orsay, IRCAM-Centre Georges Pompidou, Paris, 2002.

Detection of object resonances by vibro-acoustography and numerical vibrational mode identification

Farid G. Mitri,[a] Philippe Trompette, and Jean-Yves Chapelon
National Institute of Health and Medical Research, INSERM Unit 556—Therapeutic Ultrasound Research Laboratory, 151 Cours Albert Thomas, 69424 Lyon, cedex 03, France

(Received 25 January 2003; accepted for publication 28 July 2003)

Chalk sphere and cylinder resonance frequencies related to compressional and bending modes were detected in water, using vibro-acoustography, a relatively new imaging technique. The variable (radiation) force of low-frequency excitation, produced by intersecting two primary focused ultrasound waves with slightly different frequencies, forces the object to vibrate. The low-frequency acoustic emission field, resulting from object vibration, was detected by a hydrophone. By fixing the object at the focus of the ultrasound beam and sweeping the frequency of one of the primary beams within a chosen bandwidth, it was possible to detect some of the resonance frequencies (those related to compressional and bending modes) via variations in acoustic emission amplitude. Experimental results showed excellent agreement with finite element calculations. This method can be used to characterize the presence of heterogeneities in various media, in the field of materials science or biology. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1616921]

PACS numbers: 43.25.Qp, 43.30.Jx, 43.40.Rj [RLW] Pages: 2648–2653

I. INTRODUCTION

For many years, ultrasound techniques have been growing richer and more diversified. Apart from the usual fields of visualization (echography) and therapy, ultrasound techniques are also used in the vibratory field. That is how resonant ultrasound spectroscopy (RUS) was developed (Visscher *et al.*, 1991; Maynard, 1996). It provides information about the resonance spectrum of a structure free in space. The technique consists of connecting an ultrasound transmitter-receiver to the structure and thereby building transfer functions or frequency diagrams. Apart from the advantage of obtaining resonance frequencies for a very wide range of frequencies and working with a device independent of boundary conditions, RUS uses indirect measurement to obtain some of the characteristics of the structure's constituent material (or homogeneous equivalent material) in a frequency domain which is difficult to cover with the standard vibratory techniques.

Another equally interesting application is ultrasound stimulated vibro-acoustic spectrography (Fatemi and Greenleaf, 1998). This technique is quite recent and consists of focusing two very-high-frequency ultrasound waves, f_1 and f_2 which, if lightly shifted, produce a beating phenomenon at the focus and therefore an acoustic wave of frequency $\Delta f = f_2 - f_1$. In this way, all or part of a structure can be stimulated remotely at the following frequency Δf. First developed by Fatemi, vibro-acoustography has been used several times in materials science, such as in determining Young's modulus for an object, from its fundamental resonance frequency (Fatemi and Greenleaf, 1999), or in evaluating the mechanical characteristics of a medium by measuring the speed of vibration of a spherical target (Chen *et al.*, 2002). Another potential application is in visualizing calcifications in excised arteries (Fatemi and Greenleaf, 1998) and microcalcifications in mammary tissue (Fatemi *et al.*, 2002). This method was also introduced to help control ultrasound therapy, because the vibro-acoustic signal depends on temperature (Konofagou *et al.*, 2002).

In this paper, we explore the possibilities of RUS using the vibro-acoustic stimulation method. By combining these two techniques, and using a hydrophone, resonance frequencies of a structure immersed in a fluid (in this case water) can be obtained. This study is an extension of the work done by Fatemi and Greenleaf (1999) who calculated Young's modulus of an aluminum rod from its fundamental resonance frequency as measured using vibro-acoustography. Here, a wide frequency range is explored and a systematic identification (with finite element numerical calculations) of several vibrational modes related to resonance frequencies is proposed for two differently shaped objects (sphere and cylinder). It is shown by two distinct structural analyses (harmonic and modal) that only some vibrational modes are detected. An interpretation of the missing modes is given. It is also noted that the choice of porous, highly damped, nonlinear material of low rigidity (in this case chalk) demonstrates that vibro-acoustography can be used to detect heterogeneities similar to those found in biological tissue. The experiments and numerical simulations are in close agreement. Thus, it can be stated that by being positioned at any given resonance frequency, acoustic emission of an immersed object is optimal and consequently, even small objects could be detected with this method. Many applications, particularly those oriented towards the search for heterogeneities, are thus likely to be developed and are the subject of current investigations.

II. METHOD AND MATERIALS

A. Vibro-acoustography

To generate oscillating low-frequency radiation stress in a desired region of an arbitrary shaped object, two intersecting burst ultrasound beams are focused at slightly different

[a]Electronic mail: mitri@ieee.org

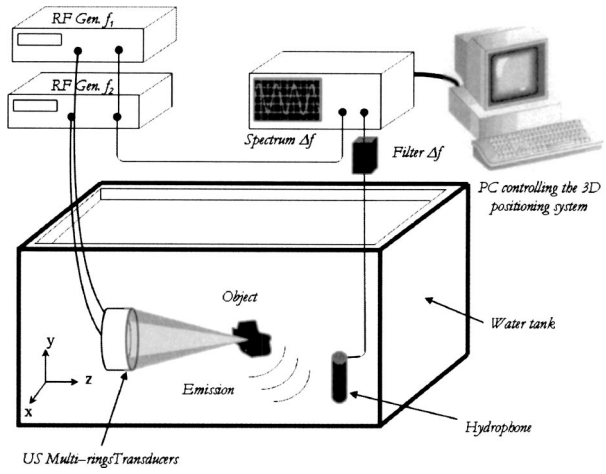


FIG. 1. Vibro-acoustic system setup.

frequencies f_1 and $f_2 = f_1 + \Delta f$ (Fig. 1). It is noted that the vibrating radiation force is only generated in the region of intersection where the ultrasound field energy is confined. This force is generally steady, since the intensity of the incident sound field does not change over time. In vibro-acoustography, the incident ultrasound is modulated in amplitude such that its energy density changes sinusoidally at low frequency. Thus the radiation force on the object is oscillatory. The time-varying radiation force (Torr, 1984), beating at frequency Δf, interacts with the object according to its mechanical properties. It produces a localized radiation stress field and causes a portion of the object, or the entire object, to vibrate at this frequency. In this interaction, the radiation force has one component in the direction of the beam and another transverse to it (Westervelt, 1951). Radiation force is expressed by (Fatemi and Greenleaf, 1998):

$$\overline{F} = d_r \int\int_S \overline{\xi}\, dS, \tag{1}$$

where d_r is the drag coefficient [or radiation force function (Yp)], $\overline{\xi}$ is the time-averaged energy density, and S is the projected area. The drag coefficient, which can be determined from the material properties (scattering and absorption) of the object and the surrounding medium, is generally frequency dependent. If the force vibration frequency corresponds to a normal mode, resonance will occur, which will produce compressional and shear wave motions inside the object. Object vibrations produces a sound field known as ''acoustic stimulated emission'' in the medium. The amplitude of the resulting acoustic emission field is (Fatemi and Greenleaf, 1998)

$$\Phi(\Delta f) = C d_r H(\Delta f)\, \mathcal{Q}(\Delta f), \tag{2}$$

where C is a constant proportional to the intensity of the primary ultrasound beams, $\mathcal{Q}(\Delta f)$ is a complex function representing the mechanical frequency response or admittance of the object at the selected point, and $H(\Delta f)$ represents the transfer function of the propagation medium and receiver, which is assumed to be fixed and known. The resonance frequency(ies) of the object can be determined by sweeping Δf within the range of interest while keeping the amplitude of the primary ultrasound beams constant, and recording the resultant acoustic emission $\Phi(\Delta f)$ from the object. $\Phi(\Delta f)$ depends directly on the product of both the drag coefficient d_r and the mechanical admittance $\mathcal{Q}(\Delta f)$ of the object. In general, the drag coefficient shows maxima and minima (Anson and Chivers, 1981; Hasegawa and Yosioka, 1969) related to certain resonances for wide variations in frequency. However, in our case, the drag coefficient is directly dependent on f_1 and f_2, and is of constant amplitude while sweeping f_2 between 2.220 and 2.275 MHz [low-frequency variations (20–75-kHz bandwidth)]. Acoustic emission variations [peaks of $\Phi(\Delta f)$] are therefore determined only from the resonances of $\mathcal{Q}(\Delta f)$.

The coupling mechanism to the objects under test is very complex with many nonlinear effects involved. The objective of this paper is not to discuss this complex mechanism in depth, which has been recently investigated (Callé *et al.*, 2002). In this work, we assume that the generation of low-frequency waves (acoustic emission) is dominated by the radiation pressure mechanism which is consistent with the results presented in the following sections.

B. Experimental setup

The corresponding experiments were performed in a water tank, which provided good ultrasonic and acoustic coupling together with free boundary conditions. A hydrophone was used to measure the resulting acoustic pressure field in the water tank.

A single circular focused piezoelectric transducer, with a diameter of 40 mm and focal length of 35 mm, was divided into two annular elements such that the areas of the two elements are equal. The two elements generated two ultrasound beams at two slightly different frequencies. The first ring was driven at 2.20 MHz, the second one at frequencies sweeping the range 2.22 to 2.275 MHz in increments of 0.5 kHz.

Radio frequency (rf) signals were obtained from two function generators (HM 8131-2 HAMEG, France) controlled by a pulse generator (8112 A, HP, Germany) and each transducer ring was fed by a rf amplifier (KMP Electronic, France). The transducer was mounted on a 3D positioning system (Micro-contrôle, France) and immersed in a tank of degassed water.

The two beams interacted at the focal region, producing an oscillatory radiation stress field on the object in an ellipsoidal volume. The spatial resolutions, defined by the diameter of the central lobe (minor axis), and the focal depth (major axis) of the transducer, were determined experimentally. A $25 \times 6\text{-mm}^2$ on-axis longitudinal plane was scanned, along which the two transducer elements were driven at 2.20 and 2.22 MHz, respectively. The emitted signal was set to repeat at a duty cycle of 318 μs-on/30 ms-off, corresponding to bursts of 700 cycles at 2.20 MHz. Acoustic pressure was measured using a needle hydrophone (PZT-Z44-0400, SEA, USA). The high-frequency signal received was acquired using a digital oscilloscope (2340 A Tektronix, USA) and the data then transferred via the IEEE-488 communication BUS to the PC controlling the positioning system. In Fig. 2, the

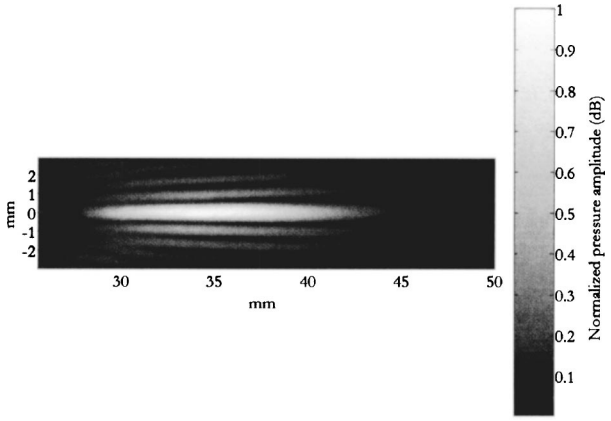


FIG. 2. Two-dimensional plot of spatial and depth resolutions of the stress field amplitude of the transducer at the focal plane.

major and minor full-widths at half maximum (FWHMs) of the ellipsoidal focal spot are equal to 10 and 1 mm, respectively, and define the volume where the low-frequency signal is confined.

Sound produced by vibrations of the object was detected by an audio hydrophone (SQ 03, SENSOR, Canada) placed in the water tank on one side of the object, not in the path of the ultrasound beam, to avoid interference with the transmitted signal. The hydrophone has a sensitivity of -162 dB *re* 1 V/μPa, and frequency response between 1 Hz and 65 kHz. The signal received was bandpass filtered (AF 420, Multimetrics, France) in the 10–80-kHz range, before acquisition using the digital oscilloscope. For greater precision relative to the resulting acoustic emission, the low-frequency signal was averaged 256 times during acquisition to eliminate noise and increase signal-to-noise ratio (SNR). The data was then transferred via the system described above.

C. Experimental procedure

Chalk objects with relatively low acoustic impedance were used because their characteristics closely resemble those of kidney stones and bone tissue.

The first object tested was a chalk sphere of the type used in lithotripsy experiments (15-mm diameter, HMT, Switzerland) and suspended like a pendulum. The corresponding mechanical properties are Young's modulus $E_0 = 4.1$ GPa, Poisson's ratio $\nu = 0.26$ (Heimbach *et al.*, 2000) and mass density $\rho = 1086\ \mathrm{Kg\,m^{-3}}$. In the experiment, the porous chalk sphere became saturated with water, so that the mechanical properties were changed (Schroeder and Shao, 1996). Sphere porosity was estimated as the ratio M_0/M, where M_0 is the measured mass in air and M is the "wetted mass" measured after immersing the sphere in water for 1 h. Porosity was found to be 25%. The Duckworth–Knudsen equation (Rice, 1977) was used to determine the effective Young's modulus E:

$$E = E_0 \exp(-bP), \tag{3}$$

where E is Young's modulus, E_0 is Young's modulus at zero porosity, P is porosity, and b is a variable porosity correction factor, set at unity. Equation (3) is only valid for a porosity of less than 50%. Therefore, the effective Young's modulus of the sphere, taking porosity into account, is $E = 3.19$ GPa, which will be used in the numerical simulations.

The second object tested was a blackboard chalk cylinder (10.38-mm diameter and 32.20-mm length). Its mass density was 1314 $\mathrm{Kg\,m^{-3}}$. The Young's modulus of the chalk was determined experimentally using a vibration test and was found to be $E_0 = 3.02$ GPa. The cylinder's porosity was 31%. The effective Young's modulus, according to Eq. (3), was then $E = 2.215$ GPa. All other parameters and procedures were identical to those used for the chalk sphere. Two series of the same experiment were run in the water tank. In the first series, the cylinder was suspended horizontally, and, in the second, vertically, with the ultrasound beam perpendicular to the axis of the cylinder. This was done to compare the two sets of measurements and verify the acoustic emission responses, which are expected to be nearly identical.

The modulated ultrasound beam was focused on the central region (axis of symmetry) to generate all the object resonances.

D. Finite element simulations

To verify the experimental results, a coupled harmonic acoustic analysis was performed using finite element software (Ansys, Inc.), and pressure distribution in the fluid was calculated in the case of a sinusoidally varying load. The driving nodal forces on both the sphere and cylinder were applied over an ellipsoidal volume corresponding to the dimensions of the focal spot. In the finite element harmonic analysis, a step frequency of 0.5 kHz was used. It was verified that the response does not vary significantly when the step value is decreased. It was assumed that the amplitude of the driving force does not vary while sweeping frequency. The coupled acoustic analysis involves modeling the fluid medium surrounding the structure, taking the fluid–structure interaction into account. The pressure amplitude has been calculated in the fluid. One particular node in the fluid was chosen and it was verified that the responses at neighboring nodes are comparatively the same.

The fluid medium was modeled using three-dimensional linear fluid elements surrounding the solid, and two-dimensional linear absorbing elements at the external bound-

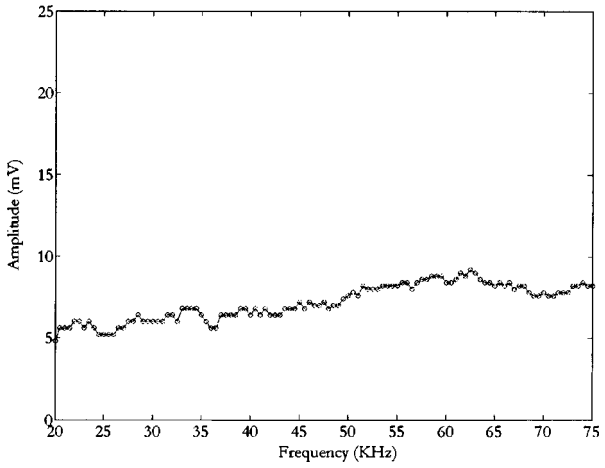


FIG. 3. Low-frequency signal amplitude in water.

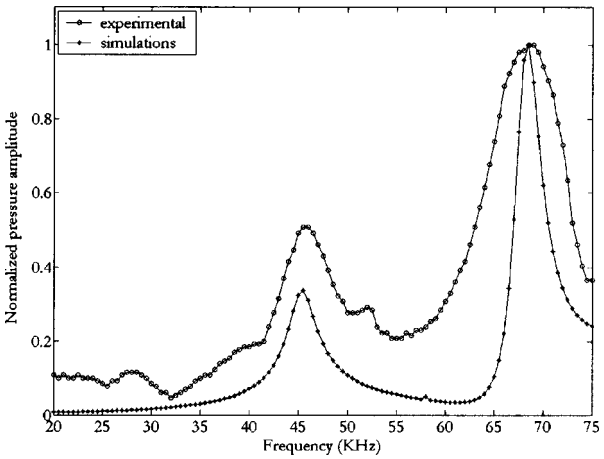


FIG. 4. Experimental and simulated vibro-acoustic amplitude response from the chalk sphere.

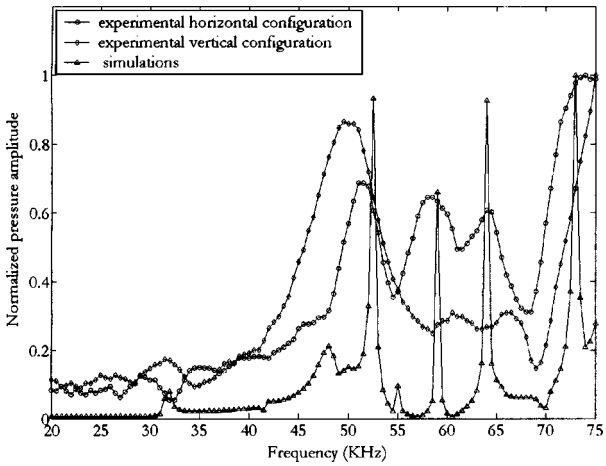


FIG. 5. Experimental and simulated vibro-acoustic amplitude response from the chalk cylinder in two configurations.

ary simulating the outgoing propagation effects of a domain extending to infinity. Absorbing elements eliminate reflections and therefore interference with vibration of the solid structure.

The chalk sphere was modeled using three-dimensional tetrahedral linear solid elements. The total numbers of nodes and elements were 2488 and 12629, respectively.

The chalk cylinder was also modeled using the same elements as were used for the sphere, and the total numbers of nodes and elements were 1081 and 4829, respectively.

To visualize the mode associated with a resonance frequency, modal analysis was performed in the fluid medium (see the appendix).

The pressure distribution was determined in the 20–75-kHz frequency range.

III. RESULTS

Each experiment was performed twice to check the repeatability of the acoustic emission response.

A. Nonobject test

Figure 3 shows the low-frequency signal amplitude in the water tank without an object. This was measured to verify that the water tank's resonance frequencies are outside the chosen bandwidth, so there was no interference phenomenon. In water and at low amplitude, ultrasound absorption is very weak, causing streaming, so that the signal detected at Δf was dominated by the nonlinear interaction of the two primary beams (Callé *et al.*, 2002). The intensity of the nonlinear parametric low-frequency field at Δf was proportional to the square of the low-frequency wave number $k=\Delta\omega/c$ (Muir and Willette, 1972). Figure 3 shows an increase in low-frequency signal amplitude versus frequency, as predicted by the theory.

TABLE I. Comparison of resonance frequencies for a chalk sphere calculated experimentally, and by simulations.

Sphere	Resonance frequencies values (kHz)	
Experimental	45.75	68.75
Simulations (Harmonic analysis)	45.5	68.5

B. Chalk sphere and cylinder models

Figure 4 shows the combined experimental and simulated frequency response of the sphere (acoustic emission amplitude versus frequency). Two main resonance frequencies are observed in the 20–75 kHz frequency range (Table I).

Figure 5 shows the combined experimental and simulated frequency response of the cylinder (acoustic emission amplitude versus frequency) in both the horizontal and vertical configurations. Five main resonance frequencies are detected in the 20–75-kHz frequency range (Table II).

IV. DISCUSSION

The method used in this paper is quite sensitive for the detection of resonance frequencies. Two and five experimental resonance frequencies are clearly detected for the sphere and cylinder, respectively. Complete interpretation of the experimental results is provided by both numerical harmonic and modal analyses using the finite element method. A coupled acoustic modal analysis, taking the fluid–structure interaction into account, was performed to visualize each mode Φ_i associated with the corresponding resonance frequency f_i for both models.

For the spherical model, two main resonance frequencies were detected at 45.5 and 68.5 kHz, and fully confirmed by the numerical harmonic analysis results (Fig. 4). The modal analysis printed out several types of mode shapes: (*i*) the dilatation modes, which are fully symmetrical (Fig. 6) and the displacement nodal components U_θ, U_ϕ have very

TABLE II. Comparison of resonance frequencies for a chalk cylinder for both configurations calculated experimentally and by simulations.

Cylinder	Resonance frequencies values (kHz)				
Experimental horizontal configuration	29	51.25	58.25	64	74
Experimental vertical configuration	31.5	49.5	60.5	66.25	>75
Simulations (Harmonic analysis)	32	52.5	59	64	73

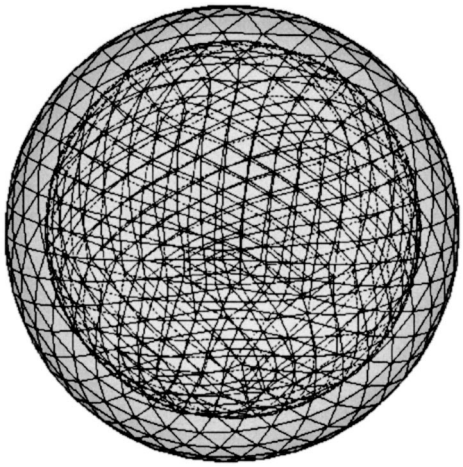


FIG. 6. Sphere dilatation mode (undeformed and deformed shapes).

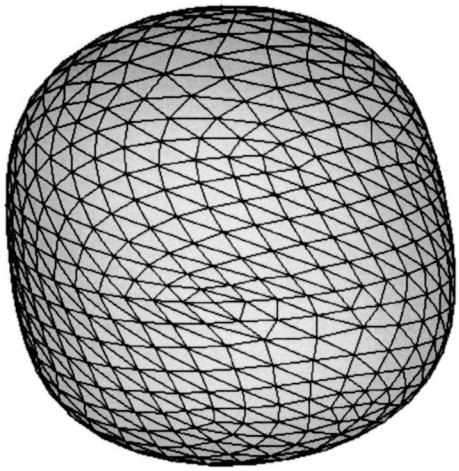


FIG. 7. Sphere torsional mode.

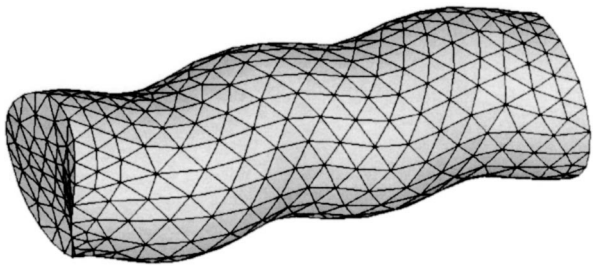


FIG. 8. Cylinder bending mode.

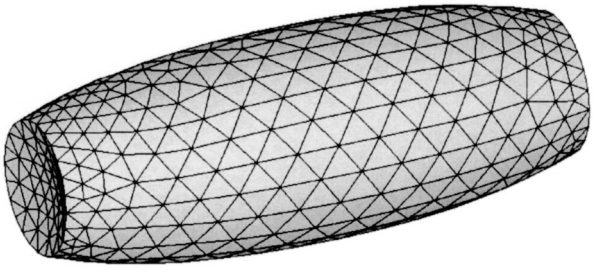


FIG. 9. Cylinder breathing mode.

small values relative to U_r, where r, θ, ϕ, are the spherical coordinates; and (*ii*) the torsional or shear modes, which are somehow difficult to describe, in which U_θ or U_ϕ may be large relative to U_r and in a different direction along the sphere's meridians and parallels (Fig. 7). It was verified that the two main resonance frequencies at 45.5 and 68.5 kHz correspond to dilatation modes (Fig. 6) generating compression waves, which can therefore be detected by the hydrophone. A third normal mode was detected by modal analysis at 58 kHz. It was found to be a torsional vibration mode (Fig. 7) generating shear waves that are not propagated in the fluid medium and are therefore not detected by the hydrophone.

For the cylinder model, harmonic analysis calculations show five resonance frequencies at 32, 52.5, 59, 64, and 73 kHz, which correlate well with the experimental results (Fig. 5). Several types of mode shapes were computed from the modal analysis: (*i*) the bending modes (Fig. 8) where the displacement components U_r are large relative to U_z and U_θ and in the opposite direction, perpendicular to the cylindrical axis, and r, z and θ are the cylindrical coordinates; (*ii*) the dilatation modes (Fig. 9), where U_r is very large relative to U_z and U_θ; and (*iii*) the torsional modes (Fig. 10) where U_θ is very large relative to U_r and U_z. Modal analysis of the cylinder is much more difficult to interpret since many complex mode shapes were observed in the frequency range. However, it was verified that the five detected resonance frequencies correspond to bending (Fig. 8) and breathing vibrational modes (Fig. 9).

The experimental results for the two configurations reveal good agreement with resonance frequency but differences in resonance amplitude, which are due to the position of the hydrophone relative to the two object orientations. The small frequency shift observed for the cylinder in Fig. 5 is explained by the effect of the water dissolving the chalk during the experiments, thus reducing the mass of the cylinder and causing a frequency shift.

The sharpness of the experimental peaks differs from that predicted by the simulations. The main reason is that damping plays an important part, since an arbitrary value for the chalk material is used in the Ansys simulations ($\xi = 0.005$), while the experimental measurements presented are naturally damped by intergranular friction. In addition, it is impossible to simulate accurately the vibro-acoustic excitation on the object, since it depends on the dimensions of the focal point and drag coefficient.

Results for the sphere correlated particularly well with

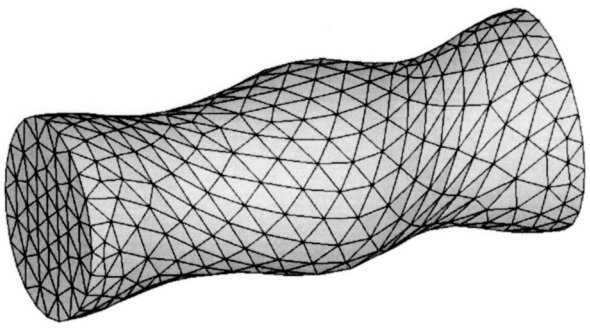


FIG. 10. Cylinder torsional mode.

the simulations because spherical geometry provides isotropic acoustic emission.

Because of the flat response of the water tank without object, the propagation medium has no significant effect on determination of the object's resonance frequencies (Fig. 3).

The temperature was fixed during the experiments; therefore its influence on the acoustic emission signal and the resonance frequencies was ignored.

The model presented here can give good quantitative results for the detection of resonance frequencies. An assumption in this model is that the fluid medium surrounding the sphere and cylinder is homogeneous. Extensive experimental verification of this method is currently under investigation, in order to detect heterogeneities in soft tissues, evaluate their viscoelastic parameters, and obtain *a priori* information on the optimal frequency to be used for imaging.

V. CONCLUSION

Vibro-acoustography provides a means of measuring accurately the resonance frequencies of immersed objects related to compressional and bending modes. A finite element model using harmonic and modal analyses has been developed to investigate the response of materials as a function of frequency. It was concluded that some resonance frequencies can be detected. The simulations correlate well with experimental results. Additional work must focus on using these detectable resonances to improve the performance of vibro-acoustography in biomedical applications, such as characterization of kidney stones during lithotripsy treatment and detection of radioactive metal seeds in brachytherapy sessions.

ACKNOWLEDGMENTS

This work was supported by Grant No. 01-2-93-0314 of an RNTS project from the French Ministry of Economy, Finance and Industry.

APPENDIX

The interaction of fluid and structure at a mesh interface causes the acoustic pressure to exert a force on the structure and the structural motion produces an effective "fluid load." The governing finite element matrix equations then become

$$[M_s]\{\ddot{u}\}+[K_s]\{u\}=\{F_s\}+[R]\{P\}, \tag{A1}$$

$$[M_f]\{\ddot{P}\}+[K_f]\{u\}=\{F_f\}+\varrho_0[R]^T\{\ddot{u}\}. \tag{A2}$$

Here $[R]$ is a coupling matrix representing the effective surface area associated with each node at the fluid–structure interface. The coupling matrix $[R]$ also takes into account the direction of the normal vector defined for each pair of adjacent fluid and structural element surfaces at the interface. The positive direction of the normal vector, as the Ansys program requires, is defined to be inward on the solid structure boundary. Both the structural and fluid load quantities produced at the fluid–structure interface are functions of the chosen nodal degrees of freedom. Placing these "load" quantities on the left-hand side of the equations and combining the two equations into a single equation yields

$$\begin{bmatrix} M_s & 0 \\ \varrho_0 R^T & M_f \end{bmatrix}\begin{Bmatrix} \ddot{u} \\ \ddot{P} \end{Bmatrix}+\begin{bmatrix} K_s & -R \\ 0 & K_f \end{bmatrix}\begin{Bmatrix} u \\ P \end{Bmatrix}=\begin{Bmatrix} F_s \\ F_f \end{Bmatrix}. \tag{A3}$$

Equation (A2) implies that nodes in a fluid–structure have degree of freedom corresponding to both displacement and pressure.

The modal matrix equation associated to Eq. (A3) is

$$\begin{bmatrix} -M_s\omega^2+K_s & -R \\ \varrho_0 R^T\omega^2 & -M_f\omega^2+K_f \end{bmatrix}\begin{Bmatrix} \tilde{u} \\ \tilde{P} \end{Bmatrix}=\begin{Bmatrix} 0 \\ 0 \end{Bmatrix}, \tag{A4}$$

where $\tilde{u}=ue^{j\omega t}$ and $\tilde{P}=Pe^{j\omega t}$. Obviously, the matrix is nonsymmetrical. The eigenvalue problem is of fourth order in ω. Thus, the eigenvalues are complex and the eigenvectors do not satisfy the usual orthogonality conditions.

Anson, L. W., and Chivers, R. C. (**1981**). "Frequency dependence of the acoustic radiation force function (Yp) for spherical targets for a wide range of materials," J. Acoust. Soc. Am. **69**, 1618–1623.

Ansys, Inc. Software.

Callé, S., Remenieras, J. P., Bou Matar, O., and Patat, F. (**2002**). "Presence of nonlinear interference effects as a source of low frequency excitation force in vibro-acoustography," Ultrasonics **40**, 873–878.

Chen, S., Fatemi, M., and Greenleaf, J. F. (**2002**). "Remote measurement of material properties from radiation force induced vibration of an embedded sphere," J. Acoust. Soc. Am. **112**, 884–889.

Fatemi, M., and Greenleaf, J. F. (**1998**). "Ultrasound stimulated vibro-acoustic spectroscopy," Science **280**, 82–85.

Fatemi, M., and Greenleaf, J. F. (**1999**). "Application of radiation force in non contact measurement of the elastic parameters," Ultrason. Imaging **21**, 141–154.

Fatemi, M., Wold, L. E., Alizad, A., and Greenleaf, J. F. (**2002**). "Vibro-acoustic tissue mammography," IEEE Trans. Med. Imaging **21**, 1–8.

Hasegawa, T., and Yosioka, K. (**1969**). "Acoustic-radiation force on a solid elastic sphere," J. Acoust. Soc. Am. **46**, 1139–1143.

Heimbach, D., Munver, R., Zhong, P., Jacobs, J., Hesse, A., Müller, S. C., and Preminger, G. M. (**2000**). "Acoustic and mechanical properties of artificial stones in comparison to natural kidney stones," J. Urol. **164**, 537–544.

Konofagou, E. E., Thierman, J., Karjalainen, T., and Hynynen, K. (**2002**). "The temperature dependence of ultrasound-stimulated acoustic emission," Ultrasound Med. Biol. **28**, 331–338.

Maynard, J. (**1996**). "Resonant ultrasound spectroscopy," Phys. Today **49**, 26–31.

Muir, T. G., and Willette, J. G. (**1972**). "Parametric acoustic transmitting arrays," J. Acoust. Soc. Am. **52**, 1481–1486.

Rice, R. W. (**1977**). "Microstructure dependence of mechanical behaviour," in *Treatise on Materials Science and Technology*, edited by R. K. MacCrone (Academic, New York), Vol. 11, pp. 200–382.

Schroeder, C., and Shao, J. (**1996**). "Plastic deformation and capillary effects in chalks," Fifth North Sea Chalk Symposium

Torr, G. R. (**1984**). "The acoustic radiation force," Am. J. Phys. **52**, 402–408.

Visscher, W. M., Migliori, A., Bell, T. M., and Reinert, R. A. (**1991**). "On the normal modes of free vibration of inhomogeneous and anisotropic elastic objects," J. Acoust. Soc. Am. **90**, 2154–2162.

Westervelt, P. J. (**1951**). "The theory of steady force caused by sound waves," J. Acoust. Soc. Am. **23**, 312–315.

The selection of layer thicknesses to control acoustic radiation force profiles in layered resonators

Martyn Hill[a)]
School of Engineering Sciences, Mechanical Engineering, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

(Received 22 November 2002; revised 2 August 2003; accepted 11 August 2003)

Ultrasonic standing waves can be used to generate radiation forces on particles within a fluid. A number of authors have derived detailed representations of these forces but these are most commonly applied using an approximation to the energy distribution based upon an idealized standing wave within a mode based upon rigid boundaries. An electro-acoustic model of the acoustic energy distribution within a standing wave with arbitrary thickness boundaries has been expanded to model the radiation force on an example particle within the acoustic field. This is used to examine the force profile on a particle at resonances other than those predicted with rigid boundaries, and with pressure nodes at different positions. A simple analytical method for predicting modal conditions for combinations of frequencies and layer thickness characteristics is presented, which predicts that resonances can exist that will produce a pressure node at arbitrary positions in the fluid layer of such a system. This can be used to design resonators that will drive particles to positions other than the center of the fluid layer, including the fluid/solid boundary of the layer, with significant potential applications in sensing systems. Further, the model also predicts conditions for multiple subwavelength resonances within the fluid layer of a single resonator, each resonance having different nodal planes for particle concentration.
[DOI: 10.1121/1.1616581]

PACS numbers: 43.25.Qp, 43.20.Ks [MFH] Pages: 2654–2661

I. INTRODUCTION

Particles within ultrasonic standing waves (USWs) experience small acoustic radiation forces that tend to drive them towards nodal or antinodal planes within that standing wave.[1,2] This can be used to agglomerate particles, or to concentrate or separate particles for analytical purposes.[3–5] Microfluidic devices based on this principle have the potential to act as the microfluidic equivalent of a centrifuge or to filter particles from samples prior to analysis.

Using these forces requires the design and construction of ultrasonic chambers in which the characteristics of the standing waves can be predicted and controlled. Typically, chambers are similar to that shown schematically in Fig. 1 and are designed such that a standing wave in a single planar dimension is predominant (although more complex geometries can be used such as cylindrical resonators[6] or crossed standing waves which move particles to nodal lines rather than nodal planes[7]).

In addition to the radiation forces due to the standing wave in the principal dimension, lateral forces will also act on particles due to variations in the primary acoustic field and interparticle interactions such as Bjerkness forces. This paper considers the dominant, primary radiation forces only.

Several models of the acoustic radiation forces within standing waves have been developed[8–10] and these have typically been used to provide analytical solutions for the force profiles within idealized standing waves. The electro-acoustic characteristics of more general planar standing waves have been modeled successfully as layered resonators.[1,11–14] This paper uses an acoustic force model, combined with an electro-acoustic model described in Refs. 13 and 14 to examine the effects of deviation from a half-wavelength standing wave pattern (due to reactive impedances at the resonator boundaries) upon the performance of a practical standing wave chamber. In particular, it examines how the thickness of the reflector and fluid layers can be chosen to determine the position of the pressure nodal plane within the fluid layer.

II. BACKGROUND

Gor'kov[15] derives the time-averaged acoustic radiation force on a spherical particle of radius a, at position x within a one-dimensional standing wave as

$$F(x)=\frac{\partial}{\partial x}\left(\frac{4\pi a^3}{3}\left(\frac{3(\rho_p-\rho_f)}{(2\rho_p+\rho_f)}E_{\mathrm{kin}}(x)-\left(1-\frac{\rho_f c_f^2}{\rho_p c_p^2}\right)E_{\mathrm{pot}}(x)\right)\right), \tag{1}$$

where c and ρ are the speed of sound and the mass density of the fluid and the particle, indicated by subscripts f and p, respectively. $E_{\mathrm{kin}}(x)$ and $E_{\mathrm{pot}}(x)$ are the time-averaged kinetic and potential energy densities at position x within the fluid.

The energy terms in this equation are typically substituted by considering a standing wave excited within a fluid of low specific acoustic impedance bounded by two reflectors with a high and purely real acoustic impedance to the

a)Electronic mail: m.hill@soton.ac.uk

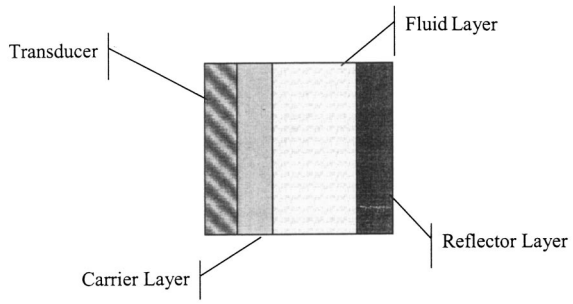


FIG. 1. Schematic diagram of layered resonator.

plane standing wave. In such cases, half-wavelength acoustic modes exist with a sinusoidal acoustic velocity profile and a consinusoidal acoustic pressure profile, and hence

$$E_{\mathrm{kin}}(x)=\varepsilon \sin^2(kx),$$
$$E_{\mathrm{pot}}(x)=\varepsilon \cos^2(kx), \tag{2}$$

where ε is the overall acoustic energy density and k is the wavenumber ($2\pi/\lambda$ where λ is the wavelength). This leads to the following commonly quoted force distribution:

$$F(x)=4\pi k\varepsilon a^3\left(\frac{\rho_p+\frac{2}{3}(\rho_p-\rho_f)}{2\rho_p+\rho_f}-\frac{c_f^2\rho_f}{3c_p^2\rho_p}\right)\sin(2kx). \tag{3}$$

However, in real resonators the purely real boundary conditions leading to this distribution do not hold under all circumstances, due to the finite extent of the reflecting boundaries. This leads to phase shifts at the boundaries and acoustic modes occur at frequencies that no longer correspond to a half-wavelength equaling the cavity depth, as is discussed in the following section.

III. THE EFFECT OF A FINITE THICKNESS REFLECTOR

Consider the reflector layer shown in Fig. 1. If it were of infinite extent, then a plane progressive wave incident on the boundary would reflect with a pressure reflection coefficient R given by[16]

$$R=\frac{r_r-r_f}{r_r+r_f}, \tag{4}$$

where $r_r(=\rho_r c_r)$ is the characteristic impedance of the reflector and r_f is the characteristic impedance of the fluid. The reflection coefficient R is real and positive for acoustic impedances higher than the fluid (''pressure doubling'' boundaries for much higher impedances) or real and negative for impedances lower than the fluid (''pressure release'' boundaries for much lower impedances).

However, if we now consider a finite reflector of thickness t_r, in steady-state operation and terminated by an impedance r_0, then the pressure reflection coefficient is[16]

$$R=\frac{r_r(r_0-r_f)+j(r_r^2-r_f r_0)\tan k_r t_r}{r_r(r_0+r_f)+j(r_r^2+r_f r_0)\tan k_r t_r}. \tag{5}$$

Such a reflection coefficient has characteristics such as those shown in Fig. 2 for an air-backed Pyrex reflector with a water fluid layer. It can be seen from this that close to $t_r/\lambda_r=0.25,0.75,1.25,...$ (i.e., well away from half-wave thickness resonances of the reflector) the reflector behaves like a ''pressure doubling'' boundary with $R=1$ and zero phase shift. However, near the thickness modes ($t_r/\lambda_r=0.5,1,1.5,...$), while the reflection coefficient magnitude remains close to unity, the phase change approaches pi radians representing a pressure release boundary. Under these conditions, there is no acoustic mode in the fluid at the frequency that corresponds to a wavelength of twice the fluid depth. In fact, when $t_r/\lambda_r=0.5$ the acoustic impedance seen looking into the reflector is that of air, as if the reflector layer does not exist.

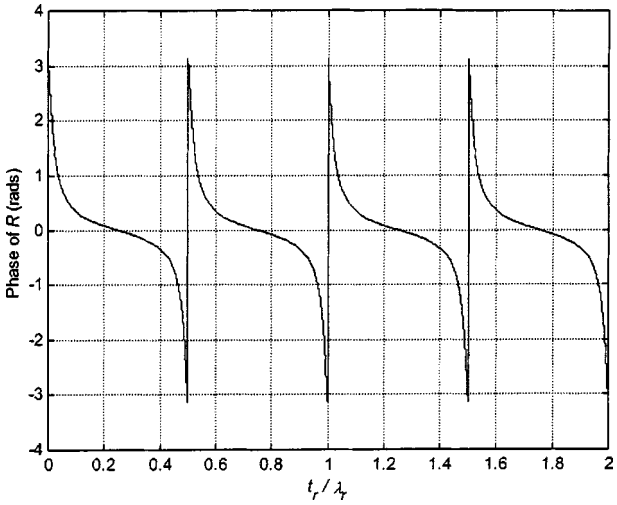


FIG. 2. Phase of the reflection coefficients for an air-backed Pyrex reflector of varying thickness.

In order to examine the effects of the reflector thickness on resonator behavior, the model transfer impedance model described by Hill *et al.*[13,14] has been used, but in order to isolate the effects of the fluid and reflector layers from those of the carrier layer and transducer, only the fluid and reflector layers are considered. To generate the results that follow, a harmonic acoustic velocity of amplitude 2×10^{-3} ms^{-1} has been imposed at the left-hand boundary of the fluid layer. Knowledge of impedances at the boundaries allows the acoustic pressure and velocity distribution throughout the fluid layer to be calculated, which is also used to derive kinetic and potential energy distributions through the layer. Such an approach is useful to illustrate the effect of coupling between isolated layers, but to design a full resonator system, a model that includes all the layers must be used. Nonetheless the author has found the approach outlined here useful to determine potential resonator performance prior to the final

TABLE I. Default values for resonator layer parameters.

Layer	Material	Density ρ (kgm^{-3})	Speed of sound c (ms^{-1})	Thickness t
Fluid	water	1000	1500	variable
Reflector	Pyrex	2200	5430	variable
Reflector termination	air	1.3	330	∞

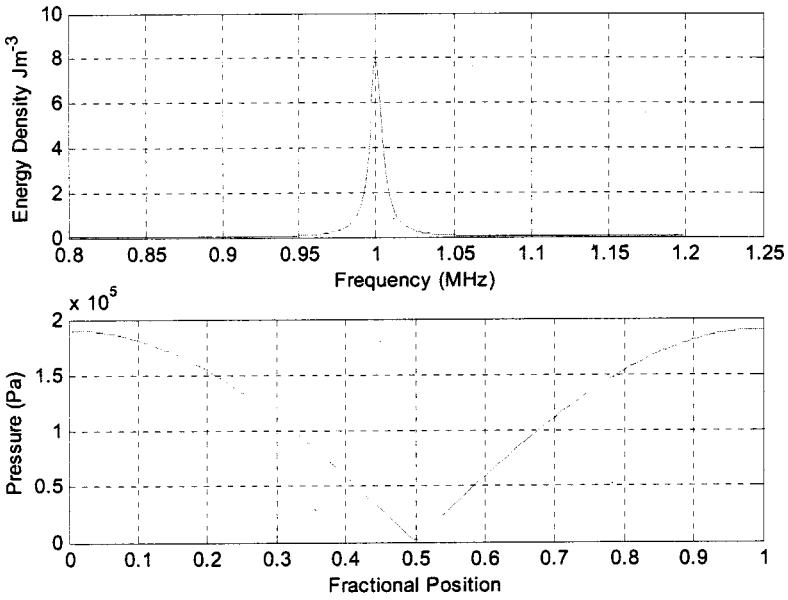


FIG. 3. Energy density against frequency and acoustic pressure profile for a half-wavelength cell.

design process. In the following section, the two-layer model is used to simulate a situation that approximates closely to the idealized half-wavelength standing wave with two perfectly rigid boundaries.

A. Half-wave chamber with quarter wave reflector

Referring to Fig. 2, if the first point of zero phase is chosen, $t_r/\lambda_r = 0.25$, then the expected half-wave resonance should be produced in the fluid.

This situation has been simulated using the reflector data shown in Table I. The chamber is designed to resonate at 1 MHz, leading to values of $t_r = 1.36$ mm and $t_f = 0.75$ mm. Figure 3 shows the variation of energy density held in the standing wave in the liquid for such a chamber.

Particles held within the standing wave will experience a force toward the central pressure node as described by Eq. (3). As the simulation model includes full derivations of the kinetic and potential energy densities across the chamber, the full acoustic force model of Eq. (1) can be implemented and is shown in Fig. 4 using data for a 2-μm diameter polystyrene sphere, as described by Tolt and Feke[17] who evaluated the force due to Eq. (3) on such a sphere. In these simulations buoyancy/gravitational forces are ignored, but for particles that are not neutrally buoyant in horizontal resonators the particles will move to a plane away from the pressure node.

To the left of the chamber the force is positive, tending to move the particle toward the central pressure node. To the right of the chamber the force is negative, again moving the particle toward the central pressure node. It can be seen that force toward the node decreases as the node is approached, but in addition it should be noted that the acoustic force on particles near the boundaries themselves are also extremely low.

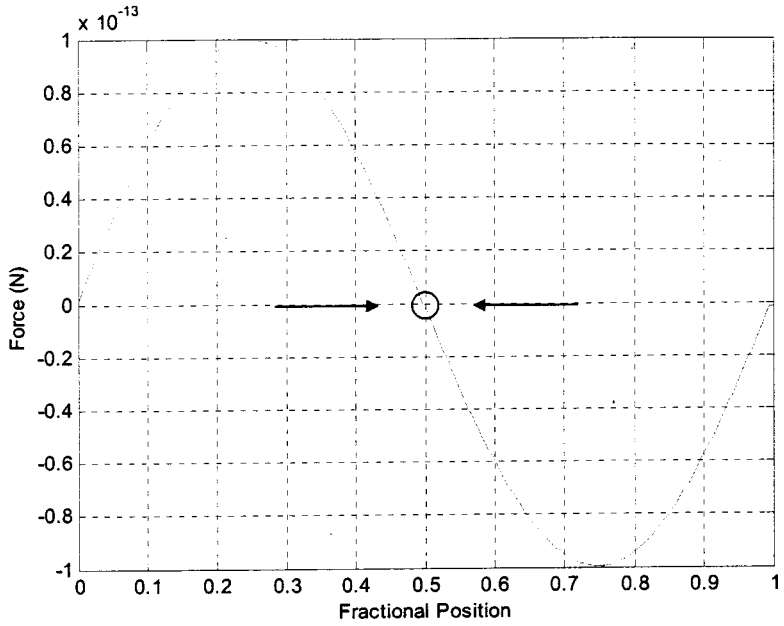


FIG. 4. Profile of acoustic radiation force on example particle for a half-wavelength cell. Arrows mark the direction of travel of the particles and the circle represents the position to which particles are driven by the acoustic force, in the absence of buoyancy forces.

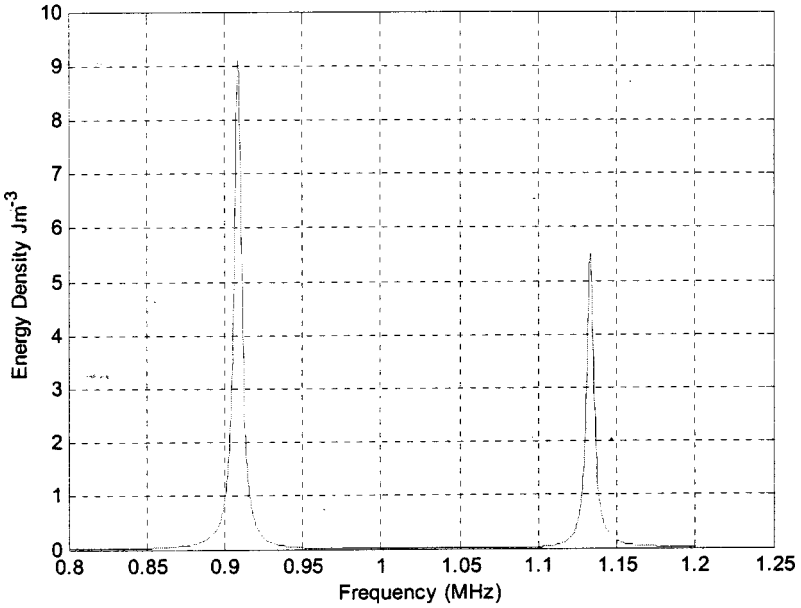


FIG. 5. Energy density against frequency for near half-wave reflector and half-wave chamber at nominal 1 MHz.

B. Half-wave chamber with near half-wave reflector

The model can be used to investigate the forces on a trapped particle when the reflector offers a complex reflection coefficient, and a value of $t_r/\lambda_r = 0.48$ has been chosen, which (referring to Fig. 2) provides a reflection coefficient with a phase of about $-\pi/2$.

The choice of dimensions now means that both the fluid and reflector layers are close to a half-wave resonances and, instead of a single resonance at 1 MHz, the coupled resonances lead to a double peak straddling the nominal resonance frequencies of the layers, as demonstrated experimentally and discussed by Hawkes *et al.*[12,14] This phenomenon occurs in simulations for the current conditions and can be seen in the energy density frequency plot of Fig. 5, where maximum energy is now stored in the standing waves at frequencies of about 0.91 and 1.13 MHz.

The acoustic pressure profile for each of these energy peaks is shown in Fig. 6 and the distribution of forces on an example particle is shown in Fig. 7.

The pressure nodes shown in Fig. 6 are no longer central, but perhaps of more significance in a practical ''half-wavelength'' system, at the upper working frequency there is now a pressure antinode to the left of the reflector boundary.

The profiles of the acoustic radiation force on a particle, Fig. 7, clearly show the effect of this. At the lower frequency [0.91 MHz, graph (a)], the particles will all tend to move toward the pressure node at a fractional position of about 0.55. However, at the higher frequency resonance [1.13 MHz, graph (b)], only particles starting at a fractional position between 0 and about 0.9 will move toward the node (at a position of about 0.45). Particles at a position of 0.9 or above will experience a positive force, tending to move them

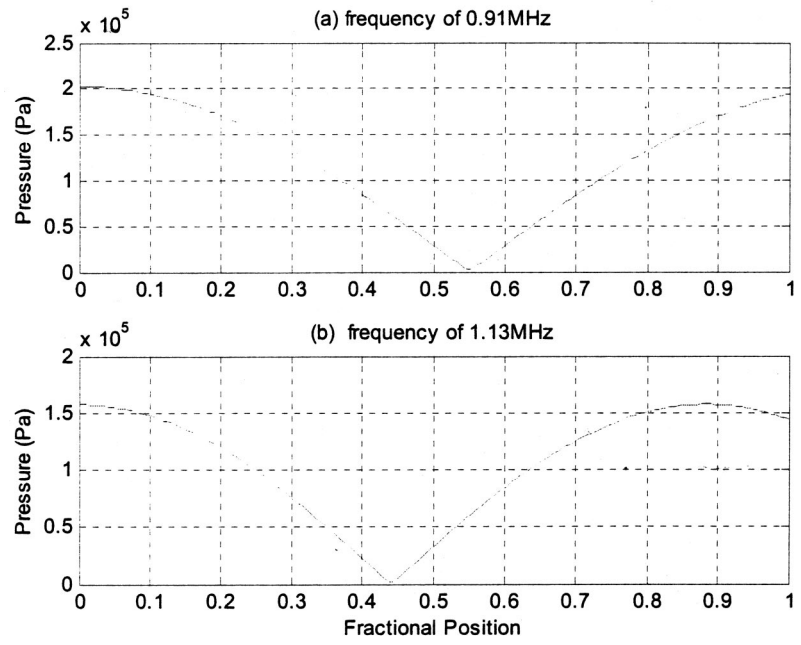


FIG. 6. Acoustic pressure profiles at each of the energy peaks for near half-wave reflector and half-wave chamber at nominal 1 MHz.

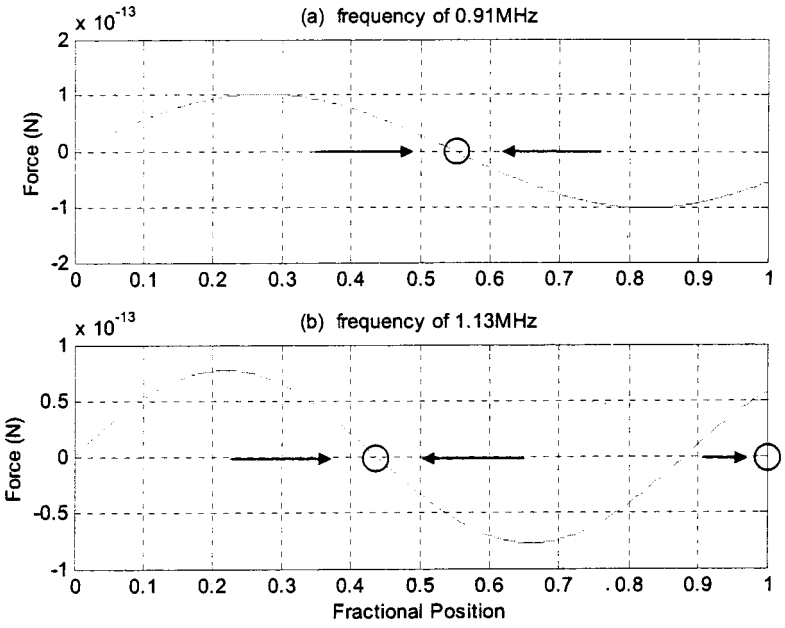


FIG. 7. Profile of acoustic radiation force on example particle at each of the energy peaks for near half-wave reflector and half-wave chamber at nominal 1 MHz.

towards the reflector boundary itself. While the phenomenon of trapping particles against the reflector itself may be undesirable in applications of particle manipulation where the design intention has been to gather particles in the center of the layer, the ability to force particles up against a boundary is potentially significant.

IV. A METHOD FOR SELECTING LAYER THICKNESSES FOR PREDETERMINED PLACEMENT OF PRESSURE NODES

The dimensions in the above example were based upon a $t_r/\lambda_r = 0.48$ reflector at a nominal 1 MHz. However, with a "half-wavelength" cell, no resonance was observed at 1 MHz. This section describes a method for selecting the dimensions of a reflector layer to ensure that an acoustic mode exists at a frequency with a predetermined wavelength (and hence nodal position) within a given fluid layer. It also enables rapid prediction of the modes and associated nodal positions existing in a predetermined fluid/reflector geometry.

The acoustic impedance looking from the transducer into the fluid layer, z_{f_in} will be

$$z_{f\_in} = \left(\frac{z_{tf} + j r_f \tan k_f t_f}{r_f + j z_{tf} \tan k_f t_f} \right) r_f, \tag{6}$$

where z_{tf} is the impedance at the boundary with the reflector. If we assume a very high impedance at the transducer boundary, the fluid layer will resonate when

$$z_{tf} = -\frac{r_f}{j \tan k_f t_f}. \tag{7}$$

The impedance looking into the reflector will be

$$z_{r\_in} = \left(\frac{z_0 + j r_r \tan k_r t_r}{r_r + j r_0 \tan k_r t_r} \right) r_r. \tag{8}$$

The combined system will resonate when the reactances at the reflector boundary are equal:

$$\mathrm{Im}\{z_{r\_in} - z_{tf}\} = 0, \tag{9}$$

which holds when

$$k_f t_f = \tan^{-1}\left(\frac{r_f (r_r^2 + r_0^2 \tan^2(k_r t_r))}{r_r \tan(k_r t_r)(r_r^2 - r_0^2)} \right). \tag{10}$$

The first four solutions to this equation are shown in Fig. 8. Also shown on the figure is the point which coincides with the quarter-wave reflector and half-wave chamber of Fig. 3 and the two points coinciding with the peaks of the "near-half-wavelength" reflector of Fig. 5.

A given design of resonator can be represented in Fig. 8 by a straight line of gradient $t_f c_r / t_r c_f$. For the "near-half-wave" example above (Figs. 5–7) $t_f c_r / t_r c_f = 1.042$ and this resonator is represented by the solid line in Fig. 9.

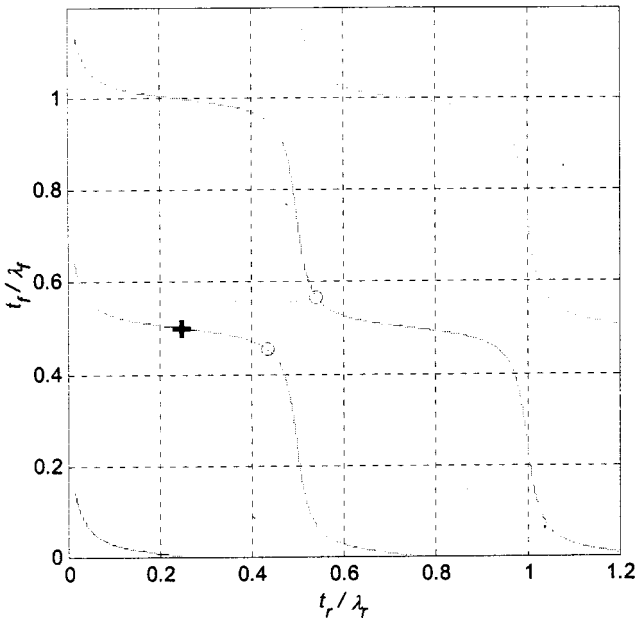


FIG. 8. The first four solutions to Eq. (10). Resonances of Fig. 3 ("+") and Fig. 5 ("○") are also marked.

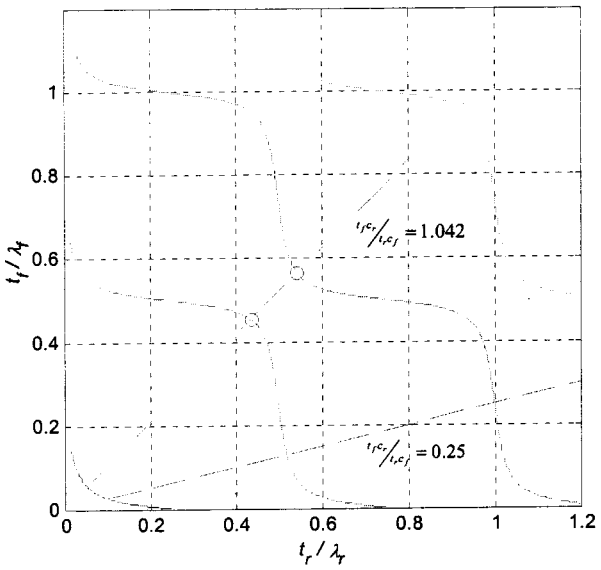


FIG. 9. Solutions to Eq. (10) with representations of ''near-half-wave'' and ''quarter-wave'' resonators. Resonances of Fig. 5 are marked as circles.

This line increases with frequency and the points at which it crosses the solutions to Eq. (10), marked with circles, correspond to $t_r/\lambda_r = 0.43$ and 0.54, i.e., $f = 0.91$ and 1.13 MHz (the peaks in Fig. 5). Figure 9 also predicts modes at $t_r/\lambda_r = 0.053$ and 0.92, i.e., $f = 0.11$ and 1.92 MHz, which can be seen in Fig. 10 in which the simulation of Fig. 5 is now run over an extended frequency range.

Figures 8 and 9 suggest that resonances can exist within the chamber for any value of t_f/λ_f, so long as an appropriate value of t_r/λ_r is chosen. This, in theory, should allow a resonator to be constructed with the pressure node placed anywhere within the resonator. In practice, precise positioning will be complicated by the finite velocity at the transducer, or carrier layer wall and variability in the controlling parameters.

Figure 11 shows the results of simulations carried out in the design of an ultrasonic microfluidic separator.[18,19] In this case a full simulation, including a transducer model, an adhesive layer, and a carrier layer, was carried out. The separator uses ultrasonic radiation forces to filter particles from fluid flowing through a channel etched out of Pyrex, which is then bonded to a silicon substrate. The aim of the simulation was to choose an appropriate etch depth for the channel to provide a high acoustic energy density within the fluid. The thickness of the chosen Pyrex wafer was 1.7 mm and hence as the fluid thickness increases in the simulations, the thickness of the Pyrex reflector layer decreases, with the sum of the thicknesses remaining at 1.7 mm. On the basis of the simulations, an etch depth of about 240 μm was selected to produce a mode of near a half-wavelength in the fluid layer with a central pressure node at a frequency of about 3.1 MHz.

Figure 11(c) also shows that over this range of fluid depths (and associated reflector thicknesses), the pressure node can be positioned anywhere between the center of the channel and the reflector boundary. As the fluid depth decreases from 250 to 110 μm the pressure node within the fluid layer approaches the reflector boundary. At $t_f = 110\ \mu$m the pressure node is positioned at the fluid/reflector boundary itself. For this depth, the ratio $t_f c_r / t_r c_f = 0.25$ and for the corresponding modal frequency [about 3.4 MHz from Fig. 11(b)] the fluid layer is a quarter wavelength in thickness and the reflector layer is a whole wavelength thick. A resonator with the ratio $t_f c_r / t_r c_f = 0.25$ is represented by the dashed line in Fig. 9 and the mode under discussion corresponds to the point at which this intersects the third solution of Eq. (10) at point $(t_r/\lambda_r = 0.25, t_f/\lambda_f = 1)$. Referring again to Fig. 11(c): at fluid depths below 110 μm the pressure minimum in the fluid remains at the reflector boundary although this is no longer a pressure node in this range as the actual node of the standing wave is within the reflector itself.

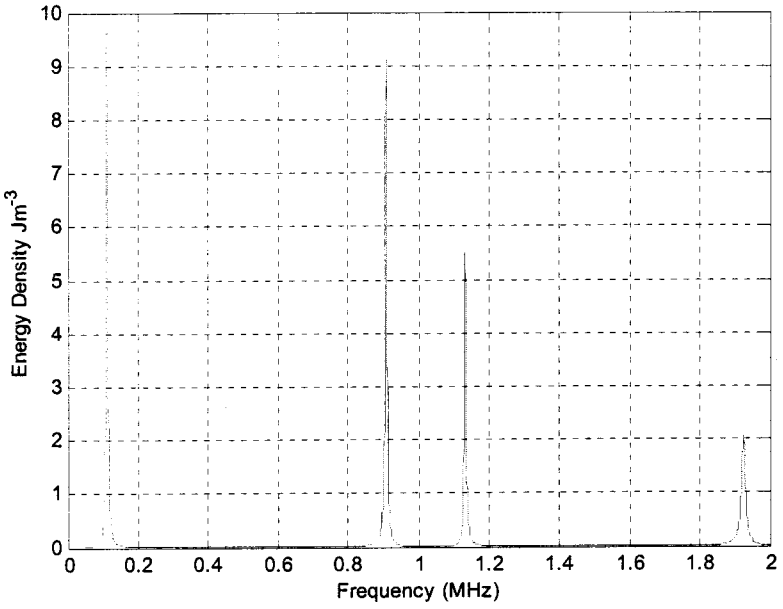


FIG. 10. Energy density against frequency for near half-wave reflector and half-wave chamber at nominal 1 MHz over an extended frequency range.

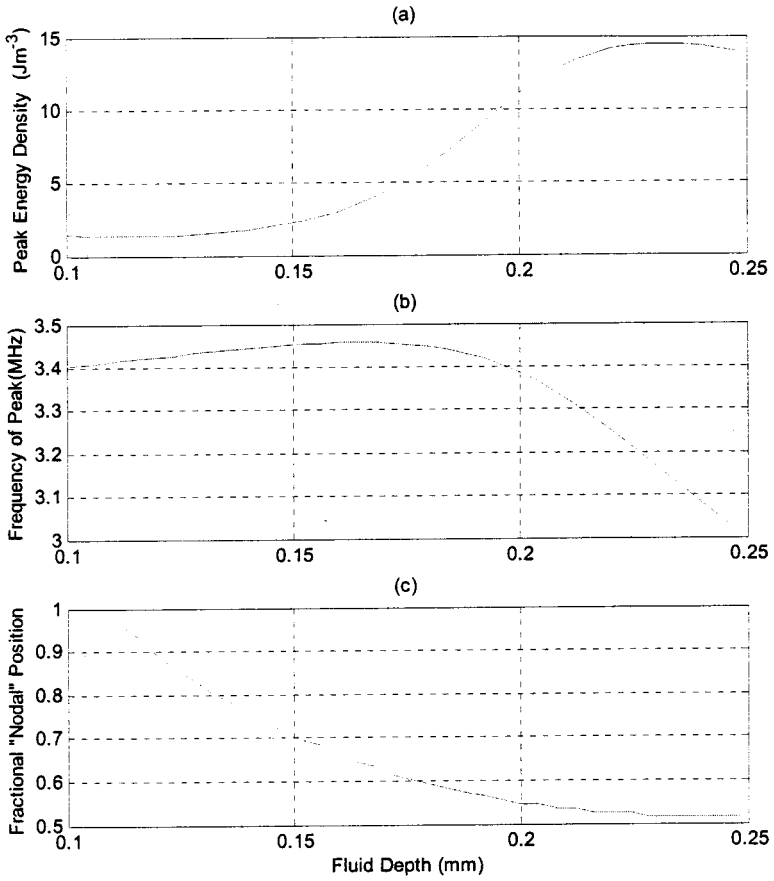


FIG. 11. Simulations carried out in the design of an ultrasonic microfluidic separator, showing (a) the change in peak energy density, (b) the frequency, and (c) the position of the pressure minimum of a node as the fluid depth changes. As the fluid depth increases, the reflector thickness decreases such that the total of the thicknesses remains at 1.7 mm.

A. Quarter-wave chamber with half-wave reflector

The phenomenon of a pressure node within a liquid at a solid boundary will occur for any multiple of half-wavelength reflector when the fluid layer is at a quarter-wavelength thickness. Referring to Fig. 8, it can be seen that the first solution to Eq. (10) with $t_r/\lambda_r = 0.5$ corresponds to a quarter-wavelength chamber. The simulated energy density spectrum, along with the acoustic pressure and particle force profiles for a quarter-wave fluid layer and a half-wave reflector are shown in Fig. 12.

In this case, the pressure node occurs at the reflector boundary and from the profile of the acoustic force on the particle remains positive across the chamber. Hence, particles within the chamber will experience a force toward that boundary. This phenomenon has been observed experimentally by Hawkes *et al.*[20] and, as the authors observe, the ability to drive particles toward the solid boundary of a fluid layer has important implications with regard to sensing applications where detection of particles at the boundary of the fluid layer is required.

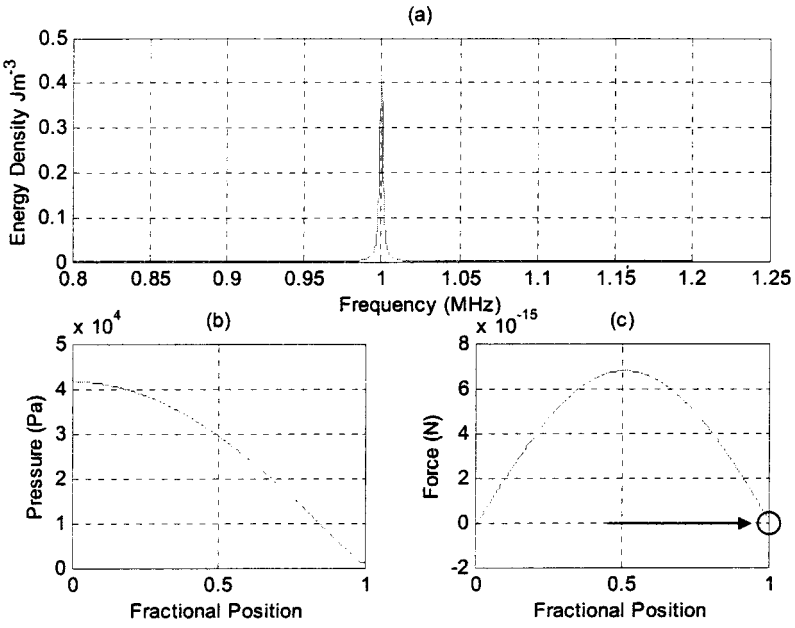


FIG. 12. (a) Frequency response, (b) pressure profile, and (c) profile of acoustic force on 2-μm-diam polystyrene particle, for a cell with a half-wavelength reflector and a quarter-wavelength fluid layer at 1 MHz.

V. CONCLUSIONS

An electro-acoustic model of a piezoelectrically driven layered resonator, which has previously been shown to predict the characteristics of experimental systems, has been expanded to model the primary acoustic radiation force on example particles. These force profiles have been shown to match widely used radiation force approximations for certain, idealized, standing wave patterns when the reflector layer may be regarded as rigid. However, under many circumstances, the radiation forces do not match this pattern, and if the resonators are driven at modes of the coupled fluid and reflector layer, the pressure nodes do not occur at the center of a nominally half-wavelength system.

An analytical equation which predicts the combination of layer thicknesses required to produce an acoustic mode with a correctly positioned pressure node has been derived and shown to have solutions which correspond to previously simulated resonators. The equation predicts that the correct combination of thicknesses can be used to place a node at any position within the fluid layer, including the fluid/reflector boundary. This is supported by simulation of a full resonator system and ties in with recent experimental observations from other authors.

ACKNOWLEDGMENTS

The author would like to acknowledge financial support from the Engineering and Physical Sciences Research Council, Dstl and the Porvair Filtration Group Ltd.

[1] M. Gröschl, ''Ultrasonic separation of suspended particles—Part I: Fundamentals,'' Acustica **84**, 432–447 (1998).
[2] W. T. Coakley, ''Ultrasonic separations in analytical biotechnology,'' Trends Biotechnol. **15**, 506–511 (1997).
[3] W. T. Coakley, J. J. Hawkes, M. A. Sobanski, C. M. Cousins, and J. Spengler, ''Analytical scale ultrasonic standing wave manipulation of cells and microparticles,'' Ultrasonics **38**, 638–641 (2000).
[4] J. J. Hawkes, M. S. Limaye, and W. T. Coakley, ''Filtration of bacteria and yeast by ultrasound-enhanced sedimentation,'' J. Appl. Microbiol. **82**, 39–47 (1997).
[5] J. J. Hawkes and W. T. Coakley, ''Force field particle filter, combining ultrasound standing waves and laminar flow,'' Sens. Actuators B **75**, 213–222 (2001).
[6] C. M. Cousins, P. Holownia, J. J. Hawkes, M. S. Limaye, C. P. Price, P. J. Keay, and W. T. Coakley, ''Plasma preparation from whole blood using ultrasound,'' Ultrasound Med. Biol. **26**, 881–888 (2000).
[7] M. Hill and E. Newsom, ''Ultrasonic standing waves for concentrating particles within a low flow-rate fluid,'' in Proceedings: USITT Colloquium Developments in Transducer Technology, Southampton, UK, 2000.
[8] L. V. King, ''On the acoustic radiation pressure on spheres,'' Proc. R. Soc. London, Ser. A **147**, 212–240 (1934).
[9] K. Yosioka and Y. Kawasima, ''Acoustic radiation pressure on a compressible sphere,'' Acustica **5**, 167–173 (1955).
[10] S. D. Danilov and M. A. Mironov, ''Radiation pressure force acting on a small particle in a sound field,'' Sov. Phys. Acoust. **30**, 280–283 (1984).
[11] H. Nowotny and E. Benes, ''General one-dimensional treatment of the layered piezoelectric resonator with two electrodes,'' J. Acoust. Soc. Am. **82**, 513–521 (1987).
[12] J. J. Hawkes, M. Gröschl, H. Nowotny, S. Armstrong, P. Tasker, W. T. Coakley, and E. Benes, ''Single half wavelength ultrasonic particle filter: Predictions of the transfer matrix multi-layer resonator model and experimental filtration results,'' J. Acoust. Soc. Am. **111**, 1259–1266 (2002).
[13] M. Hill and R. J. K. Wood, ''Modelling in the design of a flow-through ultrasonic separator,'' Ultrasonics **38**, 662–665 (2000).
[14] M. Hill, Y. Shen, and J. J. Hawkes, ''Modelling of layered resonators for ultrasonic separation,'' Ultrasonics **40**, 385–392 (2002).
[15] L. P. Gor'kov, ''On the forces acting on a small particle in an acoustical field in an ideal fluid,'' Sov. Phys. Dokl. **6**, 773–775 (1962).
[16] L. E. Kinsler, A. E. Frey, A. B. Coppens, and J. V. Saunders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982).
[17] T. L. Tolt and D. L. Feke, ''Separation of dispersed phases from liquids in acoustically driven chambers,'' Chem. Eng. Sci. **48**, 527–540 (1992).
[18] M. Hill, N. R. Harris, R. Townsend, S. P. Beeby, Y. Shen, N. M. White, and J. J. Hawkes, ''A Microfluidic Device For Ultrasonic Separation,'' in *Proceedings: Forum Acusticum*, 3rd EAA European Congress on Acoustics (Sociedad Espanola De Acustica, Sevilla, 2002), PHA-01-003-IP.
[19] N. R. Harris, M. Hill, S. P. Beeby, Y. Shen, N. M. White, J. J. Hawkes, and W. T. Coakley, ''A Silicon Microfluidic Ultrasonic Separator,'' Sens. Actuators B **95**, 425–434 (2003).
[20] J. J. Hawkes, M. Gröschl, E. Benes, H. Nowotny, and W. T. Coakley, ''Positioning particles within liquids using ultrasound force fields,'' in Proceedings 3rd EAA European Congress on Acoustics, Sevilla, 2002.

The ultrasonic weak short-pulse responses of microbubbles based on a two-frequency approximation

Chung-Yuo Wu and Jenho Tsao[a)]
Graduate Institute of Communication Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, Taiwan 10617, Republic of China

(Received 13 January 2003; revised 22 August 2003; accepted 8 September 2003)

The ultrasonic short-pulse responses of microbubbles are of interest in cavitation, transient responses, and contrast imaging. We extend the two-frequency analytic solutions of Newhouse and Shankar [J. Acoust. Soc. Am. **75**, 1473–1477 (1984)] to approximate the short-pulse responses of microbubbles in a low-amplitude field. Based on their results, there is an expected component near dc in the spectrum of bubble echoes excited by a short pulse. Here this component is named the low-frequency response, and its theoretical properties are verified experimentally. Including the fundamental and second-harmonic components, the weak short-pulse responses of microbubbles include three types of response. Our work has determined the constraint conditions under which this approximated solution can be used to analyze these short-pulse responses. This paper also provides the amplitude and spectral properties of these responses. The low-frequency response has a special bandwidth-dependent property and has potential applications in imaging and bubble sizing. [DOI: 10.1121/1.1621861]

PACS numbers: 43.25.Yw, 43.25.Zx, 43.30.Dr [AJS] Pages: 2662–2671

I. INTRODUCTION

The short-pulse responses of microbubbles to ultrasound excitation are of interest in the study of cavitation, transient responses,[1–10] and contrast imaging.[11,12] However, complete solutions are difficult to obtain due to the nonlinearity of bubble responses to such excitation. Different analytic solutions and measurement techniques have been developed for different types of bubble responses.[13–21] The nonlinear behavior is greatly influenced by the pulse length and pressure of the excitation signal. Well-developed analytic and numerical solutions are available for the low-amplitude and long-pulse case, and some transient solutions can be used to determine the threshold at which nonlinear effects occur for high-amplitude and long-pulse stimuli.[1] Some numerical results have appeared for the low-amplitude and short-pulse case.[2,6,22]

Detection techniques based on the resonance properties of microbubbles have been developed for contrast imaging to filter out the responses of intervening tissue. In a low-amplitude field, only resonant bubbles will pulsate appreciably,[13–15] and this reduces the influence of nonlinear propagation and the associated system harmonics that are generated. An attractive alternative method is to detect the subharmonic response as the pulse amplitude increases.[16,17] All the above-mentioned methods require long stimulus pulses, which reduces the resolution of the resultant images.

The increased use of ultrasound contrast imaging makes a thorough assessment of the possible biological effects of cavitation desirable.[5] Flynn,[2] Apfel,[3] and ter Haar *et al.*[4] predicted that short pulses can cause transient cavitation, in which stabilized gas pockets or microbubbles may grow rapidly and then collapse violently within a few acoustic cycles. Akulichev *et al.*[7] noted a pulse-length dependence of the transient response of microbubbles in the near-surface oceanic medium. For the frequency range (5–35 kHz) used in the experiments of Akulichev *et al.*, a measurable transient response occurred when the pulse length was less than 5–10 cycles. However, Suiter[8] found no transient responses when using a 2.6-cycle burst at 120 kHz. Pace *et al.*[9] stimulated over the frequency range 20–200 kHz with pulse lengths of 1–20 cycles, and also found no transient responses; they concluded that transient responses do exist when the pulse length is short enough, and they are observable in the absence of masking produced by the bubble distribution. The accompanying increased bandwidth when using a short pulse increases the complexity of the frequency response of the microbubbles. Morgan *et al.*[10] found that the echo intensity is a function of the physical properties of the contrast agent and the properties of the transmitted waveforms, such as signal intensity, center frequency, and transmitted pulse length.

The theoretical behavior of microbubbles exposed to an external pressure field can be described by the RPNNP equation,[23] and it comprises transient and steady-state components. Solving this equation analytically would improve our understanding of the complicated acoustic behaviors of microbubbles. Prosperetti[1] presented an analytical analysis of the transient oscillations in subharmonic responses. Miller[14] described the analytic solutions for low-amplitude and long-pulse wave fields, which were expanded by de Jong *et al.*[24] to an ideal gas bubble including the shell elasticity and friction of an Albunex® microsphere. Newhouse and Shankar[19] developed a two-frequency excitation technique for bubble size estimation, and gave approximate analytic solutions to the RPNNP equation. Due to their nonlinear behavior, when bubbles are excited by two frequencies simultaneously (f_1

a)Electronic mail: tsaor215@cc.ee.ntu.edu.tw

and f_2, where $f_1 \gg f_2$) the resulting echo signals contain sum and difference frequencies, $f_1 \pm f_2$.[18–21]

Under the steady-state conditions of the RPNNP equation, if we define the long-pulse response as a single-frequency response, then the short-pulse response can be defined as a multifrequency response with a continuous spectrum. In this paper, we extend the results of Newhouse and Shankar[19] to study the simplest case of a response resulting from stimulation with a short pulse comprising two frequencies only. For this purpose, the two frequencies are rewritten as $f_1 = f_c + \Delta f/2$ and $f_2 = f_c - \Delta f/2$, where f_c is the center frequency. If these two frequencies are not too far apart to maintain a pulse length of $T \approx 1/\Delta f$, then Δf can be treated as the effective bandwidth of the waveform. Based on the analytic solutions of Newhouse and Shankar, there is an expected component near dc in the spectrum of the bubble echo; this component is termed the low-frequency response here. Under the restriction of small external pressures, these analytic solutions can serve as a first step to characterizing the echoes of microbubbles excited by short-pulse wave forms.

The theoretical derivations of Newhouse and Shankar[19] are summarized in Sec. II. Under small external pressures, additional constraint conditions for using these analytic solutions to analyze the short-pulse responses are derived in Sec. III. The weak short-pulse responses comprise fundamental, second-harmonic, and low-frequency components. The amplitude and spectral properties of these components are also derived in Sec. III. The low-frequency response has a special bandwidth-dependent (i.e., pulse-length-dependent) property that is demonstrated by theoretical and numerical analyses in Sec. IV. These properties are also verified experimentally in Sec. V. The differences between the low-frequency and subharmonic responses are also discussed in this paper.

II. TWO-FREQUENCY APPROXIMATION

Microbubbles in fluid media exhibit several modes of vibration, among which the spherically symmetrical volume mode is the most important. The radius variation is sufficient for characterizing their nonlinear responses. The differential equation for a bubble in volume pulsation can be written as

$$\rho R\ddot{R} + \frac{3}{2}\rho\dot{R}^2 = \left(p_0 + \frac{2\sigma}{R_0}\right)\left(\frac{R_0}{R}\right)^{3\gamma} - \left(p_0 + \frac{2\sigma}{R}\right) - 4\mu\frac{\dot{R}}{R} + p_{\text{ext}}(t), \tag{1}$$

where R_0 and R are the initial and instantaneous radius of the bubble, ρ and μ are the density and viscosity of the liquid, σ is the surface tension, γ is the polytropic exponent, and p_0 and p_{ext} are the ambient and external pressure amplitudes, respectively. This equation neglects some types of damping (e.g., energy dissipation through acoustic radiation, heating of the interior gas and the resulting thermal conduction between the gas and liquid) except for the loss through viscosity at the bubble wall.[21] The methodology utilized is similar to that described in Newhouse and Shankar.[19] When the external pressure is low and resonance does not dominant the bubble response, the radius variation is small. Under these restrictions, approximate analytical solutions to the above-mentioned nonlinear differential equation can be found as follows. The instantaneous radius of bubble can be written as

$$R = R_0(1+x), \tag{2}$$

where x is the variation of bubble radius ($x \ll 1$). Substituting Eq. (2) into Eq. (1) will result in high-order terms of x^n, $n > 2$. If these terms are neglected, Eq. (1) can be expressed as

$$\rho R_0^2(1+x)\ddot{x} + \tfrac{3}{2}\rho R_0^2\dot{x}^2 = \left(p_0 + \frac{2\sigma}{R_0}\right)\left(1 - 3\gamma x + \frac{3\gamma(3\gamma+1)}{2}x^2\right) - \left(p_0 + \frac{2\sigma}{R_0}(1 - x + x^2)\right) - 4\mu(1-x)\dot{x} + p_{\text{ext}}(t). \tag{3}$$

To make the approximation valid, it is necessary that the external pressure $P_{\text{ext}}(t)$ should be small enough to ensure subharmonics,[19] the third- and higher-order harmonics being negligible. We can derive two-frequency approximation solutions as follows.

When the external pressure has two components, that is,

$$p_{\text{ext}}(t) = p_1\cos(2\pi f_1 t + \theta_1) + p_2\cos(2\pi f_2 t + \theta_2), \tag{4}$$

an approximate solution to Eq. (3) can be written as

$$x(t) = A_0 + A_1\cos(2\pi f_1 t + \phi_1) + A_2\cos(2\pi f_2 t + \phi_2) + A_3\cos(2\pi f_3 t + \phi_3) + A_4\cos(2\pi f_4 t + \phi_4) + A_5\cos(2\pi f_5 t + \phi_5) + A_6\cos(2\pi f_6 t + \phi_6), \tag{5}$$

where $f_5 = f_1 + f_2$ and $f_6 = f_1 - f_2$, and $A_1, A_2, A_3, \ldots$ are the amplitudes of the radial variations at the corresponding frequencies. From the results of Miller[14] and Newhouse and Shankar,[19] amplitude coefficients A_1 and A_2 in Eq. (5) can be expressed as

$$A_m = \frac{p_m X_m}{4\pi^2\rho f_0^2 R_0^2}, \quad m = 1,2, \tag{6}$$

where f_0 is the resonance frequency of microbubbles with radius R_0,

$$X_m = [(1-\Omega_m^2)^2 + \delta^2\Omega_m^2]^{-1/2}, \tag{7}$$

and the dimensionless frequencies are defined as

$$\Omega_m = \frac{f_m}{f_0}. \tag{8}$$

The damping constant is defined as

$$\delta = \frac{4\mu}{2\pi\rho f_0 R_0^2}. \tag{9}$$

The term $f_0 R_0$ corresponds to the polytropic exponent being adiabatic, and is approximately equal to 3.2 for air bubbles.

The other amplitude coefficients ($A_3 - A_6$) in Eq. (5) can be expressed as

$$A_3 = \tfrac{1}{2}A_1^2 Y_{11} X_{11}, \tag{10}$$

$$A_4 = \tfrac{1}{2}A_2^2 Y_{22} X_{22}, \tag{11}$$

$$A_5 = A_1 A_2 Y_{12} X_{12}, \tag{12}$$

$$A_6 = A_1 A_2 Y'_{12} X'_{12}, \tag{13}$$

where

$$X_{ij} = [(1-\Omega_{ij}^2)^2 + \delta^2 \Omega_{ij}^2]^{-1/2}, \tag{14}$$

$$Y_{ij} = \left\{ \left[\frac{\Omega_i^2 + \Omega_j^2}{2} + \frac{3}{2}\Omega_i \Omega_j + K \right]^2 + \frac{\delta^2}{4}\Omega_{ij}^2 \right\}^{1/2}, \tag{15}$$

$$\Omega_{ij} = \frac{f_i + f_j}{f_0}, \tag{16}$$

$$K = \frac{3\gamma(3\gamma+1)\left(p_0 + \frac{2\sigma}{R_0}\right) - \frac{4\sigma}{R_0}}{8\pi^2 \rho R_0^2 f_0^2}, \tag{17}$$

$$X'_{ij} = [(1-\Omega'^2_{ij})^2 + \delta^2 \Omega'^2_{ij}]^{-1/2}, \tag{18}$$

$$Y'_{ij} = \left\{ \left[\frac{\Omega_i^2 + \Omega_j^2}{2} - \frac{3}{2}\Omega_i \Omega_j + K \right]^2 + \frac{\delta^2}{4}\Omega'^2_{ij} \right\}^{1/2}, \tag{19}$$

$$\Omega'_{ij} = \frac{f_i - f_j}{f_0} \quad \text{for } ij = 11,\ 22,\ \text{or } 12. \tag{20}$$

Finally, the echo pressure P_m at a distance r and frequency f_m is

$$P_m = \frac{\rho (2\pi f_m)^2 R_0^3 A_m}{r}, \quad m = 1-6. \tag{21}$$

III. CONDITIONS FOR WEAK SHORT-PULSE RESPONSES

In this section we extend the results of Newhouse and Shankar[19] to study weak short-pulse responses. Based on Fourier theory, a short pulse can be synthesized by multiple sinusoids, with a two-frequency approximation being the simplest case. To construct the external pressure $p_{\text{ext}(t)}$ as a pulse train with a period large enough such that it approximates a short-pulse waveform, we need $p_1 \approx p_2 (= p)$ and $\Delta f (= f_1 - f_2)$ not so large as to cause spectral discontinuity of the transmitted waveform.

In Eq. (5), the fundamental responses (A_1 and A_2) are combined into a mixed response $A_F(t)$ around f_c, which is defined as the short-pulse fundamental response. The second-harmonic responses (A_3 and A_4) and the summation response (A_5) are combined into a composite second-harmonic response $A_C(t)$ around $2f_c$. Finally, since the difference response (A_6) is close to dc, it is renamed as the low-frequency response $A_L(t)$. Using these new definitions, Eq. (5) is rewritten as

$$x(t) = A_0 + A_F(t)\cos(2\pi f_c t + \phi_F) + A_C(t)\cos(4\pi f_c t + \phi_C) + A_L(t)\cos(2\pi \Delta f t + \phi_L). \tag{22}$$

It should be noted that using short-pulse excitation $p_{\text{ext}}(t)$ might not be sufficient to guarantee that the echo $x(t)$ is also a short pulse. For example, if $f_c = f_0$, $x(t)$ might be a single-frequency response only (which is not a short-pulse response). The conditions necessary to maintain $A_F(t)$, $A_C(t)$, and $A_L(t)$ as short-pulse responses are derived in the following. The amplitude and spectral properties of these three responses are also derived.

A. Fundamental response

From Eqs. (5) to (8) and (22), and letting $p_1 = p_2 = p$, the fundamental response becomes

$$A_F(t)\cos(2\pi f_c t + \phi_F) = A_1 \cos(2\pi f_1 t + \phi_1) + A_2 \cos(2\pi f_2 t + \phi_2), \tag{23}$$

where

$$A_1 = \frac{pX_1}{4\pi^2 \rho f_0^2 R_0^2}, \quad A_2 = \frac{pX_2}{4\pi^2 \rho f_0^2 R_0^2}, \tag{24}$$

$$X_1 = [(1-\Omega_1^2)^2 + \delta^2 \Omega_1^2]^{-1/2}, \quad \Omega_1 = \frac{f_1}{f_0} = \frac{f_c + \Delta f/2}{f_0}, \tag{25}$$

$$X_2 = [(1-\Omega_2^2)^2 + \delta^2 \Omega_2^2]^{-1/2}, \quad \Omega_2 = \frac{f_2}{f_0} = \frac{f_c - \Delta f/2}{f_0}. \tag{26}$$

It is apparent that the amplitude of $A_F(t)$ depends on the phase coherency of A_1 and A_2, and that bandwidth of the fundamental response is $\Delta f = f_1 - f_2$ if the spectrum of $A_F(t)$ is continuous.

To ensure that $A_F(t)$ is a short-pulse response, it is also required that $A_1 \approx A_2$. The relation between A_1 and A_2 depends primarily on f_1, f_2, and f_0. There are three possible relations among these frequencies.

(1) $f_1 \approx f_2 \approx f_0$, which is the case of a single-frequency response.

(2) One of f_1 or f_2 is close to f_0, which makes $A_1 \gg A_2$ or $A_2 \gg A_1$, and effectively this is a long-pulse response only. Note that both cases 1 and 2 do not fall within the scope of this paper.

(3) Both f_1 and f_2 are far away from f_0. It is found that if $\Omega_1 > 1 + \frac{3}{2}\delta$ or $\Omega_1 > 1 - \frac{3}{2}\delta$, then $(1-\Omega_1^2)^2 \gg \delta^2 \Omega_1^2$ in Eqs. (25) and (26). It is not difficult to satisfy this condition, since δ is always small, e.g., in a normal environment with $R_0 = 1\ \mu$m, $\frac{3}{2}\delta$ is around 10^{-1}, and when $R_0 = 10\ \mu$m, $\frac{3}{2}\delta$ is around 10^{-2} based on Eq. (9). Under these conditions, X_1 and X_2 can be approximated as

$$X_1 \approx |(1-\Omega_1^2)^{-1}|, \quad X_2 \approx |(1-\Omega_2^2)^{-1}|.$$

Then, the relative strength of A_1 and A_2 can be examined by their ratio as

$$\frac{A_1}{A_2} = \frac{X_1}{X_2} \approx \left| \frac{1-\Omega_2^2}{1-\Omega_1^2} \right| = \left| \frac{f_0^2 - (f_c^2 - f_c \Delta f + \Delta f^2/4)}{f_0^2 - (f_c^2 + f_c \Delta f + \Delta f^2/4)} \right|.$$

If $\Delta f / f_c$ is not too large, A_1 / A_2 can be approximated further as

$$\frac{A_1}{A_2} \approx \left| \frac{f_0^2 - f_c^2 + f_c \Delta f}{f_0^2 - f_c^2 - f_c \Delta f} \right|.$$

The relation between A_1 and A_2 is determined primarily by f_c and f_0 when Δf is fixed. There are three subcases.

(a) If f_c is equal to f_0, $A_1 \approx A_2$ and this equivalence is independent of Δf.

(b) If f_c is not equal to f_0, and $f_0 > f_1 > f_2$, then $f_0^2 > (f_c + \frac{1}{2}\Delta f)^2$ and $f_c \Delta f/(f_0^2 - f_c^2) < 1$. Under this condition, A_1/A_2 can be approximated as

$$\frac{A_1}{A_2} \approx \frac{1 + \dfrac{f_c \Delta f}{f_0^2 - f_c^2}}{1 - \dfrac{f_c \Delta f}{f_0^2 - f_c^2}} \approx 1 + \frac{2 f_c^2}{f_0^2 - f_c^2} B_e,$$

where $B_e = \Delta f / f_c$ is called the equivalent fractional bandwidth.

(c) If $f_0 < f_2 < f_1$, in the same way we have

$$\frac{A_1}{A_2} \approx 1 + \frac{2 f_c^2}{f_0^2 - f_c^2} B_e.$$

In cases (b) and (c), $[2 f_c^2/(f_0^2 - f_c^2)] B_e$ must be small enough to ensure $A_1 \approx A_2$. That is, B_e must be restricted within some range for a given f_c and f_0 to make this model work.

B. Composite second-harmonic response

From Eqs. (5) and (22), the composite second-harmonic response is defined as

$$\begin{aligned} A_C(t)\cos(4\pi f_c t + \phi_C) = {} & A_3 \cos(2\pi f_3 t + \phi_3) \\ & + A_4 \cos(2\pi f_4 t + \phi_4) \\ & + A_5 \cos(2\pi f_5 t + \phi_5). \end{aligned} \tag{27}$$

It comprises three components: the two second harmonics (A_3 and A_4), and the summation response (A_5). Since Δf has been restricted such that $A_F(t)$ is a short-pulse response, f_3, f_4, and f_5 will be close enough to satisfy the bandwidth requirement for $A_C(t)$ to also be a short-pulse response. We also require that A_3, A_4, and A_5 are similar to ensure that $A_C(t)$ is a short-pulse response. As $A_1 \approx A_2$, we will obtain $X_1 \approx X_2$ as well as $\Omega_1 \approx \Omega_2 \approx f_c/f_0$ from Eqs. (24) to (26). Therefore,

$$\Omega_{11} \approx \Omega_{22} \approx \Omega_{12} = \Omega_1 + \Omega_2 = \frac{2 f_c}{f_0},$$

$$X_{11} \approx X_{22} \approx X_{12}, \quad Y_{11} \approx Y_{22} \approx Y_{12}.$$

From Eqs. (10) to (12) and the above-noted approximation, A_3, A_4, and A_5 can be expressed as

$$A_3 = A_4 = \frac{1}{2} A_1^2 Y_{12} X_{12} = \frac{1}{2(4\pi^2 \rho f_0^2 R_0^2)^2} p^2 X_1^2 Y_{12} X_{12},$$

and (28)

$$A_5 = A_1^2 Y_{12} X_{12} = \frac{1}{(4\pi^2 \rho f_0^2 R_0^2)^2} p^2 X_1^2 Y_{12} X_{12}.$$

The similarity of A_3, A_4, and A_5 ensures that $A_C(t)$ is also a short-pulse response. The amplitude of $A_C(t)$ is dependent on the phase coherency of A_3, A_4, and A_5, and is proportional to A_1^2 as well as p^2 and X_1^2. As the radius of the microbubble decreases, X_1 increases as does $A_C(t)$. The bandwidth of $A_C(t)$ can be approximated as $f_3 - f_4 = 2\Delta f$, which is twice the bandwidth of $A_F(t)$.

C. Low-frequency response

From Eqs. (5), (13), and (22), the low-frequency response is defined as

$$A_L(t)\cos(2\pi \Delta f t + \phi_L) = A_6 \cos(2\pi f_6 t + \phi_6), \tag{29}$$

where $A_L(t) = A_6 = A_1 A_2 Y'_{12} X'_{12}$.

Under the conditions required for $A_F(t)$ to be a short-pulse response, $A_1 \approx A_2$ and $X_1 \approx X_2$, and hence Eq. (29) can be expressed as

$$A_L(t) = A_6 = A_1^2 X'_{12} Y'_{12} = \frac{1}{(4\pi^2 \rho f_0^2 R_0^2)^2} p^2 X_1^2 Y'_{12} X'_{12}. \tag{30}$$

The amplitude of $A_L(t)$ is then proportional to A_1^2, which is the same as that of $A_C(t)$. When the excitation pressure comprises two frequencies, there is only one component in $A_L(t)$. In a real wideband waveform, the number of components in $A_L(t)$ will increase. We can consider the position of the maximum frequency as the bandwidth of $A_L(t)$. The maximum-frequency position of $A_L(t)$ is therefore equal to the difference between the maximum frequency and the minimum frequency of the transmitted waveform. Here, the maximum-frequency position is at $f_1 - f_2$, and the bandwidth is Δf. The maximum-frequency position of $A_L(t)$ is related to Δf irrespective of the center frequency, which is an important property that allows $A_L(t)$ to be distinguished from subharmonic responses.

IV. DETECTABILITY OF THE LOW-FREQUENCY RESPONSE

The presence of $A_L(t)$ is a critical phenomenon of multifrequency excitation (i.e., it is not a long-pulse response), and so can be utilized as an indicator of a short-pulse response. Section III explains the existence of $A_L(t)$, but it does not guarantee that $A_L(t)$ can be detected in practice. Since B_e is small, $A_L(t)$ is close to dc and hence it might be filtered out by the instrumentation. The signal-to-noise ratio of $A_L(t)$ also affects the detectability, and furthermore this signal-to-noise ratio is also affected by B_e. In the following we analyze the detectability of the low-frequency response, especially related to B_e, theoretically and numerically.

A. Theoretical analysis

The bubble responses are mixed in the time domain, and hence the detectability of each component depends on their relative strength. Here the peak power ratio of the low-frequency response $A_L(t)$ to the fundamental response $A_F(t)$ is derived. Because $A_F(t)$ is affected by the coherency between A_1 and A_2, we can use either coherent or incoherent summation to obtain the peak power ratio; we chose the former method in this paper. From Eqs. (21), (24), and (30), the peak power ratio of $A_L(t)$ to $A_F(t)$ can be derived as

$$\left(\frac{P_6}{P_1+P_2}\right)^2=\left(\frac{(f_1-f_2)^2\dfrac{1}{(4\pi^2\rho f_0^2R_0^2)^2}p^2X_1^2Y'_{12}X'_{12}}{f_c^2(2)\dfrac{pX_1}{4\pi^2\rho f_0^2R_0^2}}\right)^2$$

$$=\frac{1}{4(4\pi^2\rho f_0^2R_0^2)^2}p^2X_1^2Y'^2_{12}X'^2_{12}B_e^4. \tag{31}$$

Since this peak power ratio is proportional to B_e^4, the detection of $A_L(t)$ requires that B_e is sufficiently large.

Furthermore, in Eq. (31) ρ, f_0, R_0, and p are frequency-independent constants, with only X_1, X_2, X'_{12}, and Y'_{12} being frequency dependent. X_1 and X_2 are not sensitive to the variance of bandwidth, so the bandwidth-dependent property is determined by X'_{12} and Y'_{12}. In Eqs. (18)–(20), Ω'_{12} is the common factor to show the bandwidth-dependent property of X'_{12} and Y'_{12}. Ω'_{12} can be expressed as

$$\Omega'_{12}=\frac{f_1-f_2}{f_0}=\frac{\Delta f}{f_0}=\frac{f_6}{f_0}. \tag{32}$$

From Eq. (19), Y'_{12} can be expressed as

$$Y'_{12}=\sqrt{Q+\Delta},$$

where

$$Q=\left[\frac{\Omega_1^2+\Omega_2^2}{2}-\frac{3}{2}\Omega_1\Omega_2+K\right]^2, \quad \Delta=\frac{\delta^2}{4}\Omega'_{12}.$$

Because $Q\gg\Delta$, $Y'_{12}\approx\sqrt{Q}$. Y'_{12} is less sensitive than X'_{12} to the variance of bandwidth. In Eq. (18), X'_{12} is maximal at $\Omega'_{12}=1$, i.e., $\Delta f=f_0$ in Eq. (32). Therefore, $A_L(t)$ is maximal at the same condition, which provides us with the short-pulse method for determining the microbubble radii. Both B_e and R_0 will affect X'_{12}. To illustrate the relationship between B_e and R_0 more clearly, replacing B_e into Eq. (32) yields

$$\Omega'_{12}=\frac{f_cB_e}{f_0}\approx\frac{f_cR_0B_e}{3.2}, \tag{33}$$

where $f_0R_0\approx3.2$ for air bubbles. X'_{12} is maximal at $\Omega'_{12}=1$, as well as $f_c(B_e/f_0)\approx1$ or $f_cR_0B_e\approx3.2$. After multiplying f_cR_0 by B_e, the maximum of the peak power ratio falls at the same value. Because smaller microbubbles have a higher resonance frequency, a larger B_e is needed to fit $\Delta f=f_0$ as well as to make $\Omega'_{12}=1$. For bubbles with sizes of the order of microns, which are of interest in this study, Δf must exceed hundreds of kilohertz. However, B_e has limitations as discussed in the derivation of fundamental responses, so it is impossible to reach this maximum since R_0 is too small.

To investigate the relationship between $A_C(t)$ and the bandwidth, we examine the bandwidth dependence of each term in A_3, A_4, and A_5. ρ, f_0, R_0, and p are frequency-independent constants from Eq. (28). From Eqs. (14) and (16), it can be shown that

$$X_{12}=\{[1-(2f_c)^2/f_0^2]^2+\delta^2(2f_c)^2/f_0^2\}^{-1/2}, \tag{34}$$

and so X_{12} is not bandwidth dependent. From Eqs. (15) and (16), it can be shown that

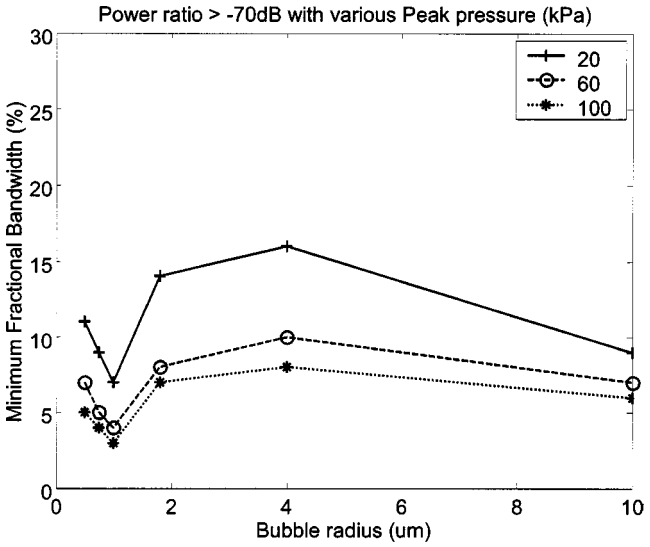


FIG. 1. The minimum value of B_e for different microbubble radii for peak pressures of 20, 60, and 100 kPa. The center frequency of the transmitted waveforms was 3 MHz.

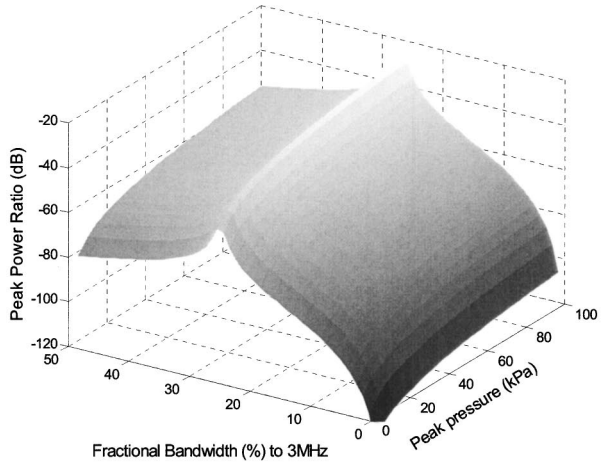
$$Y_{12}=\{[2.5(f_c^2-0.05\Delta f^2)/f_0^2+K]^2+\delta^2f_c^2/f_0^2\}^{1/2}. \tag{35}$$

Since $\Delta f<f_c$, Y_{12} is not sensitive to the variance of bandwidth. Neither X_{12} nor Y_{12} are sensitive to the variance of bandwidth, therefore, $A_C(t)$ is not sensitive to the variance of bandwidth.

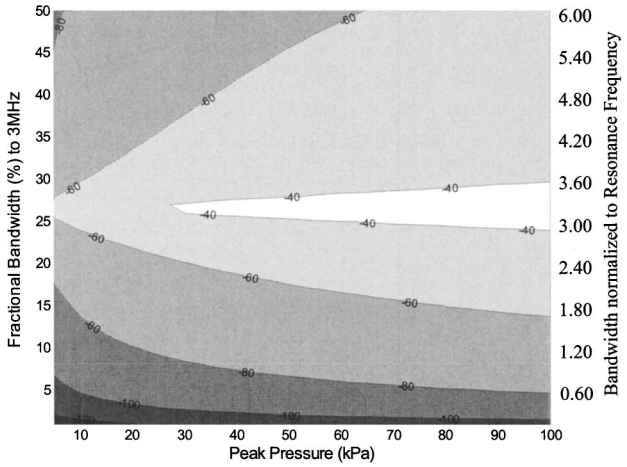
B. Numerical analysis

Equation (31) indicates that the microbubble radius, external pressure, and the bandwidth are the parameters that most affect the behavior of the system. For microbubble radius and peak pressure, we can select the minimum B_e—defined as the peak power ratio of $A_L(t)$ to $A_F(t)$—to be greater than -70 dB, as shown in Fig. 1 where f_c is 3 MHz. The minimum value of B_e can represent a criterion for determining if the microbubble response is the multifrequency response. At a fixed R_0, the peak power ratio is proportional to the peak pressure, as shown by Eq. (31). If p is larger, the minimum B_e will be smaller. This means that a larger range of pulse lengths can produce a short-pulse response. At a fixed peak pressure, the minimum B_e as a function of microbubble radius can be examined. When the radius is 1 μm, there is a local minimum B_e with a different peak pressure value. This is because the resonance frequency of this microbubble is about 3.2 MHz, which is close to f_c. The term X_1 in Eq. (31) becomes dominant, which violates the previously discussed restriction that f_c is far from f_0, and means that two frequency components cannot be considered a short-pulse waveform. As R_0 increases, f_0 decreases and the two-frequency approximation becomes valid. X'_{12} then becomes a more important factor for affecting the peak power ratio in Eq. (31). Since R_0 is larger than 4 μm, X'_{12} dominants the response.

To demonstrate the relationship between B_e and the power ratio more clearly, Fig. 2(a) presents a surface plot of the peak power ratio for different peak pressures and fractional bandwidths for a bubble radius of 4 μm and for $f_c=3$ MHz. The peak power ratios exhibit a local maximum at



(a)



(b)

FIG. 2. The peak power ratio of the low-frequency response to the fundamental response displayed as a surface plot (a) and a contour plot (b) for different peak pressures and fractional bandwidths of the transmitted waveform. Bubble radius was 4 μm and the center frequency was 3 MHz. The right-hand ordinates of the contour plot (b) display the fractional bandwidth normalized to the resonance frequency of the microbubbles.

some fixed peak pressure when B_e is 0%–50%, which result from microbubbles resonance. This is an important property of $A_L(t)$. A similar graph can be obtained for different B_e values with the peak pressure fixed, which can replace the conventional frequency response curve obtained by sweeping f_c. Figure 2(b) presents a contour plot presentation of the surface plot in Fig. 2(a). The peak power ratio again has a local maximum (or resonance effect) for each fixed peak pressure. Moreover, the resonance effect is obvious as the peak pressure increases. After B_e is multiplied by $f_c R_0$, the local maximum always has the same value (about 3.2).

Figure 3 presents a contour plot of the peak power ratio for different B_e and f_0, where $f_c = 3$ MHz and the peak pressure is 100 kPa. The maxima of the peak power ratios along f_0 for different B_e can be connected by a line as $f_0 = f_c B_e$. Therefore, to produce the maximum peak power ratios, B_e must be chosen to be proportional to the resonance frequency of microbubbles. If R_0 is too small (or equivalently f_0 is too large), there might be no suitable value of B_e for the reso-

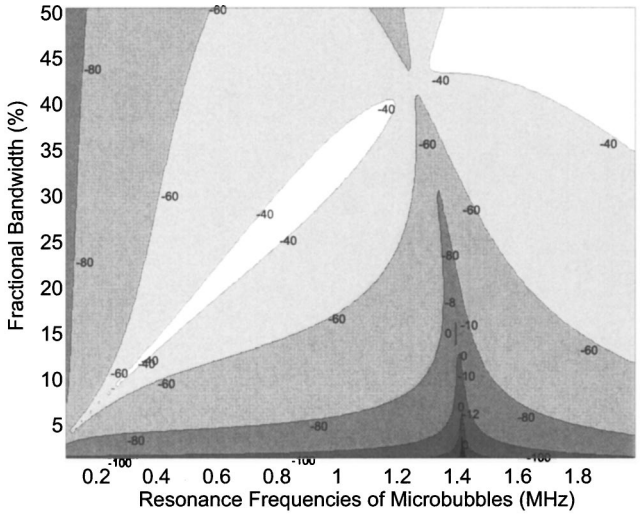


FIG. 3. The peak power ratio of the low-frequency response to the fundamental response for different resonance frequencies of microbubbles and fractional bandwidths of the transmitting waveform, displayed as a contour plot. The center frequency was 3 MHz.

nance effect to occur. There is another feature of the minima of the peak power ratios. It happens at the maximum of $A_F(t)$ when $f_c \approx f_0$, although it has violated the conditions for this model already. In practice, if we can combine the peak powers of $A_F(t)$ and $A_L(t)$, we can increase the measuring range of bubble sizes.

Figures 4 and 5 are computed with the same conditions used in Figs. 1 and 3, except now f_c is shifted to 5 MHz. The minimum B_e value now occurs at $R_0 = 0.75$ μm. Meanwhile, the effect of X_2 or X_1 is dominant for small R_0. X'_{12} starts to dominante the response when R_0 becomes larger than 2 μm, which is smaller than for the data shown in the result shown in Fig. 1. The results in Fig. 5 are similar to those in Fig. 3, in that connecting the maxima of the peak power ratios will also result in a line obeying $f_0 = f_c B_e$. Comparing the results

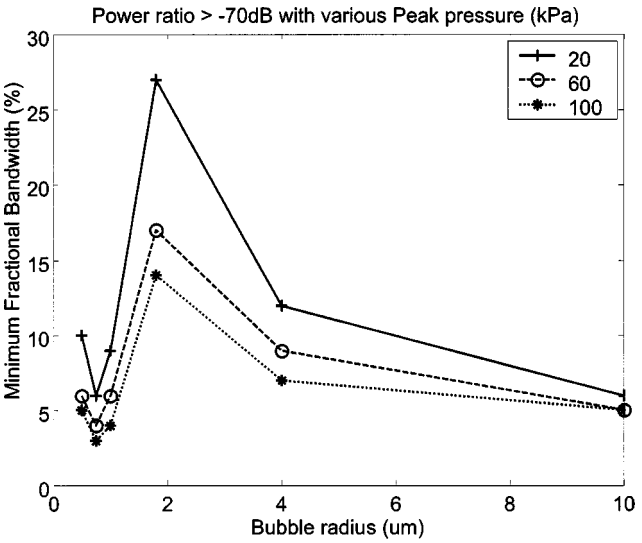


FIG. 4. The minimum value of B_e for different microbubble radii for peak pressures of 20, 60, and 100 kPa. The center frequency of the transmitted waveforms was 5 MHz.

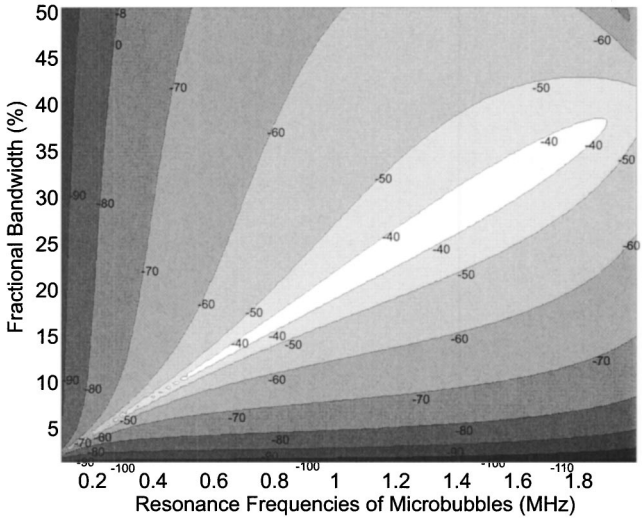


FIG. 5. The peak power ratio of the low-frequency response to the fundamental response for different resonance frequencies of microbubbles and fractional bandwidths of the transmitted waveform, displayed as a contour plot. The center frequency was 5 MHz.

of Figs. 3 and 5, reveals that for a given B_e, using a higher f_c would allow more bubbles with different sizes to resonate.

V. EXPERIMENTAL RESULTS

The low-frequency response $A_L(t)$ can be filtered out by the instrumentation (including the transducers), and hence the experimental settings should be determined carefully. The transmitted waveforms were generated by an arbitrary-function generator (TAG 1240, TTi), and fed to a power amplifier (Model 75A250, Amplifier Research). The received signals were amplified by a pulser/receiver (Model 5072PR, Panametrics) with a low-pass filter at 10 MHz and with no high-pass filtering, which is important for observing $A_L(t)$. Experimental data were recorded on a PC using a PCI-based A/D converter (PCI-9812, NuDAQ, ADLink) at a sampling rate of 20 MHz. The frequency responses displayed here were calculated by 1024-point fast fourier transforms and represent the averages of 250 presentations. The experimental setup is shown in Fig. 6. The setup includes a hollow tube with an inside diameter of 15 mm that is covered by silicon. This phantom serves as a window for observing flowing bubbles, and is totally immersed in water during experiments. The distance between the top boundary of the tube and water is 20 mm, and that between the bottom boundary of the tube and water is 40 mm. In addition to the silicon phantom, the system includes a degassing system, a roller pump, and flow tubes. Two piston probes (V325: center frequency, 2.25 MHz; V309: center frequency, 5.00 MHz; focal distance of both pistons is 20 mm) are utilized as transmitting and receiving transducers. The angle between the flow tube and the beam direction for both probes is about 45°, and both probes are focused at the same point in the middle of the flow tube. To avoid the uncertainty about the size of free-gas bubble used in the experiments, we used a commercially available source of microbubble, Levovist® (Schering AG, Berlin, Germany) whose radius is distributed around 2–3 μm.[25] We use one injection vial (2.5 g and 7 ml water)

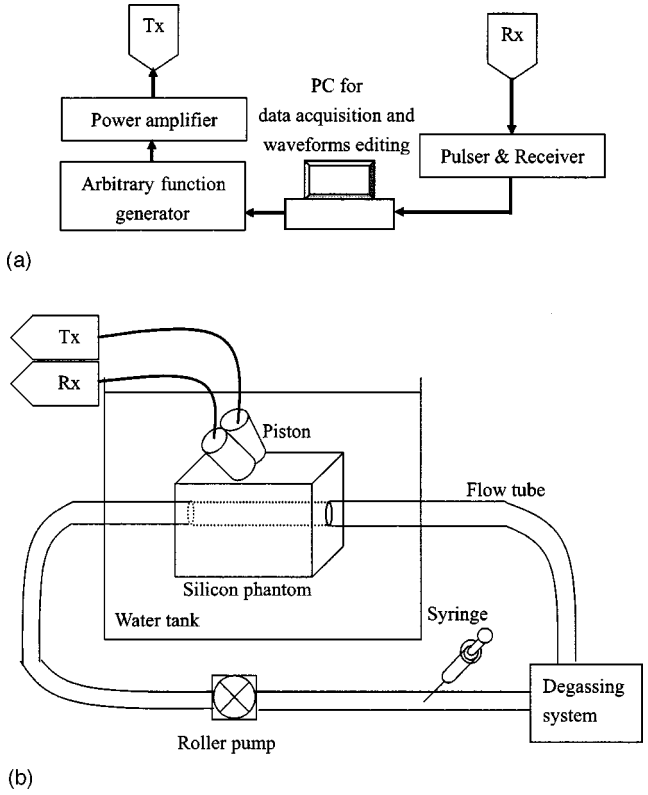


FIG. 6. Experimental instrumentation and setup.

for each experiment. Before the microbubble is injected, a degassing process is applied to ensure that the concentration of residual microbubbles is low enough so that they will not influence the subsequent experimental results. After each injection, data received within the first minute are ignored to ensure that the data analyzed correspond to a uniform distribution of microbubbles in the flow tube.

The peak pressures of the transmitted waveforms before they pass through the phantom are measured by a hydrophone (NP-1000 PVDF needle hydrophone, Model TNU001A, NTR Systems) connected to a 30 dB preamplifier (NTR Systems), which is mounted on a motorized three-dimensional positioning system. The measured pressure after compensating for the attenuation effect of the flow phantom was 455 ± 50 kPa.

Sixteen waveforms were utilized in the experiment to allow sweeping of f_c and the fractional bandwidth. f_c was swept from 2.75 to 3.50 MHz with a step increment of 0.25 MHz. The fractional bandwidth was changed from 10% to 40% in steps of 10%. The received data corresponding to these sweeps were collected during a single injection to ensure that the radii and position distributions of the microbubbles were similar. Since the received signals are influenced by the Doppler effect, their frequency responses are displayed as a rf spectrum after compensating for the gain settings of the pulser/receiver, and processing by a wall filter.[26] We also take the maximum peak pressure of these transmitted waveforms as a baseline to compensate for differences in the excitation pressure.

Interpolating the data for the four values of fractional bandwidth (10%, 20%, 30%, and 40%) allows the frequency

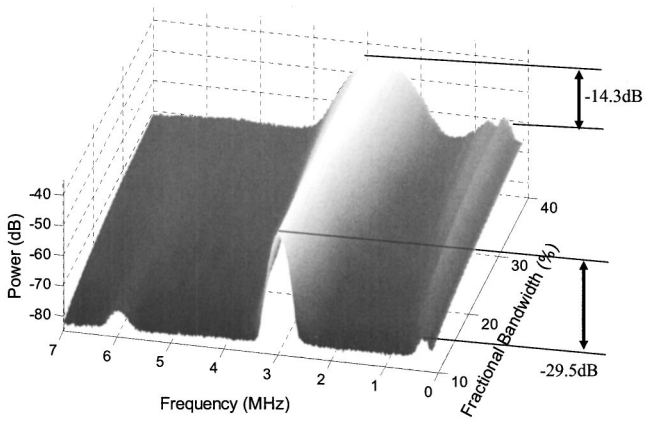


FIG. 7. Frequency responses of microbubbles for different fractional bandwidths: 10%, 20%, 30%, and 40%. The center frequency of the transmitted waveform was 3 MHz. The power ratio of the low-frequency response to the fundamental response increases from −29.5 to −14.3 dB as the fractional bandwidth increases from 10% to 40%.

responses of the microbubbles to be displayed as a surface plot (Fig. 7). After compensating for differences in the excitation pressure, the noise levels are not the identical since the signal-to-noise ratio varies with the transmitted waveforms. An increased transmitted fractional bandwidth corresponds to increased bandwidths of $A_F(t)$, $A_C(t)$, and $A_L(t)$. Moreover, the peak power ratio of $A_L(t)$ to $A_F(t)$ increases from −29.5 to −14.3 dB, but the peak power ratio of $A_C(t)$ to $A_F(t)$ increases from −31.4 to −24.1 dB only. This bandwidth-dependent property is proven theoretically in Sec. IV.

Figures 8(a) and (b) show the spectra of the microbubble responses at two center frequencies: 2.75 and 3.25 MHz, respectively. Different fractional bandwidths are given in each figure. The maximum-frequency position (also called the bandwidth) of $A_L(t)$ is equal to the difference between the maximum frequency and the minimum frequency of the transmitted wave form, as discussed in Sec. III. When the fractional bandwidth is over 30%, the spectra of $A_L(t)$ and $A_F(t)$ begin to merge, as shown by the dotted line in each figure. As f_c increases, the spectrum of $A_L(t)$ does not shift along the frequency axis. The absence of an obvious response at half the center frequency [1.375 and 1.625 MHz in Figs. 8(a) and (b), respectively] confirms that $A_L(t)$ is not the subharmonic response. The bandwidth of $A_C(t)$ increases with the fractional bandwidth also. These features are not obvious in Fig. 8(b) because the larger attenuation in the high-frequency region brings $A_C(t)$ down to the noise level.

The peak power ratio of $A_L(t)$ to $A_F(t)$ is shown in Fig. 9 as a function of fractional bandwidth. When the fractional bandwidth increases, the peak power ratio increases as given by Eq. (31). When the fractional bandwidth is over 30%, the increase in the peak power ratio tends to saturate. This might be the result of interference due to $A_F(t)$ and $A_L(t)$ overlapping. $A_F(t)$ is attenuated more as f_c increases, whereas $A_L(t)$ is invariant with f_c; therefore, the peak power ratio increases as f_c increases. As discussed in Sec. IV, the peak power ratio of $A_L(t)$ to $A_F(t)$ consists of local maximum (resonance effect) within the proper fractional bandwidth range at some fixed peak pressure. However, this is not evi-

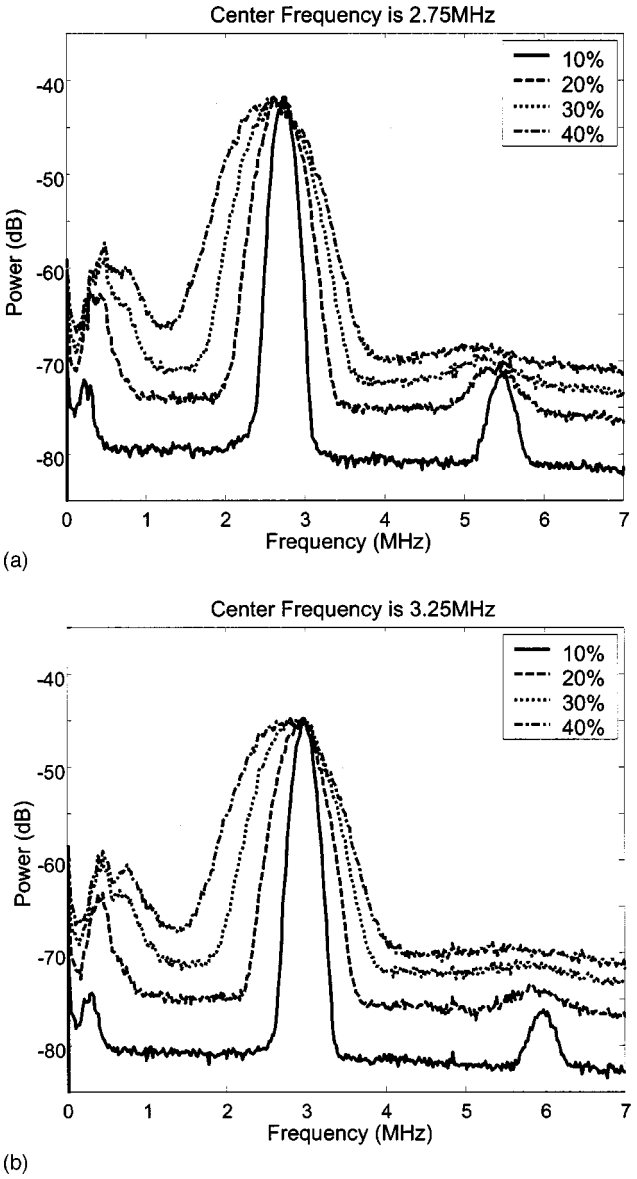


FIG. 8. Frequency responses of microbubbles for different center frequencies and fractional bandwidths of the transmitted waveform for center frequencies of 2.75 MHz (a) and 3.25 MHz (b). The solid, dashed, dotted, and dashed-dotted lines denote fractional bandwidths of 10%, 20%, 30%, and 40%, respectively, in each graph.

dent in Fig. 9, which may be attributable to two phenomena: (1) the peak pressures of the transmitted waveforms are not large enough to produce a local maximum, which tends to reside at higher-pressure regions of the contour plot in Fig. 2(b); or (2) bubbles used in the experiments vary in size resulting in different positions of the local maxima such that the resonance effect is smoothed out among these microbubbles.

Finally we examine the peak power ratio versus the peak power and the fractional bandwidth of $A_F(t)$ simultaneously, with f_c fixed at 3 MHz. Figure 10 shows the peak power ratios versus the peak power of $A_F(t)$. The solid lines represent the peak power ratio of $A_L(t)$ to $A_F(t)$, and each line shows that for a fixed fractional bandwidth, the peak power ratio increases with the peak power of $A_F(t)$. For a constant peak power of $A_F(t)$, the peak power ratio increases with the

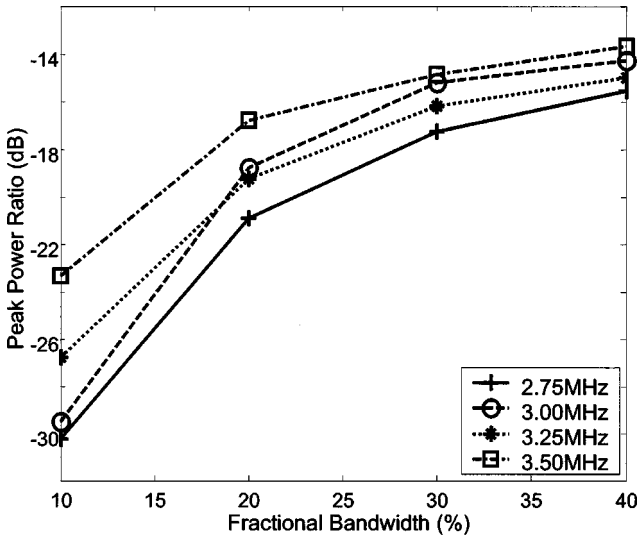


FIG. 9. Peak power ratio of the low-frequency response to the fundamental response for different fractional bandwidths of the transmitted waveform for center frequencies of 2.75 MHz (solid line), 3.00 MHz (dashed line), 3.25 MHz (dotted line), and 3.50 MHz (dashed-dotted line).

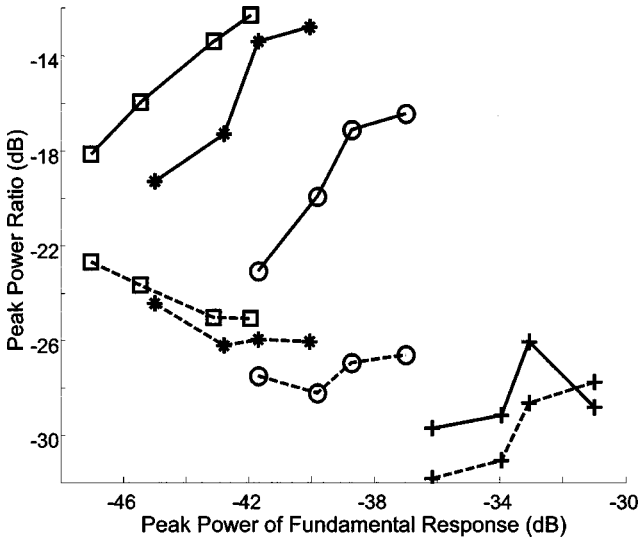


FIG. 10. Solid line represents the peak power ratio of the low-frequency response to the fundamental response. Dashed line represents the peak power ratio of the composite second-harmonic response to the fundamental response. The plus sign, circle, asterisk, and square represent fractional bandwidths of 10%, 20%, 30%, and 40%, respectively.

fractional bandwidth. These two tendencies are derived theoretically in Sec. IV.

The dashed lines in Fig. 10 indicate the peak power ratio of $A_C(t)$ to $A_F(t)$. They are consistent with two tendencies mentioned in the previous section, except for four inconsistent points. As indicated by the dashed lines for factional bandwidth of 20%–40%, the peak power ratios do not follow the weaker peak power of $A_F(t)$ consistently; this is because $A_C(t)$ is attenuated more. Furthermore, these peak power values of $A_C(t)$ may include power from noise, which would result in overestimation of the peak power ratios.

VI. CONCLUSIONS

The peak pressure and bandwidth of the transmitted waveforms, and the radius of microbubbles are the most important factors of the short-pulse responses investigated in this study. There are three restrictions to the validity of the approximation model we have used: (1) the peak pressure should not be large enough to generate higher-order harmonic or subharmonic responses, (2) the resonance frequency of microbubbles must be far from the center frequency of the transmitted waveform, since otherwise the microbubbles may resonate irrespective of whether the pulse is long or short, and (3) the two frequency components must be close enough to allow them to be considered as a single short-pulse waveform. Under these three restrictions, the weak short-pulse responses of microbubbles can be derived analytically using the two-frequency approximation method. The short-pulse responses can be divided into three components: the low-frequency response, the fundamental response, and the composite second-harmonic response. It is worth noting that if the requirement $x \ll 1$ in Eq. (2) is relaxed, then the responses of the bubbles would be more complicated and the stability of our analytic solutions might be a problem; this should be investigated in future studies.

Equation (24) shows that $A_F(t)$ is a linear wideband response. Any frequency component around the resonance frequency of the microbubbles will invalidate the short-pulse approximation, which is why short-pulse waveforms cannot be analyzed at the resonance frequency in this way. The bandwidth of $A_F(t)$ is the bandwidth Δf of the transmitted waveform. For this model to work, the conditions for short-pulse responses must be established based on restricting the bandwidth of $A_F(t)$.

$A_C(t)$ contains three components: the summation response and the second-harmonic responses. We have shown that its strength is not sensitive to the variation of bandwidth. The bandwidth of $A_C(t)$ is twice the bandwidth of $A_F(t)$. It is difficult to distinguish these three components due to their spectral overlap and since they are both proportional to the square of the fundamental responses. Furthermore, the power of $A_C(t)$ is less than the power of $A_F(t)$ and $A_L(t)$ due to attenuation present in experimental conditions. These factors make experimentation on $A_C(t)$ difficult.

$A_L(t)$ is a component that was predicted to be present during two-frequency excitation, and this was investigated in this study. Its existence can provide us with an important indicator for determining whether multiple frequencies are present in the response. This is the first step to studying the short-pulse response. Its bandwidth increases in proportion to the bandwidth of the transmitted waveform. Equation (31) shows that the detectability of $A_L(t)$ depends greatly on the bandwidth of the transmitted waveform and its power ratio to $A_F(t)$. $A_L(t)$ may be filtered by instrumentation when the bandwidth is too small. It is proportional to the square of the peak pressure, and we have also shown that its strength is bandwidth dependent. It is maximal at $\Delta f = f_0$, and its bandwidth-dependent property may enable determination of the bubble sizes (rather than using the swept-frequency method). If we can combine the peak powers of $A_F(t)$ and $A_L(t)$, we can extend the range of bubble sizes that can be measured. Because the attenuation is less at lower frequencies, and furthermore a smaller excitation sound pressure can be used to generate $A_L(t)$, it may be useful for

improving the signal-to-noise ratio in medical imaging, especially in depth.

The subharmonic response, which resides at $f_c/2$, is another response that may appear in the low-frequency region. The subharmonic response may occur for higher-pressure excitation due only to the threshold effect. $A_L(t)$ resides at the difference of the excitation frequencies irrespective of the center frequency of the transmitted waveform. The occurrence of subharmonics can be verified by sweeping the center frequency. We confirmed that no subharmonics were present in any of our experiments. The detailed relationships between $A_L(t)$ and the subharmonic response, such as the coherency when they reside at the same frequency and the bandwidth-dependent properties, need further investigation.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Yi-Hong Chou for his comments regarding the clinical importance of contrast agents. This work was supported by NSC Grant No. NSC 91-2314-B-075-122.

[1] A. Prosperetti, ''Nonlinear oscillations of gas bubble in liquids: Transient solutions and the connection between subharmonic signal and cavitation,'' J. Acoust. Soc. Am. **57**, 810–821 (1977).
[2] H. G. Flynn, ''Generation of transient cavities in liquids by microsecond pulses of ultrasound,'' J. Acoust. Soc. Am. **72**, 1926–1932 (1982).
[3] R. E. Apfel, ''Acoustic cavitation: A possible consequence of biomedical uses of ultrasound,'' Br. J. Cancer **45**, 140–146 (1982).
[4] G. R. ter Harr, S. Daniels, and K. Morton, ''Evidence for acoustic cavitation *in vivo*: Thresholds for bubble formation with 0.75-MHz continuous wave and pulsed beams,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **UFFC-33**, 162–198 (1986).
[5] J. B. Fowlkes and L. A. Crum, ''Cavitation threshold measurements for microsecond length pulses of ultrasound,'' J. Acoust. Soc. Am. **83**, 2190–2201 (1988).
[6] H. G. Flynn and C. C. Church, ''Erratum: Transient pulsations of small gas bubbles in water,'' J. Acoust. Soc. Am. **84**, 985–998 (1988).
[7] V. A. Akulichev, V. A. Bulanov, and S. A. Klenin, ''Acoustic sensing of gas bubbles in the ocean medium,'' Sov. Phys. Acoust. **32**, 177–180 (1986).
[8] H. R. Suiter, ''Pulse length effects on the transmissivity of bubbly water,'' J. Acoust. Soc. Am. **91**, 1383–1387 (1992).
[9] N. G. Pace, A. Cowley, and A. M. Campbell, ''Short pulse acoustic excitation of microbubbles,'' J. Acoust. Soc. Am. **102**, 1474–1479 (1997).
[10] K. E. Morgan, P. A. Dayton, D. E. Kruse, A. L. Klibanov, G. H. Brandenburger, and K. W. Ferrara, ''Changes in the echoes from ultrasonic contrast agents with imaging parameters,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 1537–1548 (1998).
[11] J. Ophir and K. J. Parker, ''Contrast agents in diagnostic ultrasound,'' Ultrasound Med. Biol. **15**, 319–333 (1989).
[12] P. J. Frinking, A. Bouakaz, J. Kirkhorn, F. J. ten Cate, and N. de Jong, ''Ultrasound contrast imaging: Current and new potential methods,'' Ultrasound Med. Biol. **26**, 965–975 (2000).
[13] D. G. Tucker and V. G. Welsby, ''Ultrasonic monitoring of decompression,'' Lancet I, 1253 (1968).
[14] D. L. Miller, ''Ultrasonic detection of resonant cavitation bubbles in a flow tube by their second-harmonic emissions,'' Ultrasonics **19**, 217–224 (1981).
[15] N. de Jong, ''Improvements in ultrasound contrast agents,'' IEEE Eng. Med. Biol. Mag., 72–82 (1996).
[16] P. M. Shankar, P. D. Krishna, and V. L. Newhouse, ''Advantages of subharmonic over second harmonic backscatter for contrast-to-tissue echo enhancement,'' Ultrasound Med. Biol. **24**, 395–399 (1998).
[17] W. T. Shi, F. Forsberg, J. S. Raichlen, L. Needleman, and B. B. Goldberg, ''Pressure dependence of subharmonic signals from contrast microbubbles,'' Ultrasound Med. Biol. **25**, 275–283 (1999).
[18] J. R. Klepper, ''Ultrasonic methods for detection and sizing of venuous gas emboli,'' IEEE Ultrasonics Symposium, 1992, pp. 27–29.
[19] V. L. Newhouse and P. M. Shankar, ''Bubble size measurements using the nonlinear mixing of two frequencies,'' J. Acoust. Soc. Am. **75**, 1473–1477 (1984).
[20] R. M. Schmitt, H. J. Schmidt, O. Kripfgans, and C. Poser, ''A novel ultrasonic resonance spectrometer for microbubble sizing,'' IEEE Ultrasonics Symposium, 1992, pp. 941–946.
[21] A. D. Phelps, D. G. Ramble, and T. G. Leighton, ''The use of a combination frequency technique to measure the surf zone bubble population,'' J. Acoust. Soc. Am. **101**, 1981–1989 (1997).
[22] K. E. Morgan, J. S. Allen, J. E. Chomas, P. A. Dayton, and K. W. Ferrara, ''Experimental and theoretical analysis of individual contrast agent behavior,'' IEEE Ultrasonics Symposium, 1999, pp. 1685–1688.
[23] T. G. Leighton, *The Acoustic Bubble* (Academic, New York, 1994), Chap. 4.
[24] N. de Jong, R. Cornet, and C. T. Lancee, ''Higher harmonic of vibrating gas-filled microspheres. 1. Simulations,'' Ultrasonics **32**, 447–453 (1994).
[25] M. Bloomley, ''Clutter reduction using harmonic power angio with Levovist helps in assessment of liver metastases,'' Medica Mundi **42**, 15–18 (1998).
[26] P. Fish, *Physics and Instrumentation of Diagnostic Medical Ultrasound* (Wiley, New York, 1990), Chap. 11.

Measurements of near-surface bubble plumes in the open ocean with implications for high-frequency sonar performance

Mark V. Trevorrow[a)]
Defence Research & Development Canada—Atlantic, P.O. Box 1012, Dartmouth, Nova Scotia B2Y 3Z7, Canada

(Received 9 October 2002; revised 2 August 2003; accepted 3 September 2003)

This study examines near-surface bubble data obtained with a self-contained 200-kHz inverted echo-sounder deployed at Ocean Station Papa (NE Pacific, 1400 km west of Vancouver Is.) over an 81-day period in the spring of 1996. The instrument operated continuously, recording calibrated volume scattering profiles from near-surface bubbles with 3-s and 30-cm resolution. The data show the frequent occurrence of bubbles organized into vertical, plume-like structures, presumably drawn downwards by turbulence and other near-surface circulations. Average bubble plume penetrations of up to 15 m were observed, with maximum penetrations up to 25 m. Within the plumes, the backscatter cross section exhibited an exponential decay with depth, with e-folding scale in the range 0.5 to 3 m, increasing proportionally to the square of average plume depth. Using standard models for bubble scattering, and incorporating recent acoustic resonator measurements of bubble-size distributions along with actual bubble plume data, high-frequency near-surface sonar performance models were developed. These models show that on a ping-to-ping basis the bubble plume structures can induce significant spatial variations in the reverberation level and path-integrated extinction losses to near-surface targets. [DOI: 10.1121/1.1621008]

PACS numbers: 43.30.Ft [WMC] Pages: 2672–2684

I. INTRODUCTION

The performance of high-frequency (HF, >10 kHz) sonars in the near-surface ocean is strongly affected by acoustic scattering from small air bubbles created by breaking waves. High-resolution HF sonars are important in this environment for such applications as torpedo and mine detection, obstacle avoidance (e.g., boats and whales), or fisheries surveys. The properties of near-surface bubbles have been the subject of considerable research over the past 20 years, motivated both by naval interests in understanding surface reverberation (e.g., Hall, 1989; Dahl, 2003) and for climatological reasons arising from the fact that bubbles mediate air–sea gas exchange (e.g., Farmer *et al.*, 1993). Breaking-wave induced bubbles typically begin to appear above 6–10-knots wind speed (Thorpe, 1982; Crawford and Farmer, 1987; Dahl *et al.*, 1997; Dahl, 2003), and thus are a common occurrence in the open ocean. In the immediate vicinity of a breaking wave the bubble densities can be very high, with total void fraction up to roughly 10% (Monahan and Lu, 1990; Lamarre and Melville, 1994; Dahl and Jessup, 1995); however, once injected into the near-surface ocean these bubbles are dispersed and advected by surface waves, turbulence, and other circulation phenomena (see Monahan and Lu, 1990 for a description of bubble generation and dispersal processes). The bubble density and depth of penetration generally increase as the wind speed increases; however, there are significant spatial variations due to the patchiness of wave breaking and near-surface circulations. Most previous HF sonar studies in the near-surface zone, while acknowledging the cloud- or plume-like structure of near-surface bubble clouds, have usually chosen to describe the bubble effects in a time-averaged, horizontally uniform sense (e.g., Novarini and Bruno, 1982; Thorpe, 1986; Crawford and Farmer, 1987; Wu, 1988; Hall, 1989; Nützel *et al.*, 1994). One exception is a recent HF modeling effort by Norton *et al.* (1998), which confirmed the importance of the bubble plume structures on near-surface propagation. Similarly, both McDonald (1991) and Henyey (1991) utilized bubble plume structures in their modeling of low-frequency (<2 kHz) surface reverberation. The focus of this present study is to extract high-resolution data on bubble spatial structures, and then to assess the implications of these bubble structures for the operation of HF sonars at near-horizontal geometries in the near-surface ocean.

High-frequency acoustic backscatter systems are sensitive tools for the investigation of near-surface bubbles. These applications have generally focused on the characteristics of natural oceanic bubble clouds (e.g., Thorpe, 1982, 1986), or on the use of these bubbles as tracers of near-surface fluid flows such as surface and internal gravity waves (Thorpe and Hall, 1983; Trevorrow, 1995, 1998), turbulence (Garrett *et al.*, 2000), and Langmuir circulation (Zedel and Farmer, 1991). In the open ocean, autonomous inverted echo-sounder (IES) devices have been used previously for investigation of bubble clouds (e.g., Thorpe, 1986; Zedel and Farmer, 1991; Vagle and Farmer, 1992). IES systems have the advantage of remotely sensing bubble properties while positioned safely away from the destructive motions of the ocean surface. With advances in electronics technology these systems can provide increased spatial and temporal resolution along with increased recording capacity. The main limitation in the use of IES systems is that bubble extinction and multiple-scattering processes make measurements at medium to high

[a)]Electronic mail: mark.trevorrow@drdc-rddc.gc.ca

bubble densities (i.e., void fractions $>1\times10^{-5}$) difficult or impossible (see Vagle and Farmer, 1998).

This study focuses on an 81-day deployment of a 200-kHz IES in the open Pacific Ocean during the spring of 1996. Data on near-surface bubble distributions were collected continuously with 3-s and 25-cm resolution. Working within the limitations of the instrument, some detailed measurements of bubble plume structure and occurrence are presented in Sec. III. Section IV presents a simple model for HF bubble backscatter and extinction, and then examines reverberation and target detection levels with emphasis on the importance of the bubble spatial structures.

II. INSTRUMENTATION AND METHODS

The focus of this paper is the data from a self-contained 200-kHz IES, which was deployed and operated continuously from 0130UT 01 March until 1400UT 20 May 1996. The mooring location was Ocean Station Papa (OSP, N50° W145°), situated approximately 1400 km west of Vancouver Island in the NE subarctic Pacific. The IES was deployed at the top of a subsurface mooring at a nominal depth of 43 m in a total water depth of approximately 4100 m. OSP has been the site of oceanographic and ecological monitoring since the late 1950's.

This 200-kHz IES was a prototype instrument developed at the Institute of Ocean Sciences (Sidney, BC, Canada). It was designed to be a simple, self-contained system capable of deployments up to 1 year in length. Thus, low power consumption and large data storage capability were the dominant design factors, necessitating some compromises in overall performance. This same instrument was deployed on the Oregon coast in study of internal solitary waves (Trevorrow, 1998) and an updated version of this instrument is described in Lemon *et al.* (2001). The electronics and batteries were contained in a 25-cm-diameter by 80-cm-long cylindrical pressure housing, supported by streamlined syntactic foam floatation with a strong righting moment. The 200-kHz transducer was mounted externally on the top, oriented vertically upwards. The transducer -3-dB beamwidth was 4.6°. The IES operated continuously, transmitting a 0.5-ms pulse at 1-s intervals. Backscatter amplitude was acquired with 8-bit resolution (48-dB dynamic range) in 21.5-cm depth bins spanning 10 to 53 m (upward) range. No time-varying gain was used, which turned out to be beneficial for detection of near-surface bubble plumes which have increasing scattering strength with (upward) range. The receiver gain was configured as a compromise to resolve backscattering from both meso-zooplankton and near-surface bubble plumes; thus, its dynamic range spanned volume backscatter levels between -85 and -30 dB (re: 1 m^{-1}). To reduce the overall data storage requirements, range-binned data from three successive pings (3 s in time) were averaged before storage. The raw sounder data were stored internally on two standard PC hard drives.

Using standard echo-sounder relations (e.g., Medwin and Clay, 1998), the raw IES amplitude, A, can be converted to *volumetric backscatter strength* (S_v, decibels re: 1 m^{-1}) vs range (upwards) from the sounder, r, i.e.,

$$S_v(r) = 20\log_{10}[A(r)] + K + 40\log_{10}[r] + 2\cdot\alpha\cdot r - 10\log_{10}[U(r)], \quad (1)$$

where K is a calibration coefficient, r is range ($=\frac{1}{2}\cdot c\cdot t$, where c is sound speed and t is time after transmission), α is the acoustic absorption, and $U(r)$ is the insonified volume. S_v is the decibel equivalent of the volumetric backscatter cross section, s_v (m^{-1}), and any arithmetic operations (i.e., averaging) are performed on the cross-section data. The calibration coefficient K includes transmit power level, transducer sensitivity, preamplifier gains, and A/D conversion factors. The acoustic absorption is frequency- and environment dependent, and for the typical water conditions at OSP has a value of 0.044 dB·m^{-1} at 200 kHz (Francois and Garrison, 1982). In terms of the *equivalent solid angle* of the transducer, φ, and the pulse duration, $\tau(=500\ \mu s)$, the insonified volume is given by

$$U(r) = \tfrac{1}{3}\varphi[(r+\tfrac{1}{4}c\tau)^3 - (r-\tfrac{1}{4}c\tau)^3]. \quad (2)$$

For this IES the insonified volume increased from 0.2 m^3 at 10-m range to 3.4 m^3 at 42 m (the nominal ocean surface), and is thus small compared to the spatial scales of ocean surface waves and bubble plumes under study.

The IES was calibrated using as reference the backscatter target strength of tungsten–carbide spheres following methods outlined in Vagle *et al.* (1996). Given the calibrations and geometric variations in volume scattering strength, both the noise threshold and clipping levels were range dependent. With a three-ping average, the equivalent noise amplitude was approximately 0.5 digital counts, corresponding to S_v from -92 dB at 10-m range to -76 dB at the ocean surface. Similarly, the A/D conversion clipping level of 250 counts corresponded to S_v from -43 dB at 10-m range to -28 dB at the ocean surface. Because of this limitation, features with higher volumetric cross sections, such as the ocean surface echo and denser portions of the bubble plumes, were underestimated.

Strictly, in the presence of near-surface bubbles Eq. (1) should include a correction for bubble extinction. Specifically, the $S_v(z)$ should be corrected for the integrated extinction loss from bubbles at greater depths (see Vagle and Farmer, 1992). However, for this particular IES this correction is unnecessary due to the relatively low $S_v(z)$ at the A/D clipping limit. Based on bubble extinction models described in a later section, at the maximum measurable value of $S_v(z)$ near -30 dB the extinction cross section is at most 0.04 m^{-1} (0.17 dB·m^{-1}). This, in combination with the relatively short vertical propagation distance ($<$ a few meters) through the denser bubble regions, makes extinction loss corrections small.

III. OBSERVATIONS

A. Oceanographic setting

Ocean Station Papa lies approximately 1400 km west of Vancouver Island in the NE subarctic Pacific, near the south-

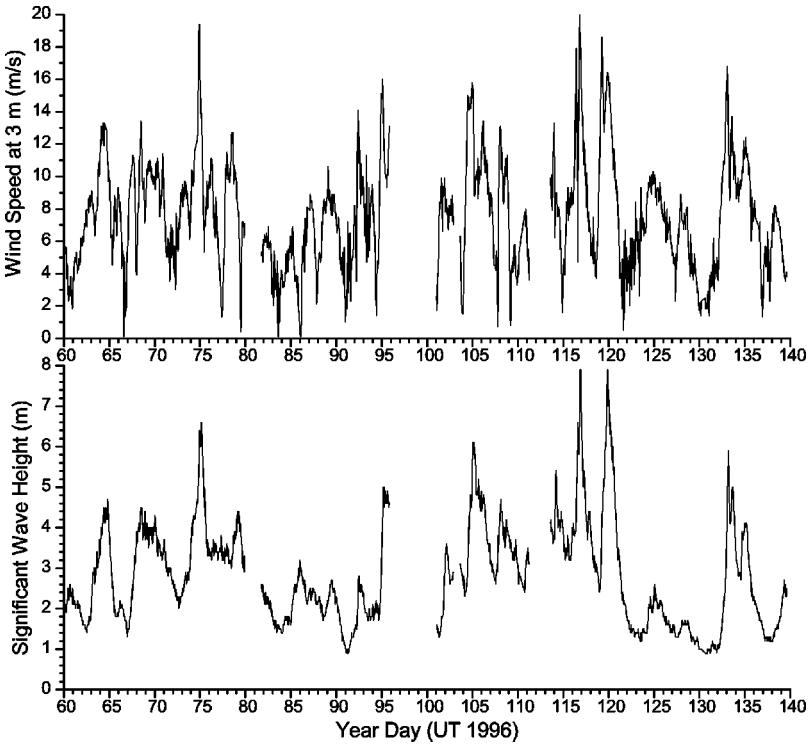


FIG. 1. Wind speed and significant wave height from Environment Canada ocean weather buoy 46004, located at N50° 56′, W135° 52′. 1 March 1996 is year day 61. Data courtesy Dr. J. Gower at the Institute of Ocean Sciences.

east boundary of the Alaskan Gyre. This 81-day time series of 200-kHz IES data covered a range of conditions spanning the late winter and spring of 1996. A companion time series of wind speed and significant wave-height data from Ocean Weather Buoy 46004 is shown in Fig. 1. Unfortunately, while this is the nearest ocean weather buoy to OSP, it was located 650 km to the east and thus cannot be used for direct comparisons to the OSP acoustic data. However, owing to the generally ENE movement of weather systems in the NE Pacific, the *general* characteristics of the wind and waves are expected to be similar between the two locations. During this 80-day period the wind speeds ranged up to $20\,\mathrm{m\cdot s^{-1}}$ and the significant wave height varied from 1 to 8 m, with no discernible seasonal trend in either. Average values of wind speed and significant wave height were $7.5\,\mathrm{m\cdot s^{-1}}$ and 2.8 m, respectively. It should be noted that this average wind speed significantly exceeds the generally accepted threshold for breaking-wave activity of roughly $3\,\mathrm{m\cdot s^{-1}}$ (Dahl and Jessup, 1995; Dahl *et al.*, 1997; Dahl, 2003), implying that on average white-capping and subsurface bubble plumes should be present. The air–sea temperature difference at the weather buoy varied between $-5°$ and $+1.4\,°\mathrm{C}$, with mean of $-0.82\,°\mathrm{C}$, implying on average unstable atmospheric boundary layer conditions. Thorpe (1982) showed that atmospheric boundary layer stability influences the character of subsurface bubble plumes, with deeper, more columnar clouds forming under unstable sea-surface conditions. Again, due to the separation of the weather buoy from OSP, a detailed assessment of the effects of atmospheric boundary layer stability cannot be performed on these data.

Oceanographically, the near-surface conditions varied from well-mixed to thermally stratified, as shown in Fig. 2. In this March to May period the near-surface waters show a transition from a 100-m-deep surface mixed layer to a thermally stratified profile. Due to solar heating through the spring season the near-surface temperature increased by approximately 1.5°, while the salinity remained approximately constant. This behavior is consistent between both the specific 1996 CTD profiles (taken during the deployment and recovery cruises) and the long-term averages; however, both profiles in 1996 were approximately $0.5\,°\mathrm{C}$ warmer than the corresponding monthly averages (such temperature variability is not unusual for this location—see Whitney and Freeland, 1999). A similar 1.5° surface water temperature increase over this same period was observed at the weather buoy location; however, the weather buoy water temperatures were approximately $1.6\,°\mathrm{C}$ warmer than observed with two CTD profiles at OSP in 1996. The 2 March 1996 near-surface (upper 20 m) conditions yield a sound speed of $1472\,\mathrm{m\cdot s^{-1}}$, increasing to $1476\,\mathrm{m\cdot s^{-1}}$ by 20 May.

B. Inverted echo-sounder observations

The IES data consist of a continuous time series of volume scatter strength vs range profiles looking upward toward the ocean surface. These profiles (*records*) were averaged over 3 pings or 3 s in time, smoothing out the effects of short waves but leaving relatively large excursions (up to ± 2 m) in surface height due to longer waves. The first stage of processing involves extracting the range to the ocean surface, which is then used as reference for further analysis. Zedel (1994) discusses techniques for detecting the ocean surface using inverted echo-sounders. The simplest technique is to apply an amplitude threshold, which works adequately in situations with minimal bubble presence. However, under more typical conditions the relatively high gain and limited dynamic range of this prototype IES cause the near-surface echoes from the bubble plumes to exceed the A/D clipping

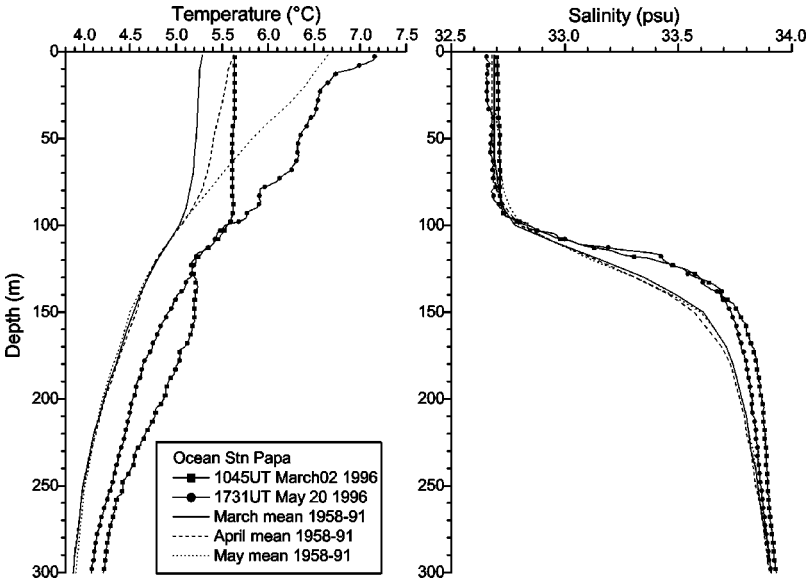


FIG. 2. Comparison of temperature and salinity vs depth profiles at Ocean Station Papa. Data for 1996 were taken during the IES deployment and recovery cruises. The climatological means were derived from weather-ship data (1958–86) and repeat oceanographic ship visits. Data courtesy Dr. H. Freeland at the Institute of Ocean Sciences.

threshold, triggering the simple threshold algorithm well below the actual surface. With this IES this problem occurred with S_v levels exceeding approximately -28 dB (re: 1 m^{-1}). A pragmatic solution in such cases was to make use of the *mirror-image* property of the combined surface and bubble plume echo, in essence averaging the range to the upper and lower boundaries of the total surface echo. This estimate was then corrected for the specular surface reflection of the transmitted pulse by subtracting the pulse length. This algorithm agreed with the threshold technique in the absence of bubbles, and was generally robust against high bubble densities and deep plumes. However, problems with the mirror-image algorithm arose in situations where the upper edge of the surface echo extended past the upper limit of data recording (53-m range). Fortunately, this only occurred in a few isolated events (1.1% of total period) where water currents pulled the IES (at the top of the mooring) down below 50-m depth. Data from these events were excluded from further analysis. Another problem observed occasionally was that extreme near-surface bubble densities induced acoustic masking of the surface and reflected bubble echoes (see Nützel *et al.*, 1994). The characteristic of such events was an anomalous drop in the mean surface height by several meters over a period of approximately 30 s. These surface-masking events occurred $<1\%$ of the period, and were excluded from the analysis through manual data inspection.

By remapping the volume scatter profile to align with the detected surface, longer surface wave displacements can be removed. This allows averaging of data records and display in depth vs time coordinates, as shown in Fig. 3. This 2-h echogram is typical of the stronger bubble plume penetration events observed with this IES, but it is not extreme. The echogram shows a large number of individual bubble *plumes* extending to depths of 5 to 20 m, separated by regions of lower backscatter (herein called *null zones*) which sometimes extend to within 1 to 2 m of the surface. The volume scatter strength generally decreases with increasing depth within each plume, with larger S_v levels (usually >-30 dB re: 1 m^{-1}) in the uppermost meter. Typically 10

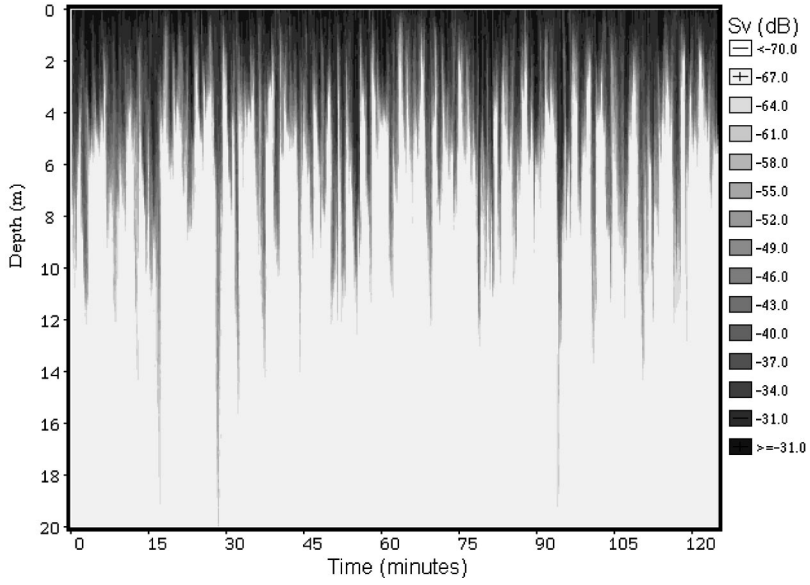


FIG. 3. Volume scatter strength vs depth and time echogram starting 0600UT 7 March 1996. Depth profiles are referenced to the detected surface height within each record (3 s), then averaged over five records. S_v data are not corrected for bubble extinction.

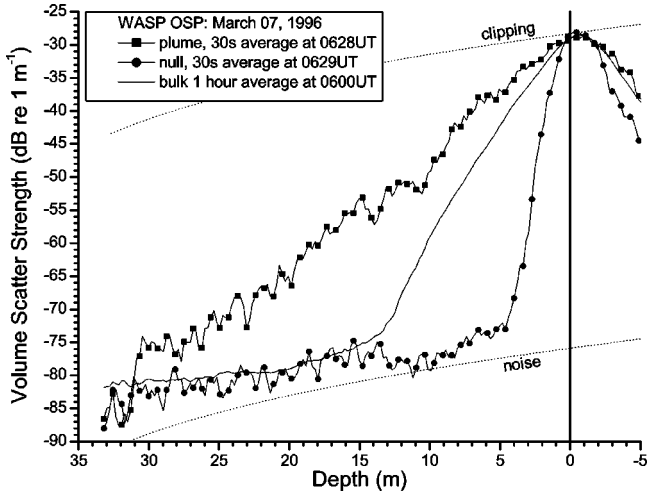


FIG. 4. Comparison of volume scatter strength versus depth profiles for a bubble plume and null zone near 0628UT 7 March (near minute 28 in Fig. 3). Plume and null zone profiles averaged over ten records, compared to 1-h (1200 record) average. Mean surface height is 43.2 m, used as origin of depth scale. Noise and A/D clipping limits indicated with dotted lines.

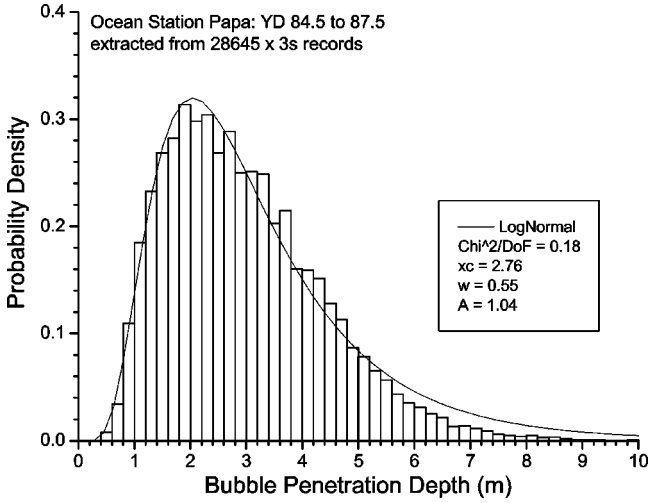


FIG. 5. Probability density of bubble penetration depth in 3-s records taken from a 3-day period starting 1200UT 24 March 1996. Distribution has mean = 2.96 m, standard deviation = 1.41 m, and maximum value = 10.55 m. Best-fit log-normal distribution plotted for comparison.

to 40 plumes per hour are observed, with duration in the beam of between 20 to 90 s. Note that the duration in beam is a function of both the horizontal size of the plumes and the horizontal near-surface current. Unfortunately, this simple IES had no current measuring capability, nor was there any other current meter in the surface mixed layer; thus, it is not possible to extract true horizontal plume dimensions. A rough estimate can be made by assuming a near-surface current of 20 $\mathrm{cm}\cdot\mathrm{s}^{-1}$ and assuming the plumes are much wider than the IES beam, in which case the horizontal dimension of these plumes is roughly 4 to 18 m.

Figure 4 examines a specific (yet not extraordinary) bubble plume event, comparing volume scatter strength profiles from the center of a plume and its adjacent null zone taken near minute 28 in Fig. 3. Note the symmetry of the three profiles around a point 37 cm above the specified surface. This is the mirror-image property referred to earlier. The null zone exhibits scattering only slightly above the noise threshold (showing sporadic echoes from meso-zooplankton) to within 5 m of the surface, where it rapidly increases to the clipping limit. Conversely, the plume profile shows a measurable scattering level to 30-m depth and an approximately linear increase (in decibels) towards the surface. The contrast in backscatter level between the plume and null zone profile reaches up to 40 dB. The linear variation in decibels is equivalent to an exponential variation in backscatter cross-section, i.e.,

$$s_v(z) = s_v(0)\exp[-z/d], \tag{3}$$

where z is depth (positive downwards) and d is the e-folding depth scale. The best-fit line to the plume profile in Fig. 4 has $s_v(0) = 1.15\times10^{-3}$ m^{-1} and $d = 2.54$ m, with the coefficient of correlation = 0.993. The 1-h averaged profile lies roughly midway between the plume and void zone profiles, and falls to near the noise threshold below 13-m depth. For $S_v > -60$ dB a linear fit to the 1-h averaged profile has $s_v(0) = 1.98\times10^{-3}$ m^{-1} and $d = 1.39$ m.

To quantify the *bubble penetration depth*, a threshold of $S_v = -50$ dB (re: 1 m^{-1}) was used to define the bottom of the bubble layer. This threshold effectively separated bubble and zooplankton scattering contributions, at the cost of truncating the more tenuous portions of the plumes. As will be shown in the next section, at scattering levels below -50 dB the acoustic extinction due to bubbles is minimal and the total void fraction ($<2\times10^{-9}$) is negligible. The bubble penetration depth and e-folding scale (fitted to s_v vs depth as above) were then calculated for each 3-s data record in the entire data set. From this database, a wide variety of statistics on the bubble plumes was extracted. The basic distribution of bubble penetration depths is shown in Fig. 5. These data were taken from a 3-day period where the bubble penetration depth statistics were reasonably constant. Specifically, the stationarity of this data period was examined by looking at trends and variance in bubble plume parameters (mean, standard deviation, and skewness of bubble penetration depth, and e-folding scale) within 2-h blocks. Trends in and variance of the parameters were all $<10\%$. The overall bubble penetration depth distribution is positively skewed (coefficient of skewness = 0.73) and is consistent with a log-normal distribution (reduced chi-squared = 0.18).

Overall, the bubble layer shows an interesting inverse relation between the bubble densities and e-folding depth scales, as shown by Fig. 6. In this case the extrapolated $s_v(0)$ is taken as a proxy for bubble density (as will be shown later, it is directly proportional). The extrapolated $s_v(0)$ values are truncated at 10 m^{-1}, as this corresponds to a void fraction near 2×10^{-3}, comparable to densities found under breaking waves. The figure shows that deeper plumes tend to be associated with low surface bubble densities, whereas the higher bubble densities have much shorter e-folding scales. Extreme bubble densities ($s_v(0) > 1.0$ m^{-1}) are rare, occurring in less than 1% of this data set, and they are generally associated with e-folding depth scales <0.6 m. The overall mean $s_v(0)$ and e-folding scales for this period are 0.23 m^{-1} and 0.60 m, respectively.

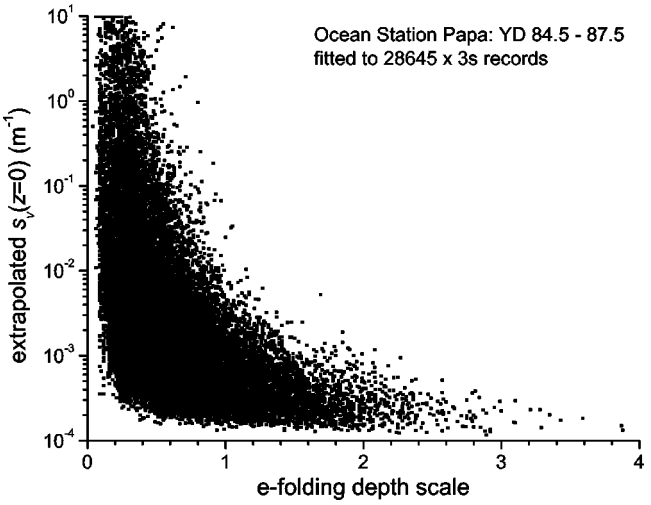


FIG. 6. Relation between extrapolated surface volumetric scattering cross section and e-folding depth scale using data from a 3-day period starting 1200UT 24 March 1996 (same period as Fig. 5).

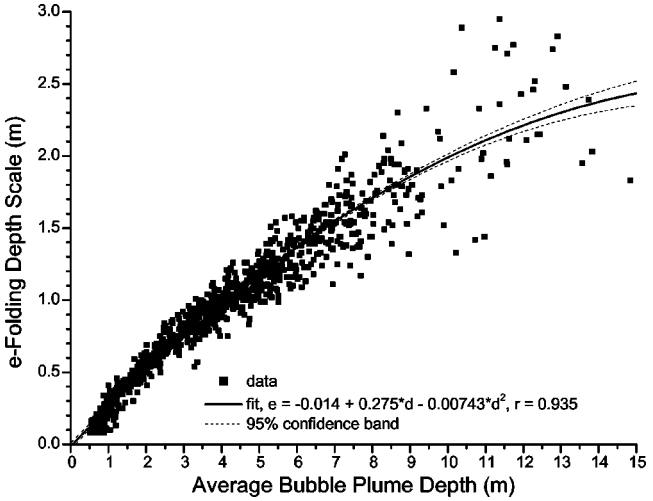


FIG. 8. Relation between averaged bubble plume duration and average bubble plume depth, using data from 936 2-h blocks spanning year days 61 to 141. Least squares fit parabola parameters shown in figure label, with correlation coefficient = 0.935.

For further analysis, the 81-day time series was broken into 2-h blocks. This period was chosen to give a significant number of bubble plumes per block yet be short relative to changes in surface wind and wave conditions. Individual bubble *plumes* were identified as those events where the bubble penetration depth exceeded the mean plus 1 standard deviation. For each plume so identified, the maximum penetration (hereafter called *plume depth*), duration, and e-folding scale at the time of maximum penetration were then extracted. Figure 7 shows the resulting time series of average plume depth, maximum plume depth, and average bubble penetration depth within each 2-h block spanning the entire data set. As indicated by these bubble plume measures, these data cover a variety of wind and wave conditions. There is some similarity between the bubble plume activity and the surface winds and waves shown in Fig. 1, for example with peaks near year days 95, 104, and 120 and calm periods during year days 82–92 and 121–131. But, there are

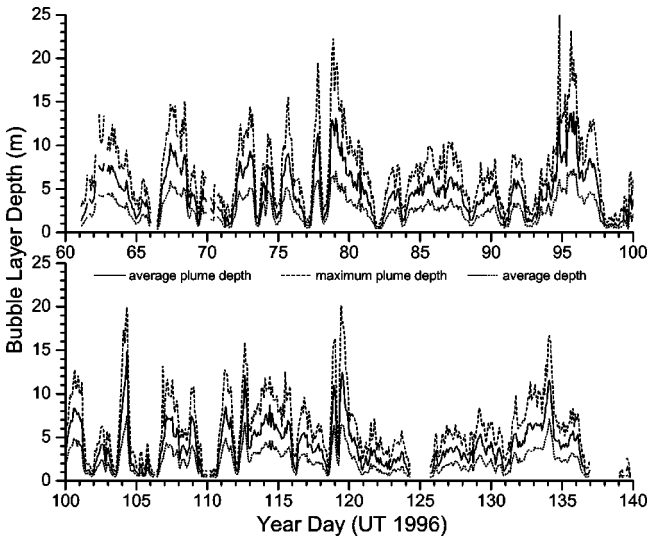


FIG. 7. Time series of average and maximum plume depths and average bubble penetration depth in 2-h blocks spanning year days 61 to 140 (1 March to 19 May), 1996.

significant differences, as should be expected for two locations 650 km apart. Also common to both the bubble plume and surface winds and wave data is the lack of any discernible seasonal trend.

Figure 7 shows that measurable bubble plume activity is present throughout the majority of this 81-day data set. Looking at the data as a whole, the averaged bubble penetration and plume depths were 2.44 and 6.72 m, respectively. *Significant* bubble plume activity, defined as where the average plume depth exceeded 3 m, occurred during 61% of this 81-day operating period. The average plume depth exceeded 10 m during 3.5% of this period. The maximum plume depth recorded was 25.0 m on year day 94.8 (3 April). Figure 7 also shows that there is a clear proportionality between the average bubble penetration depth and the average and maximum bubble *plume* depths over the entire data set. Additionally, the average and maximum bubble plume depths within all the 2-h blocks were highly correlated (correlation coefficient $r^2 = 0.97$). This proportionality between plume depth measures suggests that the log-normal distribution function (i.e. Fig. 5) is generally applicable over the entire observation period.

The e-folding depth scale within each bubble plume were also related to the average bubble plume depth, as shown in Fig. 8. The best (highest correlation coefficient)-fit relation between these variables was parabolic, with a reasonably good fit for average bubble plume depths up to roughly 8 m, but with increasing scatter for larger plume depths. These e-folding depth scales are similar to time-averaged values of 0.7 to 1.5 m reported by Crawford and Farmer (1987).

Finally, as shown in Fig. 9 the plume duration generally increased with plume depth, but there was much scatter. No reasonable polynomial fit to these data could be found. The *a priori* expectation was that duration would be related to the depth parameters for plumes of similar shape; i.e., that deeper plumes would also be wider. A major contributor to this lack of correlation is the variation in near-surface cur-

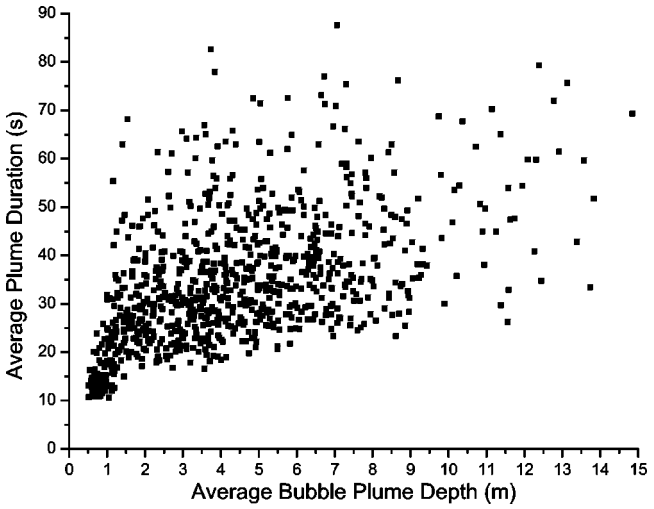


FIG. 9. Relation between averaged bubble plume duration and average bubble plume depth, using date from 936 2-h blocks spanning year days 61 to 141.

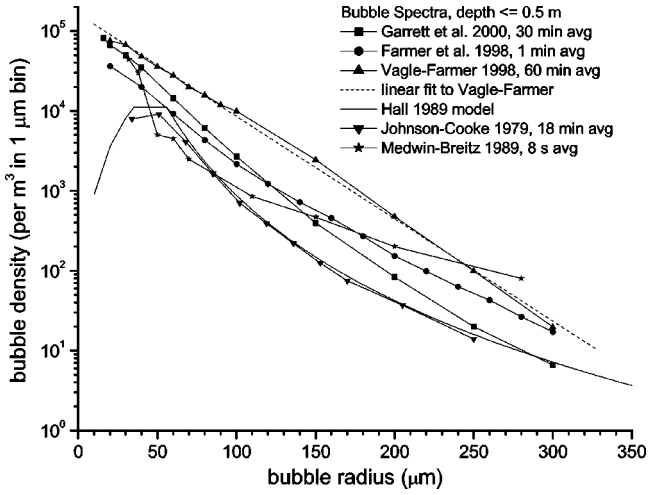


FIG. 10. Comparison of measured oceanic bubble size spectra for wind speeds near $12\,\mathrm{m\cdot s^{-1}}$, selected from or extrapolated to depths $\leqslant 0.5$ m. Averaging times as noted. Johnson and Cooke used a photographic technique, while the others are based on acoustic resonator devices. Hall's model is based on Johnson and Cooke data.

rent, which is expected to vary from 5 to $40\,\mathrm{cm\cdot s^{-1}}$ (Dahl and Jessup, 1995 observed $10\,\mathrm{cm\cdot s^{-1}}$, Thorpe and Hall, 1983 reported $30\,\mathrm{cm\cdot s^{-1}}$), but need not be related to the local surface wind and wave forcing that drives the bubble penetration. Without simultaneous measurements of the near-surface currents, little more can be concluded from the measured bubble plume duration.

IV. IMPLICATIONS FOR HIGH-FREQUENCY SONAR PERFORMANCE

To proceed from these 200-kHz IES backscatter measurements to more general sonar performance predictions, a form for the bubble-size distribution must be determined. Near-equilibrium oceanic microbubbles have a density vs size spectrum $n(a)$, where a is radius (normalized to 1-μm radius bin), typically spanning 10 to 400 μm in radius. Figure 10 shows a comparison of measured bubble-size spectra in the open ocean near 0.5-m depth at wind speeds of $12\,\mathrm{m\cdot s^{-1}}$. These size spectra decrease rapidly from peaks near 20- to 50-μm radius (Johnson and Cooke, 1979; Medwin and Breitz, 1989; Vagle and Farmer, 1992, 1998; Farmer *et al.*, 1998), although the exact form is highly variable and dependent on the interplay of processes such as bubble dissolution, buoyancy, turbulence, fractionation, and coalescence (Thorpe, 1982; Garrett *et al.*, 2000). Early work with photographic and acoustic techniques (reviewed by Wu, 1988) were matched (in a time-averaged sense) by a power-law dependence on bubble radius, with exponent near -4.4. However, other studies have shown a variety of different bubble spectral shapes under different conditions. For example, Medwin and Breitz (1989) found a power-law exponent of -2.7 for larger bubbles ($a>100$ μm) directly under (<25 cm below) breaking waves, suggestive that the breaking-wave *source spectrum* contains a greater proportion of large bubbles than longer period averages at greater depth reported by others. Breaking-wave events are highly transient and spatially localized, whereas for the deeper bubble plume penetration events under consideration here the size spectra have time to reach equilibrium. Recent measurements with broadband (5 to 200 kHz) acoustic resonator devices (Vagle and Farmer, 1998; Farmer *et al.*, 1998; Garrett *et al.*, 2000) show size-spectral variations closer to exponential in radius. For example, in Fig. 10 a linear regression (in log-density versus linear radius) to the Vagle–Farmer data from the open Pacific is well-fit ($r=-0.998$) by the relation

$$n(a)=1.6\times 10^5 \exp[-a/34\ \ \mu\mathrm{m}]$$
$$(re:\ 1\ \mu\mathrm{m\ bin})\ \ \text{for}\ \ a>20\ \ \mu\mathrm{m}. \qquad (4)$$

The acoustic resonator measurements generally show higher densities of smaller bubbles ($a<50$ μm), presumably highlighting measurement limitations in the older photographic techniques. These high densities of smaller bubbles have been confirmed with coincident inverted echo-sounder measurements at 200 and 400 kHz (Vagle and Farmer, 1998). Since these smaller bubbles are important to HF sonar performance, and the older power-law models are limited in accounting for smaller bubbles, the Vagle–Farmer model [i.e., of the form in Eq. (4)] will be used.

Most previous HF acoustic reverberation models have assumed a horizontally uniform bubble layer, with a vertical exponential depth scale and surface bubble density dependent only on wind speed (e.g., Hall, 1989; Dahl *et al.*, 1997; Dahl, 2003). However, the IES measurements discussed here clearly show that the bubble concentrations and e-folding depth scales are temporally and spatially variable. The model proposed here keeps the Farmer–Vagle size-spectral shape and an exponential decay with depth, but uses a surface bubble density, n_0, and e-folding depth, d, parameters extracted from the IES data on a per record (3-s) basis. The observed temporal variations in n_0 and d can then be converted to spatial variations by assuming a uniform horizontal current (i.e., equivalent to Taylor's hypothesis). Thus, the model must expressly allow for horizontal variations in bubble parameters and generally include variable source and target locations.

Using well-known models for bubble scattering (e.g., following Medwin and Clay, 1998), the *extinction* and *backscattering* cross section per unit volume, $s_e(f)$ and $s_v(f)$ (units of m^{-1}), can be calculated as integrals over the product of the bubble-size spectra with the acoustic scattering cross section for a single bubble, i.e.,

$$s_e(f,z)=\int_0^\infty \frac{4\pi a^2 \delta (k_R a)^{-1}}{[(f_R/f)^2-1]^2+\delta^2} n(a,z)da, \tag{5}$$

and

$$s_v(f,z)=\int_0^\infty \frac{a^2}{[(f_R/f)^2-1]^2+\delta^2} n(a,z)da, \tag{6}$$

where δ is the bubble total damping constant (dependent on a and f, see Medwin and Clay, 1998 for evaluation) and k_R is the acoustic wave number at the resonant frequency. The relationship between frequency and bubble size at resonance is given by

$$f_R=(2\pi a_R)^{-1}\sqrt{3\gamma P_0/\rho_w}, \tag{7}$$

where P_0 is the static pressure on the bubble, γ is the ratio of specific heats at constant pressure and volume of the enclosed gas ($=1.4$ for air), and ρ_w is the density of the surrounding water. For example, with a 200-kHz sonar at 1-m depth, the resonant bubble radius is 17 μm. For HF sonars both the resonant and nonresonant scattering regimes are important; thus, the limits for numerical integration of Eqs. (5) and (6) are taken as 1 and 500 μm. The extinction cross section determines the extra transmission loss, TL_{bub} (in dB), for acoustic signals propagating through a bubbly region, i.e.,

$$TL_{bub}(f,z)=4.34\int_{path} s_e(f,z)dl, \tag{8}$$

where l is the distance along the acoustic path. This loss term will be significant for the long, near-horizontal paths through the denser portions of the bubble plumes. As mentioned earlier, this extinction correction should have been applied to the measured $S_v(z)$ from the IES. However, this turned out to be unnecessary because of the A/D clipping limit and the much shorter vertical path lengths. For the bubble penetration threshold of $S_v=-50$ dB the extinction cross section lies near 4×10^{-4} m^{-1} (1.5×10^{-3} $dB\cdot m^{-1}$), which is less than the seawater absorption for an HF sonar.

In the case of a single frequency and with a specific bubble size-spectral shape, the integrals in Eqs. (5) and (6) can be evaluated to produce relations between $s_e(z)$, $s_v(z)$, and $n_0(z)$. For example, at 200 kHz the ratio of s_v to n_0 at the surface is 9.21×10^6 m^{-2} (*re*: 1 μm bin). Conversion of the 200-kHz IES data to n_0 then allows bubble backscatter and extinction calculations to be performed at other frequencies. Of direct relevance to HF sonar performance is that the ratio of $s_e(z)$ to $s_v(z)$ is both depth- and frequency dependent, as shown in Fig. 11. For 200 kHz this ratio lies in the range 25 to 40, and approaches 80 for frequencies near 50 kHz. Clearly, sonars operating at 200 kHz and higher will be less susceptible to bubble extinction losses than those near 50 kHz. A related parameter that can be derived from n_0 is the total void fraction, VF, defined

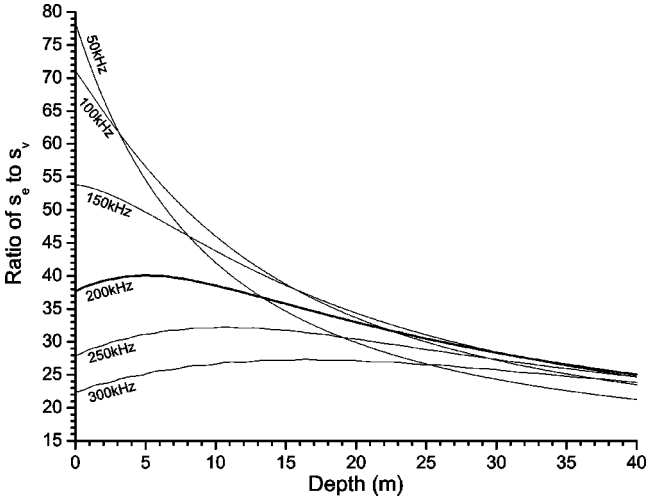


FIG. 11. Depth dependence of ratio between volumetric extinction and backscatter cross sections for bubbles as a function of sonar frequency. Calculations use Eqs. (5)–(7) assuming bubble size-spectral shape as given by Eq. (4).

$$VF=\int_0^\infty \frac{4}{3}\pi a^3 n(a)da. \tag{9}$$

Thus, through manipulations of Eqs. (4)–(6) and (9), measurements of $s_v(z)$ can be converted to VF for given bubble densities. For example, at 5-m depth, the bubble penetration threshold of -50 dB at 200 kHz corresponds to a VF of 2.2×10^{-9}, and the IES clipping limit of -28 dB corresponds to a VF of 3×10^{-7}. These VF are negligible compared to values of order 1 to 10% found under breaking waves (Lamarre and Melville, 1994), and nearly 2 orders of magnitude less than the upper VF limit of 1×10^{-5} recommended for sonar backscatter measurements (Vagle and Farmer, 1998).

These general expressions for bubble scattering and extinction can be adapted for the case of near horizontal sonar performance in near-surface bubble layer. The backscatter target strength, TS (dB *re*: 1 m^2), as a function of distance, x, for a vertical slice through the bubble layer is given by a depth integral of $s_v(f,x,z)$ multiplied by the insonified horizontal area, i.e.,

$$TS=10\log_{10}\left[\frac{1}{2}c\tau x\varphi\int_0^h s_v(f,x,z)dz\right], \tag{10}$$

where c is the sound speed, τ is the pulse length, φ is the transducer horizontal beamwidth (to -3 dB, in radians), and h is an effective bubble layer depth (the above assumes $x \gg h$ so that $r\approx x$). However, for near-horizontal paths the distance through the bubble layer is relatively long, so that the bubble extinction now becomes significant. In this model contributions from surface-reflected paths are assumed negligible because of the large bubble extinction losses encountered within the uppermost meter. The horizontal and vertical variations of s_v, s_e, and sonar beam pattern must be included together in the calculation of the total target strength

of the bubble layer at any given range. This is done by combining several terms of the sonar equation into a vertical integral of the form

$$\mathrm{TS}-\mathrm{TL}_{\mathrm{bub}}-\mathrm{BDL}(\theta)=10\log_{10}[\sigma],$$

where

$$\sigma=\frac{1}{2}c\tau x\varphi\int_0^{\infty}s_v(x,z)D(\theta)\exp\left[-2\int_{\mathrm{path}}s_e(x,z)dl\right]dz, \tag{11}$$

where $\mathrm{BDL}(\theta)$ is the two-way beam deviation loss vs angle θ from the transducer axis, evaluated as a function of range and depth given the sonar geometry. In the absence of particular beam-pattern measurements, for the purposes of this modeling study the vertical beam pattern is assumed to follow the standard line-source relation, i.e.,

$$\mathrm{BDL}(\theta)=10\log_{10}[D(\theta)],$$

$$\text{where}\quad D(\theta)=\left[\frac{\sin(\frac{1}{2}kL_T\sin(\theta))}{\frac{1}{2}kL_T\sin(\theta)}\right]^4, \tag{12}$$

where k is the acoustic wave number and L_T is the transducer vertical dimension. In Eq. (11) the bubble backscatter contribution will be larger near the surface, but this is balanced by much stronger extinction for the near-surface paths, such that significant reverberation contributions come from a region from the surface extending down 1 to 2 e-folding depths, depending on grazing angle. The sonar equation is then used to convert the *apparent* TS in Eq. (11) to received reverberation level, RL, i.e.,

$$\mathrm{RL}(r)=\mathrm{SL}+10\log_{10}[\sigma]-40\log_{10}[r]-2\cdot\alpha\cdot r, \tag{13}$$

where SL is the sonar source level.

An alternate formulation for bubble layer reverberation based on a semiempirical approach is advocated by Dahl *et al.* (1997) and Dahl (2003). This model features a horizontally uniform bubble layer, incoherent summation of direct and surface-reflected multipaths, a resonant bubble scattering approximation, and includes both backscatter and extinction effects. For this Dahl model the target strength and bubble transmission loss terms become

$$\mathrm{TS}+\mathrm{TL}_{\mathrm{bub}}=10\log_{10}\left[\frac{1}{2}c\tau x\varphi\frac{\sin\phi}{8\pi}\frac{\delta_r}{\delta_R}\times(1+8\beta\exp(-2\beta)-\exp(-4\beta))\right], \tag{14}$$

where $\beta=\beta_v/\sin\phi$, ϕ is the acoustic grazing angle, δ_r is the bubble damping due to reradiation at resonance ($=0.0136$), δ_R is the total damping constant at resonance, and the quantity β_v is related to the depth-integrated density of resonant-sized bubbles. Dahl *et al.* (1997) suggest an approximation $\delta_R=0.002\,55\cdot f^{1/3}$. Note that in the case of $\beta_v\geqslant 1$ and $\phi<30°$, this relation reduces to a $\sin\phi$ grazing-angle dependence and becomes independent of bubble density. This so-called *saturation effect* has been observed previously (Nützel *et al.*, 1994; Dahl *et al.*, 1997). The β_v parameter has been estimated from a large set of surface scattering data as a function of frequency (f, kHz) and wind speed at 10-m height (U, $\mathrm{m\cdot s^{-1}}$), with result

$$\log_{10}\beta_v=-6.45+0.47\cdot U+0.85\cdot\log_{10}[f]. \tag{15}$$

On the basis of comparisons with field data, this Dahl model is considered accurate between 10 and 100 kHz and grazing angles from approximately 2°–45° where bubble backscattering dominates (see Dahl, 2003 for a more detailed description of the source data).

To proceed further with the modeling some generic sonar operating parameters and geometry must be assigned. Perhaps the most demanding and structure-sensitive application is that of an HF obstacle avoidance sonar oriented near-horizontally to search for a near-surface object (perhaps a whale or floating mine). Clearly, the detailed sonar performance results will be dependent on the sonar, mounting geometry, and target parameters chosen. By choosing typical sonar parameters and applying them equally to both horizontally uniform and bubble plume cases, a realistic assessment of the importance of the bubble plume structures should emerge. The HF sonar is modeled with a 1° horizontal by 30° vertical beam pattern oriented horizontally from a depth of 5 m (as from the bow of a vessel). The transmit source level is taken as 210 dB (*re*: 1 μPa at 1 m) with a pulse length $\tau=500$ μs. The integrals in Eq. (11) were evaluated from 0 to 20-m depth with $dz=0.2$ m, and from 10 to 300-m horizontal range with $dx=1.0$. In this geometry the grazing angles vary from about 27° at close range to 1° at the maximum range.

Clearly, the modeled sonar performance will be dependent on the bubble parameters chosen. As a realistic first step bubble data were chosen from a period of *average* bubble activity (specifically a period where the bubble penetration and plume depths were close to the 81-day averages). Thus, data from the period 1730-1830UT 24 March (near year day 84.7) were chosen, which is a subset of the data used in Figs. 5 and 6. The horizontally uniform bubble *layer* case was generated using the surface bubble density and e-folding depth parameters extracted from the 1-h averaged profile. The spatially varying bubble *plume* case was generated using the measured n_0 and d values from individual 3-s records from the same period, converted to horizontal distance using a current magnitude of $20\ \mathrm{cm\cdot s^{-1}}$. Reverberation vs range curves from 20 separate bubble-plume realizations (''pings'') were calculated, each with start time advanced by 50 records (150 s) relative to the previous. Finally, a 20-ping reverberation average was calculated.

In all these models the complicating effects of surface waves and near-surface sound-speed stratification are purposely ignored. This is not because such effects are insignificant, but rather for the purposes of contrasting the effects of bubble spatial structures against horizontally uniform models, the inclusion of these effects would unnecessarily complicate the picture. Clearly, the focusing and shadowing effects due to surface waves will be significant, as shown by Norton *et al.* (1998); however, the corrections would be the same for both cases (uniform layer vs plumes). Under the moderate to strong wave-breaking and bubble injection conditions observed herein, thermo-haline stratification can be

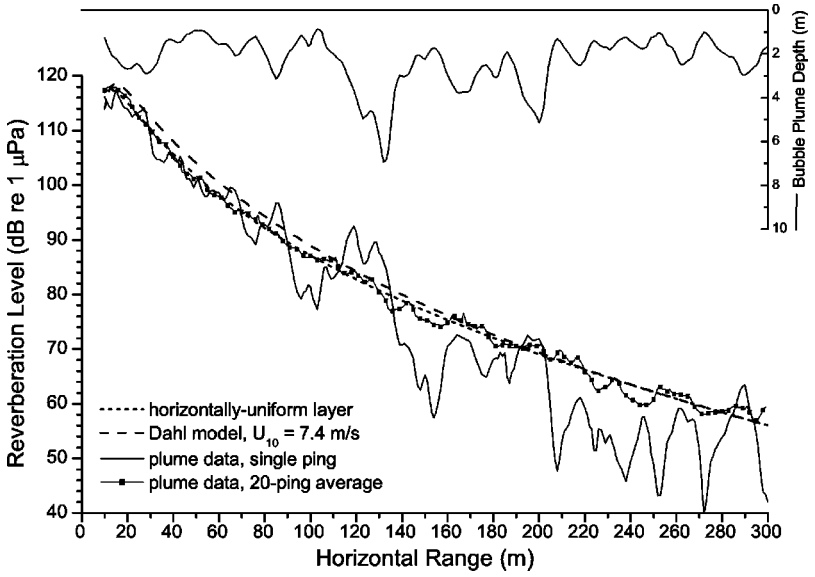


FIG. 12. Comparison between horizontally uniform bubble *layer* and *plume* reverberation models as a function of range. Models calculated for a 100-kHz sonar at 5-m depth using a horizontally oriented 1° horizontal by 30° vertical beam. Horizontally uniform layer result based on 1-h averaged profile 1730-1830UT 24 March, hence using $n_0 = 10\,680$ per m^3 and $d = 0.67$ m. Bubble plume model based on Eqs. (3)–(13) with n_0, d extracted from 3-s IES data records over the same time period, converted to horizontal range using current $= 20$ cm·s^{-1}. Upper panel shows bubble penetration depth data corresponding to the single ping. Dahl model [Eqs. (14) and (15)] evaluated using $U = 7.4$ m·s^{-1}.

ignored because the near-surface (upper 10 to 20 m) waters are typically well-mixed. However, the bubble plumes themselves have the potential to induce significant sound-speed changes. Using models outlined in Medwin and Clay (1998), the observed bubble densities (the parameter n_0 reached upwards of 1×10^6 m^{-3} in 1-μm bin) predict sound-speed anomalies at the ocean surface of up to $+50$ m·s^{-1} at 100 kHz, decreasing asymptotically at higher frequencies and decreasing exponentially with depth. Clearly, this will induce significant but localized downward refraction, which will greatly complicate predictions within the upper meter. However, these near-surface paths are subject to large bubble extinction losses and thus may only make minor contributions to the total scattering. In principle, the effects of surface waves and sound-speed anomalies could be incorporated into the above model by replacing straight-line paths with ray-tracing results.

Figure 12 shows the comparison of predicted reverberation levels at 100 kHz, along with the Dahl reverberation prediction using $U = 7.4$ m·s^{-1}($\beta_v = 0.054$ at 100 kHz). First, note the agreement between the Dahl prediction and the horizontally uniform layer case, lending confidence in the accuracy of the bubble reverberation model outlined herein. Additionally, the 20-ping averaged plume reverberation is in close agreement with the horizontally uniform prediction (based on a 1-h averaged profile), providing a useful consistency check. Thus, the importance of the bubble plumes boils down to significant differences between the *single-ping* and (temporally or spatially) averaged results. Figure 12 shows one arbitrarily chosen ping. Although the overall range dependence of the *single-ping* plume reverberation curve is similar to the horizontally uniform cases, there are large range variations that correlate strongly with the bubble plume structures. In particular, the modeled reverberation near plume maxima is typically 3 to 8 dB higher than predicted by the horizontally uniform models, whereas within the null zones, where there is strong acoustic shadowing, the plume-case reverberation drops by up to roughly 20 dB. Moreover, the excursions in reverberation level increase with horizontal range, due to the decreasing grazing angle allowing greater shadowing effects. These variations are important, as they imply the visibility of a target on a single-ping basis will be strongly dependent on its location relative to the bubble plumes. In many instances the bubble reverberation level is much better (lower) than predicted by horizontally uniform models; however, reverberation is only one part of sonar performance.

Figure 13 shows a comparison between the horizontally uniform and plume cases of the signal excess (i.e., target echo strength minus reverberation level) for an object with TS $= -15$ dB located at 1.0-m depth at the given horizontal range. Signal excess is a useful quantifier of target detectability, combining the effects of reverberation and bubble extinction loss along the path to the target. This TS is a generic value for objects like seamines, dense fish schools, or small whales. Clearly, although the overall range dependence is similar, the single-ping plume results show important range-dependent features at all three frequencies. For example, within the 7-m-deep plume and its accompanying shadow region near 120-m range, the signal excess drops to below zero for the 50- and 100-kHz results and drops by roughly 20 dB in the 200-kHz case. This dropout is due to high extinction loss within the plume, with (in this particular example) the one-way bubble extinction loss to a target at 130-m range (and 1-m depth) being 5.8, 25.0, and 71.9 dB for the 200-, 100-, and 50-kHz sonars, respectively. Conversely, at some ranges the signal excess is higher than predicted by the horizontally uniform model. For example, the horizontally uniform model indicates that a 50-kHz sonar would be ineffective beyond 120-m range; however (in this particular example), the individual ping prediction shows a window of detectability at 230–280 m. Clearly, the details of the signal-excess anomalies will change from ping-to-ping. The 200-kHz sonar is clearly less susceptible to the effects induced by bubble plume structures, which is a consequence of the relatively lower bubble extinction loss at higher frequencies (see Fig. 11).

Finally, the model calculations were repeated for more intense (but not extreme) oceanic bubble conditions, such as

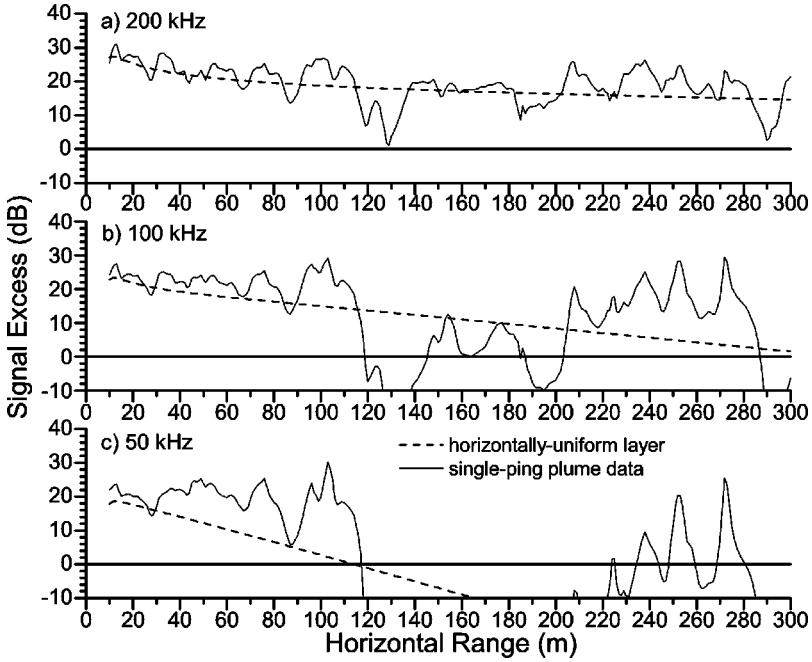


FIG. 13. Comparison of signal excess (echo strength minus reverberation level) versus range for a −15-dB (re: 1 m^2) target located at 1.0-m depth, at three different operating frequencies. Dashed lines are predictions for horizontally uniform layer and solid lines are for single-ping bubble plume case. Sonar geometry and bubble parameters are the same as in Fig. 12.

from the period examined in Figs. 3 and 4 (0600-0700UT 7 March 1996). In this case the average bubble penetration and plume depth (4.59 and 7.78 m, respectively) were roughly 1.8 times those in the previous example, with bubble plumes (on average) now extending below the sonar depth. A comparison between horizontally uniform and single-ping predictions at 100 and 200 kHz is shown in Fig. 14. Qualitatively the results are similar to the previous simulation, but with stronger bubble-plume shadowing dropouts in the reverberation curve (now up to roughly 50 dB in the 100-kHz reverberation) and stronger range-dependent anomalies in the signal-excess prediction. Note that the 200-kHz reverberation curve is similar in level and structure to that at 100 kHz, especially near the peaks, but the reverberation dropouts in plume shadows are less pronounced at 200 kHz. Although there are only minor differences in reverberation at the two frequencies, the 200-kHz sonar offers significantly greater detection range in both the single-ping and horizontally uniform predictions. However, on a single-ping basis there is still significant spatial variation in the signal excess, with both positive and negative anomalies.

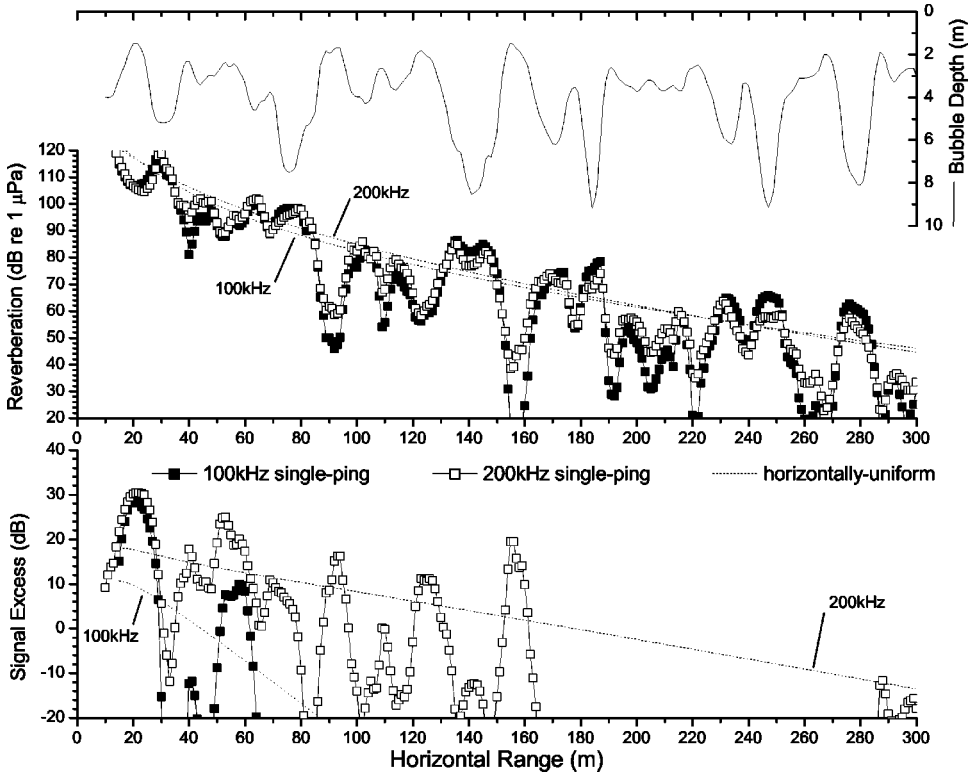


FIG. 14. Comparison between horizontally uniform and single-ping predictions of reverberation level and signal excess at 100 and 200 kHz, for bubble conditions found at 0600-0700UT 7 March 1996. Horizontally uniform model based on 1-h averaged $s_v(z)$ profile from Fig. 3, hence using $n_0 = 18\,230$ per m^3 and $d = 1.39$ m. Single-ping plume model based on Eqs. (3)–(13) with n_0, d extracted from 3-s IES data records within the same time period, converted to horizontal range using current = 20 cm $\cdot$ s^{-1}. Upper panel shows bubble penetration depth data corresponding to the single ping.

V. SUMMARY DISCUSSIONS

Bubble plumes were consistently observed with a 200-kHz IES in the near-surface ocean, with typical penetration depths from 2 to 15 m and maximum penetrations up to 25 m. Generally, the deeper portions of these bubble plumes were rather tenuous in terms of bubble densities, with corresponding volumetric backscatter strengths in the range -60 to -30 dB (re: 1 m^{-1}) and void fractions in the range 10^{-10} to 10^{-7}. However, the acoustic contrast between bubble plume and null zone HF backscatter at vertical incidence was observed to reach up to 40 dB. Within the near-surface bubble layers, including the plumes and at least up to medium bubble densities ($S_v \leqslant -28$ dB re: 1 m^{-1}, the limit of IES measurement), the bubble backscatter cross section was observed to decrease exponentially with depth, with e-folding scales from roughly 0.2 to 3.0 m. These depth parameters varied rapidly in time (and presumably in spatial dimension) as the bubble plumes moved in and out of the IES beam. The e-folding depth scale was observed to be highly correlated with the square of the bubble penetration depth, and inversely related to the surface bubble density. The observed bubble plume duration was in the range 20–90 s, generally increasing with average plume depth. However, no consistent relation between plume duration and bubble depth scales could be found as the duration was also controlled by near-surface currents, which were not measured in this field trial. Significant bubble penetrations, i.e., with average plume depth greater than 3 m, occurred over 61% of this late-winter to spring period. This depth is significant as it is below the reach of direct breaking-wave injection, which is generally confined to the upper meter (Lamarre and Melville, 1994), and thus other mechanisms (turbulence, convection, Langmuir circulation) must be responsible for this deeper advection. Under conditions with relatively constant bubble activity over a 3-day period, the penetration depths were shown to follow a log-normal distribution. Moreover, the consistency over the entire 81-day data set between measures such as average and maximum penetration depth and e-folding depth scale suggests that this distribution is more generally applicable.

This particular IES instrument and experimental setup had several shortcomings which placed restrictions on the data analysis. The relatively high gain and limited dynamic range (48 dB) of the echo-sounder resulted in A/D clipping of the denser portions of the bubble layer to $S_v(z) < -28$ dB (re: 1 m^{-1}). Using modern electronics this could be improved through the use of 12- or 16-bit A/D converter electronics. With more appropriate gain settings backscatter measurements up to $S_v(z) \approx -10$ dB (void fraction up to 10^{-5}) should be possible with improved resolution of the ocean surface echo. Furthermore, with modern digital storage hardware it should be possible to record individual 1-s ping data rather than 3-s averages, which would allow tighter measurement of shorter surface waves. One significant limitation in the overall data collection was the lack of surface wind and wave monitoring at the IES mooring location. In this particular 1996 deployment it was considered too difficult and expensive to install a surface buoy on the mooring. The lack of colocated surface meteorology prevented any analysis of the relationships between bubble penetration and parameters such as surface wind, wave height, and atmospheric boundary-layer stability. Additionally, the lack of any near-surface current measurement prevented any direct assessment of the bubble plume horizontal scales. This could be remedied in several ways, either with a current meter on a surface buoy, or through acoustic observation of the near-surface bubble advection using two orthogonal sidescan sonars (as in Thorpe and Hall, 1983) from the IES instrument package.

While similar observations of bubbles in plume-like structures in the near-surface ocean have been reported previously, the significance of these structures for HF sonar performance has been generally overlooked (at least in the open literature). Considerable work has been performed in the measurement and modeling of HF surface reverberation at low grazing angles (e.g., Nützel *et al.*, 1994; Dahl *et al.*, 1997; Dahl, 2003), with the general goal of predicting the (temporally or spatially) averaged reverberation level vs grazing angle and wind speed. In this study a simple model for bubble-induced reverberation using time-averaged bubble density profiles was able to match closely the previous model results (i.e., Fig. 12). This agreement is remarkable given the differences in modeling assumptions; specifically, the Dahl model used a resonant bubble approximation and a coherent summation of direct and surface-reflected paths, whereas Eq. (11) used a single direct path and included nonresonant bubble scattering contributions. Additionally, ping-averaged reverberation and target signal-excess predictions derived from actual bubble plume structures matched the predictions based on horizontally uniform bubble density profiles. This suggests that the *average* sonar performance can be realistically predicted using horizontally uniform models. However, it should be noted that active sonar detection occurs in the context of the actual, not time-averaged bubble spatial structures, and that the performance on any given ping will be significantly different from the spatially- or temporally averaged predictions.

A generic HF sonar and target geometry was devised that was both realistic and strongly sensitive to the presence of near-surface bubbles. When the observed bubble plume structures were used, important deviations in *single-ping* reverberation level and target signal excess were found. For example, with a forward-looking 100-kHz sonar at 5-m depth under average bubble conditions, the predicted variations in bubble plume reverberation level varied from $+8$ to -30 dB relative to the horizontally uniform predictions. The reverberation peaks were coincident with the bubble plumes, and reverberation dropouts occurred in shadow zones behind the plumes. Under stronger bubble penetration conditions (mean plume depths near 8 m), the reverberation dropouts were even stronger (exceeding 40 dB). As these reverberation curves provide a background for target detection, this suggests that target visibility, even for targets below the bubble plumes, will be sensitive to the location of the target relative to the bubble plume structures. When the effects of bubble extinction loss are included for near-surface targets, the situation becomes even more complicated. In the case of a target at 1-m depth, anomalies up to ± 30 dB were found in

single-ping target signal excess with respect to horizontally uniform predictions. In this context, a potentially fruitful avenue for further work is to assess the probabilities of detection and false alarm that might be induced by such bubble plume structures. The reader should note that this particular sonar geometry was chosen to highlight the impacts of bubble plume structures, and thus represents a somewhat extreme case. The anomalies in reverberation and signal excess diminish as the sonar and/or target move to greater depth.

Another important result from the modeling was improved performance in the presence of bubbles at higher sonar frequencies (e.g., 200–400 kHz). Both horizontally uniform and single-ping predictions showed higher signal excess to greater range for 200 kHz versus the 50- and 100-kHz results. For example under the average bubble conditions examined in Figs. 12 and 13, the predicted range for target detectability (based on a criterion of positive signal excess in the horizontally uniform models) was 115 m for the 50-kHz sonar as compared to 300 m or more for the 100- and 200-kHz sonars. The reason for this is a lower susceptibility to bubble-mediated extinction losses at frequencies above 100 kHz (see Fig. 11), accentuated at these small grazing angles by the relative long paths through the denser portions of the bubble plumes.

ACKNOWLEDGMENTS

The instrument development and moorings programs were supported during 1994–98 by Fisheries and Oceans Canada and the Canadian Joint Global Ocean Flux Study. Throughout this period the author is grateful for the guidance and support of David Farmer, as well as collaborations with colleagues Svein Vagle, Craig McNeil, and Nick Hall-Patch, all at the Institute of Ocean Sciences. The author also acknowledges the early work of Paul Johnston (then at IOS) for development and maintenance of the IES instrument.

Crawford, G., and Farmer, D. (**1987**). ''On the spatial distribution of ocean bubbles,'' J. Geophys. Res. **92**(C8), 8231–8243.

Dahl, P., and Jessup, A. (**1995**). ''On bubble clouds produced by breaking waves: An event analysis of ocean acoustic measurements,'' J. Geophys. Res. **100**(C3), 5007–5020.

Dahl, P., Nützel, B., Schmidt, A., Herwig, H., and Terray, E. (**1997**). ''Simultaneous acoustic and microwave backscattering from the sea surface,'' J. Acoust. Soc. Am. **101**(5), 2583–2595.

Dahl, P. (**2003**). ''The contribution of bubbles to high-frequency sea surface backscatter: A 24-h time series of field measurements,'' J. Acoust. Soc. Am. **113**(2), 769–780.

Farmer, D., McNeil, C., and Johnson, B. (**1993**). ''Evidence for importance of bubbles in increasing air–sea gas flux,'' Nature (London) **361**, 620–623.

Farmer, D., Vagle, S., and Booth, D. (**1998**). ''A free-flooding acoustical resonator for measurement of bubble size distributions,'' J. Atmos. Ocean. Technol. **15**, 1132–1146.

Francois, R., and Garrison, G. (**1982**). ''Sound absorption based on ocean measurements. II. Boric acid contribution and equation for total absorption,'' J. Acoust. Soc. Am. **72**, 1879–1890.

Garrett, C., Li, M., and Farmer, D. (**2000**). ''The connection between bubble size spectra and energy dissipation rates in the upper ocean,'' J. Phys. Oceanogr. **30**, 2163–2171.

Hall, M. (**1989**). ''A comprehensive model of wind-generated bubbles in the ocean and predictions of the effects on sound propagation at frequencies up to 40 kHz,'' J. Acoust. Soc. Am. **86**(3), 1103–1117.

Henyey, F. (**1991**). ''Acoustic scattering from ocean microbubble plumes in the 100-Hz to 2-kHz region,'' J. Acoust. Soc. Am. **90**, 399–405.

Johnson, B., and Cooke, R. (**1979**). ''Bubble populations and spectra in coastal waters: A photographic approach,'' J. Geophys. Res. **84**(C7), 3761–3766.

Lamarre, E., and Melville, W. K. (**1994**). ''Sound-speed measurements near the ocean surface,'' J. Acoust. Soc. Am. **96**(6), 3605–3616.

Lemon, D., Gower, J., Trevorrow, M., and Clarke, M. (**2001**). ''The acoustic water column profiler: A tool for long-term monitoring of zooplankton populations,'' Proceedings MTS/IEEE Oceans 2001 (Honolulu, HI, 5–8 Nov.), pp. 1904–1909.

McDonald, B. E. (**1991**). ''Echoes from vertically striated subresonant bubble clouds: A model for ocean surface reverberation,'' J. Acoust. Soc. Am. **89**(2), 617–622.

Medwin, H., and Breitz, N. (**1989**). ''Ambient and transient bubble spectral densities in quiescent seas and under spilling breakers,'' J. Geophys. Res. **94**(C9), 12751–12759.

Medwin, H., and Clay, C. (**1998**). *Fundamentals of Acoustical Oceanography* (Academic, San Diego).

Monahan, E., and Lu, M. (**1990**). ''Acoustically relevant bubble assemblages and their dependence on meteorological parameters,'' IEEE J. Ocean. Eng. **15**(4), 340–349.

Norton, G., Novarini, J., and Keiffer, R. (**1998**). ''Modeling the propagation from a horizontally directed high-frequency source in shallow water in the presence of bubble clouds and sea surface roughness,'' J. Acoust. Soc. Am. **103**(6), 3256–3267.

Novarini, J., and Bruno, D. (**1982**). ''Effects of the sub-surface bubble layer on sound propagation,'' J. Acoust. Soc. Am. **72**(2), 510–514.

Nützel, B., Herwig, H., Koenigs, P., and Monti, J. (**1994**). ''Acoustic backscatter measurements in the North Sea: 3–18 kHz,'' J. Acoust. Soc. Am. **95**(5), 2488–2494.

Thorpe, S. (**1982**). ''On the clouds of bubbles formed by breaking wind-waves in deep water and their role in air-sea gas transfer,'' Philos. Trans. R. Soc. London, Ser. A **304**, 155–210.

Thorpe, S., and Hall, A. (**1983**). ''The characteristics of breaking waves, bubble clouds, and near-surface currents observed using side-scan sonar,'' Cont. Shelf Res. **1**(4), 353–384.

Thorpe, S. (**1986**). ''Measurements with an automatically recording inverted echo-sounder, ARIES, and the bubble clouds,'' J. Phys. Oceanogr. **16**, 1462–1478.

Trevorrow, M. (**1995**). ''Measurement of ocean wave directional spectra using Doppler sidescan sonar arrays,'' J. Atmos. Ocean. Technol. **12**(3), 603–616.

Trevorrow, M. (**1998**). ''Observations of internal solitary waves near the Oregon coast with an inverted echo-sounder,'' J. Geophys. Res. **103**(C4), 7671–7680.

Vagle, S., and Farmer, D. (**1992**). ''The measurement of bubble-size distributions by acoustical backscatter,'' J. Atmos. Ocean. Technol. **9**(5), 630–644.

Vagle, S., Foote, K., Trevorrow, M., and Farmer, D. (**1996**). ''Absolute calibrations of monostatic echo-sounder systems for bubble counting,'' IEEE J. Ocean. Eng. **21**(3), 298–305.

Vagle, S., and Farmer, D. (**1998**). ''A comparison of four methods for bubble size and void fraction measurements,'' IEEE J. Ocean. Eng. **23**(3), 211–222.

Whitney, F., and Freeland, H. (**1999**). ''Variability in upper ocean water properties in the NE Pacific Ocean,'' Deep-Sea Res., Part II **46**, 2351–2370.

Wu, J. (**1988**). ''Bubbles in the near-surface ocean: A general description,'' J. Geophys. Res. **93**(C1), 587–590.

Zedel, L., and Farmer, D. (**1991**). ''Organized structures in subsurface bubble clouds: Langmuir circulation in the open ocean,'' J. Geophys. Res. **96**(C5), 8889–8900.

Zedel, L. (**1994**). ''Deep ocean wave measurements using a vertically oriented sonar,'' J. Atmos. Ocean. Technol. **11**(1), 182–191.

High-frequency acoustic scattering from turbulent oceanic microstructure: The importance of density fluctuations

Andone C. Lavery,[a] Raymond W. Schmitt, and Timothy K. Stanton
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 27 July 2002; accepted for publication 11 August 2003)

Acoustic scattering techniques provide a unique and powerful tool to remotely investigate the physical properties of the ocean interior over large spatial and temporal scales. With high-frequency acoustic scattering it is possible to probe physical processes that occur at the microstructure scale, spanning submillimeter to centimeter scale processes. An acoustic scattering model for turbulent oceanic microstructure is presented in which the current theory, which only accounts for fluctuations in the sound speed, has been extended to include fluctuations in the density as well. The inclusion of density fluctuations results in an expression for the scattering cross section per unit volume, σ_v, that is explicitly dependent on the scattering angle. By relating the variability in the density and sound speed to random fluctuations in oceanic temperature and salinity, σ_v has been expressed in terms of the temperature and salinity wave number spectra, and the temperature-salinity co-spectrum. A Batchelor spectrum for temperature and salinity, which depends on parameters such as the dissipation rates of turbulent kinetic energy and temperature variance, has been used to evaluate σ_v. Two models for the temperature-salinity co-spectrum have also been used. The predictions indicate that fluctuations in the density could be as important in determining backscattering as fluctuations in the sound speed. Using data obtained in the ocean with a high resolution vertical microstructure profiler, it is predicted that scattering from oceanic microstructure can be as strong as scattering from zooplankton. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1614258]

PACS numbers: 43.30.Ft, 43.30.Re, 43.30.Pc [WMC] Pages: 2685–2697

I. INTRODUCTION

High-frequency acoustic scattering techniques are commonly used to obtain highly informative, though often qualitative, images of the biological and physical processes that occur in the ocean interior. For example, these techniques have been applied, with varying success, to the assessment of the distribution of zooplankton and fish (Holliday and Pieper, 1995; Medwin and Clay, 1998). It is also common to observe high-frequency acoustic scattering images of internal waves (Proni and Apel, 1975; Haury *et al.*, 1979; Farmer and Smith, 1979; Sandstrom *et al.*, 1989; Trevorrow, 1998; Farmer and Armi, 1999; Orr *et al.*, 2000). However, the exact interpretation of the scattering images in terms of a physical or biological process remains poorly understood. It is possible that the scattering arises from turbulent microstructure created by breaking internal waves, biological organisms or other particulates acting as passive tracers of the different physical processes, or, as is most likely, a combination of these processes.

In this paper we focus on the scattering of sound from turbulent oceanic microstructure. There is a significant body of evidence suggesting that high-frequency acoustic scattering from oceanic microstructure can be as strong as that due to zooplankton (Munk and Garrett, 1973; Proni and Apel, 1975; Goodman and Kemp, 1981; Thorpe and Brubaker, 1983; Goodman, 1990; Seim *et al.*, 1995; Seim, 1999). This observation has led to an interest in the possibility of using acoustic scattering techniques to remotely infer physical properties of the scattering medium. Such measurements could be valuable, for example, for the remote determination of fundamental ocean mixing parameters. The smallest length scales before mixing is complete, referred to as the microstructure scale, are set by molecular diffusion. Large gradients in the media properties can occur at these submillimeter to centimeter scales with the potential of creating measurable acoustic scattering returns. The acoustic wave lengths of interest are commensurate with the scale set by the physical process, corresponding to high-frequency sound spanning tens of kHz to a few MHz.

There are many alternative techniques for acquiring information regarding the physical and biological processes that occur in the ocean interior. However, most of these techniques are limited by either the volume sampled or the spatial and temporal resolution of the measurements. For example, free-falling vertical microstructure probes have extremely high spatial resolution, and yet are inherently one-dimensional in nature. The primary advantage of using acoustic techniques is the possibility of synoptically imaging large volumes of the ocean interior without compromising on the high spatial resolution of the measurements. The primary challenge involves the interpretation of the received signal in terms of relevant physical and biological parameters.

Accurate scattering models are key to the interpretation of the received acoustic signals. With such models, though typically also in conjunction with supporting physical and biological information, it may be possible to distinguish between the complex scattering signatures characteristic of dif-

[a]Electronic mail: alavery@whoi.edu

ferent physical and biological processes (Stanton *et al.*, 1994; Trevorrow, 1998). Under sufficiently well-characterized environments it may even be possible to infer properties of the processes that give rise to scattering (Lhermitte and Lemmin, 1993; Goodman *et al.*, 1992; Stanton *et al.*, 1994; Menemenlis and Farmer, 1995; Oeschger and Goodman, 1995). Of most interest to the current investigation, if the contribution from turbulent oceanic microstructure can be isolated, it may be possible to use scattering models to determine parameters such as the dissipation rates of turbulent kinetic energy or temperature variance.

Over the years a number of investigations have predicted that volume scattering resulting from fluctuations in oceanic temperature (arising from sound speed fluctuations alone) can be significant (Kraichnan, 1953; Munk and Garrett, 1973; Proni and Apel, 1975; Goodman and Kemp, 1981; Goodman, 1990). These predictions are supported by highly suggestive, though infrequently conclusive, evidence (Thorpe and Brubaker, 1983). More recently, it has been predicted that the effects of salinity fluctuations (again arising from sound speed fluctuations alone) can also be important (Seim *et al.*, 1995; Seim, 1999), particularly at higher frequencies and at locations where salinity plays a significant role in determining the vertical density stratification. To date though, acoustic scattering models specifically developed for turbulent oceanic microstructure have only included the effects of sound speed fluctuations, ignoring the effects of density fluctuations. Yet temperature and salinity microstructure gives rise to small scale fluctuations in both the sound speed and the density.

In this paper we show that under some circumstances it is critical to include the effects of density fluctuations to accurately predict acoustic scattering by turbulent oceanic microstructure. It has been common practice to consider fluctuations in the medium density in both the fields of medical ultrasound and atmospheric turbulence. The theory behind these seemingly disparate fields is broadly similar, though there are differences in the details of the application of the theory to the different fields (Chernov, 1960; Tatarski, 1961; Morse and Ingard, 1968; Ishimaru, 1978; Goodman and Kemp, 1981; Waag, 1984). The theory is based on far-field weak scattering for which the Born approximation can be used to evaluate the scattering cross section per unit volume, σ_v. However, unlike for medical ultrasound and atmospheric turbulence, the contribution to scattering from oceanic microstructure due to fluctuations in the density has typically been neglected. We have included the density contribution to σ_v for oceanic microstructure and have found that it can be comparable to the contribution from sound speed fluctuations, leading to increases in σ_v as large as a factor of 4 (corresponding to a 6-dB increase in the volume scattering strength), under certain conditions. In addition, the inclusion of density fluctuations leads to an expression for σ_v which contains an explicit dependence on the scattering angle.

This paper is organized as follows. In Sec. II, we derive an expression for σ_v that includes contributions from both density and sound speed fluctuations. Then, in Sec. III, by assuming a universal Batchelor spectrum for temperature and salinity, two expressions for σ_v are derived. One expression uses an upper bound for the temperature-salinity co-spectrum (Bendat and Piersol, 1986), and represents the case of a perfect correlation between temperature and salinity fluctuations. In addition, we have developed an expression for the co-spectrum that is based on Stern's theory (1968). In Sec. IV, general predictions for scattering from turbulent oceanic microstructure are made and the range of model input parameters for which the density contribution is important are isolated. We also make acoustic scattering predictions based on microstructure data obtained using a high resolution vertical microstructure profiler (Schmitt *et al.*, 1988). From these data, all the model parameters necessary for making scattering predictions at relevant scales can be determined. Finally, in Sec. V, the results are summarized and recommendations are made as to the conditions under which it is important to include the density term in predictions of acoustic scattering from oceanic microstructure.

II. SCATTERING FROM TURBULENT OCEANIC MICROSTRUCTURE

A general expression for the scattering cross section per unit volume, σ_v, due to random fluctuations in the compressibility and density of a weakly scattering medium is derived in Sec. II A. The resulting expression for σ_v is not specific to oceanic microstructure, and alternative derivations can be found in a number of standard text books (Chernov, 1960; Tatarski, 1961; Morse and Ingard, 1968; Ishimaru, 1978). In order to facilitate the application of this formulation to oceanic microstructure, the expression for σ_v in terms of the medium compressibility and density is then expressed in terms of the medium density and sound speed, though the expression that is derived is still not specific to oceanic microstructure, but holds for any weakly scattering random medium. Assumptions specific to oceanic microstructure are made in Sec. II B, where the temperature and salinity dependence of density and sound speed are explicitly included in the expression for σ_v.

A. Weak scattering by random media

A sound wave traveling through a region (of volume V) containing turbulent microstructure will be scattered from the changes in the medium density and compressibility. Suppose the density, $\rho(\mathbf{r},t)$, and compressibility, $\kappa(\mathbf{r},t)$, in the region fluctuate randomly in space and time, deviating from the average values of the medium density, ρ_0, and compressibility, κ_0. From this point on, we assume that the density and compressibility do not change significantly during the time of the measurement, and therefore that $\rho(\mathbf{r})$ and $\kappa(\mathbf{r})$ are time independent. Any temporal changes that might occur are simply considered as different realizations of the real random fields $\rho(\mathbf{r})$ and $\kappa(\mathbf{r})$. For a weakly scattering medium in which the fluctuations in the compressibility and density are small, such as that produced by turbulent microstructure, the Born approximation can be used. In this case, there are well known solutions to the wave equation, and the far field scattered pressure wave is given by (Morse and Ingard, 1968, p. 411, Eq. 8.1.14)

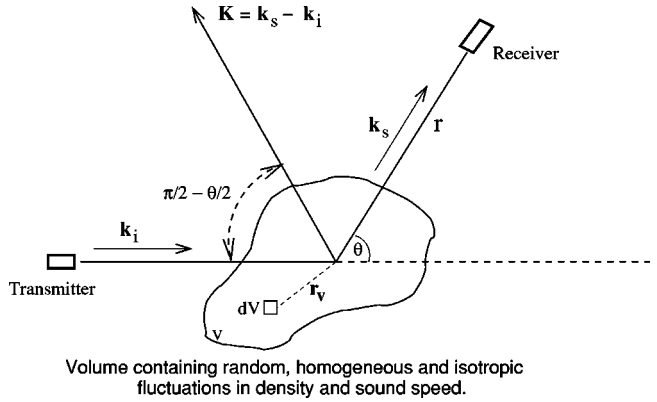


FIG. 1. Scattering geometry for an incident plane wave scattering from a bounded region of volume V containing random, stationary, homogeneous and isotropic fluctuations in the density and sound speed. $\mathbf{r}_v$ is the position vector of an infinitesimal volume element, $\hat{\mathbf{k}}_i$ is a unit vector along the direction of the incident wave, $\hat{\mathbf{k}}_s$ is a unit vector along the direction of the scattered wave, and $\mathbf{K}=k(\hat{\mathbf{k}}_s-\hat{\mathbf{k}}_i)$. The angle between the incident and scattered wave vectors is given by θ $(\hat{\mathbf{k}}_i\cdot\hat{\mathbf{k}}_s=\cos\theta)$, and $K=|\mathbf{K}|=2k\sin(\theta/2)$. K is maximum at backscattering where $\theta=\pi$.

$$p_s(r)=P_0\frac{e^{ikr}}{r}f, \tag{1}$$

where k is the incident acoustic wave number ($=2\pi/\lambda$, where λ is the acoustic wave length), r is the range from the scattering volume to the receiver, and P_0 is the incident acoustic pressure at some known reference range. The scattering amplitude, f, is a measure of the efficiency with which sound is scattered and is given by (Morse and Ingard, 1968, p. 413, Eq. 8.1.20):

$$f=\frac{k^2}{4\pi}\int_V \gamma(\mathbf{r}_v)e^{-i\mathbf{K}\cdot\mathbf{r}_v}d\mathbf{r}_v\,, \tag{2}$$

where

$$\gamma(\mathbf{r}_v)=\gamma_\kappa(\mathbf{r}_v)+\gamma_\rho(\mathbf{r}_v)\hat{\mathbf{k}}_i\cdot\hat{\mathbf{k}}_s=\gamma_\kappa(\mathbf{r}_v)+\gamma_\rho(\mathbf{r}_v)\cos\theta, \tag{3}$$

$\mathbf{r}_v$ is the position vector of any volume element relative to some arbitrary center point chosen within the volume V, $\hat{\mathbf{k}}_i$ is a unit vector along the direction of the incident wave, $\hat{\mathbf{k}}_s$ is a unit vector along the direction of the scattered wave, and $\mathbf{K}=k(\hat{\mathbf{k}}_s-\hat{\mathbf{k}}_i)$ (Fig. 1). The angle between the incident and scattered wave vectors is given by θ $(\hat{\mathbf{k}}_i\cdot\hat{\mathbf{k}}_s=\cos\theta)$, and $K=|\mathbf{K}|=2k\sin(\theta/2)$. The wave vector $\mathbf{K}$ is often referred to in the literature as the Bragg wave vector (Goodman, 1990). K is maximum at backscattering where $\theta=\pi$. The terms γ_κ and γ_ρ describe the variations in the medium compressibility and density and are given by (Morse and Ingard, 1968, p. 409, Eq. 8.1.11)

$$\gamma_\kappa=\frac{\kappa-\kappa_0}{\kappa_0}=\frac{c_0^2}{c^2}\frac{\rho_0}{\rho}-1 \quad\text{and}\quad \gamma_\rho=\frac{\rho-\rho_0}{\rho}=1-\frac{\rho_0}{\rho}, \tag{4}$$

where $c^2=1/\kappa\rho$ and c_0 is the mean sound speed. The contribution to scattering from fluid velocity has been ignored in the derivation of Eq. (2). This contribution has been investigated by others and, for oceanic microstructure, can be shown to be smaller than the other contributions by various orders of magnitude (Goodman, 1990). In fact, for backscattering, fluid velocity fluctuations do not contribute to scattering at all. It has also been assumed that the scattering volume is bounded and in the far field of both the transmitter and the receiver and that the transmitter and receiver are in the far field of the scattering volume.

The solution to the wave equation for bounded weakly scattering targets given by Eq. (2) has been used on many occasions as a starting point for developing scattering models for random media variability; it has been used for scattering of electromagnetic waves from atmospheric turbulence (Chernov, 1960; Tatarski, 1961; Morse and Ingard, 1968; Ishimaru, 1978), scattering of ultrasonic waves from different tissues in the field of medical ultrasound (Waag, 1984; Waag *et al.*, 1985, 1989), and it has also been used previously for scattering of high-frequency sound from temperature microstructure in the ocean (Munk and Garrett, 1973; Goodman and Kemp, 1981; Goodman, 1990). In addition, it has been used as a starting point for a number of acoustic scattering models for weakly scattering zooplankton (McGehee *et al.*, 1998; Stanton *et al.*, 1998; Chu and Ye, 1999; Lavery *et al.*, 2002). In contrast to the models developed for individual zooplankton, scattering from random media fluctuations is typically described in terms of the statistical properties of the medium. Thus, the scattering cross section per unit volume, σ_v, with units of inverse length, is given by (Ishimaru, 1978, p. 332, Eq. 16-10a)

$$\begin{aligned}\sigma_v&=\frac{r^2\langle p_s p_s^*\rangle}{VP_0^2}\\&=\frac{1}{V}\langle ff^*\rangle\\&=\frac{1}{V}\frac{k^4}{2^4\pi^2}\int_V\int_V B_\gamma(\mathbf{r}_1,\mathbf{r}_2)e^{-i\mathbf{K}\cdot(\mathbf{r}_1-\mathbf{r}_2)}d\mathbf{r}_1 d\mathbf{r}_2\,,\end{aligned} \tag{5}$$

where $\langle\ldots\rangle$ represents an ensemble average, and $\mathbf{r}_1$ and $\mathbf{r}_2$ are integration vectors over the volume V. The term $B_\gamma(\mathbf{r}_1,\mathbf{r}_2)\equiv\langle\gamma(\mathbf{r}_1)\gamma^*(\mathbf{r}_2)\rangle=\langle\gamma(\mathbf{r}_1)\gamma(\mathbf{r}_2)\rangle$ is the spatial correlation function of the real random field, γ, that describes the physical properties of the medium in the volume V. Note that the term volume scattering coefficient (denoted s_v) is more commonly used in the literature (instead of σ_v) when referring to an aggregation of discrete scatterers (Urick, 1980; Medwin and Clay, 1998).

Following the bulk of the published literature on scattering from turbulent microstructure we assume that the medium properties are homogeneous, meaning that the mean value of γ is constant and that the correlation function, $B_\gamma(\mathbf{r}_1,\mathbf{r}_2)$, does not change when the pair of points $(\mathbf{r}_1,\mathbf{r}_2)$ is displaced by the same amount in the same direction (Tatarski, 1961, p. 15). This latter assumption requires some justification since vertical stratification in the ocean is prevalent. However, the vertical stratification can be overturned by sufficiently large velocity shears. Typically, once overturning has occurred, physical structures that are aligned vertically or horizontally are destroyed, and the medium becomes turbulent, resulting in random homogeneous fluctuations in the

medium properties, at least locally within the scattering resolution volume. In the ocean interior the principal cause of mechanical mixing, or the source of high velocity shear resulting in overturning events and turbulent microstructure, is the presence of internal waves, which is where the assumptions we make here are most likely valid. Consequently, with the assumption that the fluctuations are locally homogeneous: $B_\gamma(\mathbf{r}_1,\mathbf{r}_2)=B_\gamma(\mathbf{r}_1-\mathbf{r}_2,0)=B_\gamma(\mathbf{r})$, where the substitution $\mathbf{r}=\mathbf{r}_1-\mathbf{r}_2$ has been made. Letting $\mathbf{x}=(\mathbf{r}_1+\mathbf{r}_2)/2$, and performing the integral over $\mathbf{x}$,

$$\sigma_v=\frac{k^4}{2^4\pi^2 V}\int_V\int_V B_\gamma(\mathbf{r})e^{-i\mathbf{K}\cdot\mathbf{r}}d\mathbf{r}d\mathbf{x}$$
$$=\frac{(\pi/2)\,k^4}{(2\pi)^3}\int_V B_\gamma(\mathbf{r})e^{-i\mathbf{K}\cdot\mathbf{r}}d\mathbf{r}. \tag{6}$$

For wave front curvature effects to be unimportant (Goodman, 1990; Waag *et al.*, 1985), there must be many characteristic length scales of the media variability encompassed within the first Fresnel radius, R_F. For a point transmitter and receiver, the first Fresnel radius refers to the locus of points for which the phase difference that arises from the path length difference between this locus of points and the center of the volume V is equal to $\pi/2$. At backscattering $R_F=\sqrt{R\lambda/2}$, where R is the range from the transmitter to the center of the volume V [an expression for R_F for forward scattering is given in Flatte *et al.* (1979, p. 91, Eq. 6.2.2)]. Next, we assume that the volume V is sufficiently large that many correlation lengths of the random media properties are encompassed by the volume. These assumptions allow the integral over the volume V in Eq. (6) to be converted to an integral over all space:

$$\sigma_v=\frac{(\pi/2)\,k^4}{(2\pi)^3}\int B_\gamma(\mathbf{r})e^{-i\mathbf{K}\cdot\mathbf{r}}d\mathbf{r}=\frac{\pi}{2}k^4\Phi(\mathbf{K}). \tag{7}$$

$\Phi(\mathbf{K})$ is the Fourier transform of the spatial correlation function of the medium variability, or simply the 3D wave number spectrum of the medium variability evaluated at $\mathbf{K}$. If we further assume that the medium is isotropic, then $B_\gamma(\mathbf{r})=B_\gamma(r)$ and $\Phi(\mathbf{K})=\Phi(K)$. The effects of anisotropy on scattering have been considered by Goodman (1990), who included a vertical wave number scaling factor, α (the ratio of vertical to horizontal spatial scale size), which resulted in an increase (α is typically <1) in the predicted scattering by a factor of α^{-2}. If the degree of anisotropy is small, the scattering cross section per unit volume is given by

$$\sigma_v=\frac{\pi}{2}k^4\Phi(K). \tag{8}$$

As mentioned earlier, a similar expression for σ_v can be found in a number of standard text books and has been used widely in the literature. However, when applied specifically to oceanic microstructure, the expression for B_γ inside the integral is typically given in terms of sound speed fluctuations alone, and the contribution from density fluctuations is ignored.

In order to facilitate the application of this formalism to the specific case of oceanic microstructure, we proceed to obtain an expression for σ_v which is explicitly written in terms of both density and sound speed fluctuations. This requires the evaluation of the spatial correlation function B_γ in terms of sound speed and density fluctuations. Since we are considering a weakly scattering medium, $\rho=\rho_0+\rho'$ and $c=c_0+c'$, where the inherent variations in the medium density, ρ', and sound speed, c', satisfy $c'/c_0\ll 1$ and $\rho'/\rho_0\ll 1$. From Eq. (4),

$$\gamma_\kappa=\left(1+\frac{c'}{c_0}\right)^{-2}\left(1+\frac{\rho'}{\rho_0}\right)^{-1}-1\approx -2\frac{c'}{c_0}-\frac{\rho'}{\rho_0}$$

and

$$\gamma_\rho=\frac{\rho'}{\rho_0}\left(1+\frac{\rho'}{\rho_0}\right)^{-1}\approx\frac{\rho'}{\rho_0}, \tag{9}$$

where only first order terms in the primed variables have been kept. Making use of these relationships, the trigonometric relationship $(1-\cos\theta)=2\sin^2(\theta/2)$, as well as the definition of γ [Eq. (3)], the spatial correlation function of media variability becomes

$$B_\gamma(r)=4\left(\left\langle\frac{c'}{c_0}\frac{c'^*}{c_0}\right\rangle+2\sin^2\left(\frac{\theta}{2}\right)\left\langle\frac{c'}{c_0}\frac{\rho'^*}{\rho_0}\right\rangle\right.$$
$$\left.+\sin^4\left(\frac{\theta}{2}\right)\left\langle\frac{\rho'}{\rho_0}\frac{\rho'^*}{\rho_0}\right\rangle\right). \tag{10}$$

Using this expression for B_γ, σ_v [Eq. (8)] becomes

$$\sigma_v=2\pi k^4\left(\Phi_c(K)+2\sin^2\left(\frac{\theta}{2}\right)\Phi_{c\rho}(K)\right.$$
$$\left.+\sin^4\left(\frac{\theta}{2}\right)\Phi_\rho(K)\right), \tag{11}$$

where $\Phi_c(K)$, $\Phi_\rho(K)$, and $\Phi_{c\rho}(K)$ are the 3D wave number spectra of the sound speed fluctuations, density fluctuations, and the correlation between sound speed and density fluctuations. In arriving at this expression, it should be noted that all cross-terms between density and sound speed fluctuations have been included.

No assumptions specific to scattering from oceanic microstructure have been made in deriving this general expression, which holds (theoretically) for any weakly scattering random medium. To date, various equivalent forms of Eq. (11) have been used to describe scattering by atmospheric turbulence and organ tissues. However, all formulations specific to scattering from oceanic microstructure have only included fluctuations in the sound speed, resulting in simply the first term in this expression. The inclusion of density fluctuations results in two extra terms, both of which have an explicit angular dependence. It is clear from the wave equation where this angular dependence arises since the term involving density depends on the *gradient* of the density. We turn now to the specific case of scattering from turbulent oceanic microstructure.

B. Application to turbulent oceanic microstructure

The most convenient parameters for describing scattering from turbulent oceanic microstructure are temperature

TABLE I. The following parameters were used for the prediction of acoustic scattering from oceanic microstructure (Flatte *et al.*, 1979: p. 5, Eq. 1.1.6): $a=3.19\times10^{-3}$ ($^\circ C^{-1}$), $b=0.96\times10^{-3}$ (psu^{-1}), $\alpha=0.13\times10^{-3}$ ($^\circ C^{-1}$), $\beta=0.8\times10^{-3}$ (psu^{-1}), $\nu=0.15\times10^{-5}$ ($m^2\,s^{-1}$), $D_T=1.38\times10^{-7}$ ($m^2\,s^{-1}$), $D_S=0.95\times10^{-9}$ ($m^2\,s^{-1}$), and $L_e=145.26$. Modest values of ϵ (1×10^{-8} W kg^{-1}) and χ_T ($1\times10^{-6}\,^\circ C^2\,s^{-1}$) were chosen, giving values of $k^*=29$ (cpm), $k_d=769$ (cpm), and $k_{ds}=9207$ (cpm). Acoustic frequencies ranging from approximately 10 kHz ($k=42$ cpm, $\lambda=15$ cm) to 2 MHz ($k=8379$ cpm, $\lambda=0.75$ mm) were used to make the scattering predictions. Based on these parameters, most of the predictions lie in the viscous-convective range, where $k>k^*$. The following three cases were investigated (all parameters dimensionless unless otherwise stated). Note that the values for $R_{\rho c}$ in this table were evaluated at backscattering ($\theta=\pi$).

Case no.	$R_{\rho c}$	$R_{\rho c}^{-1}$	R_c	R_ρ	$\delta(^\circ C\,psu^{-1})$	$R_{\rho c}/L_e$ ($\times10^{-2}$)	$\frac{\max(\sigma_v^S)}{\max(\sigma_v^T)}$ [Eq. (34)]
1	1	1	1.911	0.094	0.575	0.6	10
2	3.477	0.288	6.646	0.325	2	2.39	0.72
3	0.165	0.287	0.548	0.0268	3.49	0.20	367.3

and salinity, since these are the most commonly measured, and consequently better mapped and understood oceanographic parameters. The aim of this section is to express Eq. (11) in terms of the wave number spectra of temperature and salinity, instead of density and sound speed. Fluctuations in the sound speed and density are related to the small scale fluctuations in temperature, T', and salinity, S', by

$$\frac{c'}{c_0}=\frac{1}{c_0}\frac{\partial c}{\partial T}T'+\frac{1}{c_0}\frac{\partial c}{\partial S}S'=aT'+bS', \tag{12}$$

where

$$a\equiv\frac{1}{c_0}\frac{\partial c}{\partial T}\quad\text{and}\quad b\equiv\frac{1}{c_0}\frac{\partial c}{\partial S}, \tag{13}$$

and

$$\frac{\rho'}{\rho_0}=\frac{1}{\rho_0}\frac{\partial \rho}{\partial T}T'+\frac{1}{\rho_0}\frac{\partial \rho}{\partial S}S'=-\alpha T'+\beta S', \tag{14}$$

where

$$\alpha\equiv-\frac{1}{\rho_0}\frac{\partial \rho}{\partial T}\quad\text{and}\quad \beta\equiv\frac{1}{\rho_0}\frac{\partial \rho}{\partial S}. \tag{15}$$

Here α is the coefficient of thermal expansion and β is the coefficient for saline contraction. The importance of salinity versus temperature in determining scattering from sound speed and density variations can be gauged by examining the vertical changes in sound speed and density,

$$\frac{1}{c_0}\frac{dc}{dz}=a\frac{\partial T}{\partial z}+b\frac{\partial S}{\partial z}\quad\text{and}\quad\frac{1}{\rho_0}\frac{d\rho}{dz}=-\alpha\frac{\partial T}{\partial z}+\beta\frac{\partial S}{\partial z}, \tag{16}$$

and forming the sound speed and density ratios:

$$R_c\equiv\frac{a}{b}\delta\quad\text{and}\quad R_\rho\equiv\frac{\alpha}{\beta}\delta,\quad\text{where}\quad\delta=\frac{\partial T/\partial z}{\partial S/\partial z}. \tag{17}$$

Here R_c and R_ρ indicate the relative importance of temperature versus salinity in determining the vertical sound speed and density gradients, and play a critical role in determining the contribution to scattering from salinity relative to temperature. If $-1<R_c<1$, then salinity plays a more dominant role than temperature in determining vertical sound speed variations. If $-1<R_\rho<1$, salinity plays a more dominant role than temperature in determining vertical density variations. Typical values of these parameters for an open ocean environment are given in Table I. From these values, $R_c=3.32\ \delta$ and $R_\rho=0.16\ \delta$. There is typically larger variability in the parameter δ than in a, b, α, or β. However, it can be seen that the range of values of δ ($-6.25<\delta<6.25$) for which $-1<R_\rho<1$ is significantly wider than the range of values of δ ($-0.30<\delta<0.30$) for which $-1<R_c<1$. In addition, in order to assess the importance of including fluctuations in the density when calculating acoustic scattering from turbulent oceanic microstructure, it is necessary to compare the fractional change in the sound speed from salinity, b, and the coefficient of haline contraction, β. Since these terms are approximately equal, we conclude that it is important to include the contribution to scattering from density fluctuations.

In order to evaluate σ_v in terms of temperature and salinity fluctuations, it is necessary to evaluate B_γ in terms of T' and S'. Making use of Eqs. (10), (12), and (14),

$$B_\gamma(r)=4[A^2\langle T'T'^*\rangle+B^2\langle S'S'^*\rangle+2AB\langle T'S'^*\rangle], \tag{18}$$

where

$$A=a-\alpha\sin^2\left(\frac{\theta}{2}\right)\quad\text{and}\quad B=b+\beta\sin^2\left(\frac{\theta}{2}\right). \tag{19}$$

Using this expression for B_γ, σ_v becomes

$$\sigma_v=2\pi k^4(A^2\Phi_T(K)+B^2\Phi_S(K)+2AB\Phi_{ST}(K)), \tag{20}$$

where $\Phi_T(K)$, $\Phi_S(K)$, and $\Phi_{ST}(K)$ are the 3D wave number spectra of temperature, salinity, and the temperature-salinity co-spectrum, all evaluated at the wave number K.

This is a general expression for the scattering cross section per unit volume describing scattering from stationary, homogeneous, and isotropic turbulent oceanic microstructure. Unlike previous formulations of scattering specific to turbulent oceanic microstructure, which only included fluctuations in the sound speed, the scattering cross section per unit volume derived here also includes the contributions from variability in the density. It has typically been assumed (Tatarski, 1961; Goodman and Kemp, 1981; Thorpe and Brubaker, 1983; Goodman, 1990; Seim *et al.*, 1995; Seim, 1999) that scattering from oceanic microstructure is dominated by sound speed fluctuations, that is, $\rho'/\rho_0\ll c'/c_0$. In

this case it follows from Eq. (9) that $\gamma_\kappa \approx -2c'/c_0 = -2aT'$ and $\gamma_\rho \approx 0$. If the effects of salinity are ignored, this is a reasonable assumption since for typical open ocean environments $\alpha/a \approx 0.03$. As a result, $\sigma_v = 2\pi k^4 a^2 \Phi_T(K)$, which is the first term of Eq. (20) with the coefficient A replaced by a. Seim *et al.* (1995) developed this theory one step further, and included the contributions to scattering from fluctuations in the sound speed originating from both temperature and salinity microstructure. In this scenario, $\gamma_\rho = \rho'/\rho_0 \approx 0$, $\gamma_\kappa \approx -2c'/c_0 = -2(aT' + bS')$, and the scattering cross section per unit volume is given by

$$\sigma_v = 2\pi k^4 (a^2 \Phi_T(K) + b^2 \Phi_S(K) + 2ab\Phi_{ST}(K)). \quad (21)$$

This expression is identical in form to Eq. (20) with the coefficient A replaced by a, and B replaced by b. However, the coefficients A and B contain terms involving the coefficient of thermal contraction (α), the coefficient of haline expansion (β), and the bi-static scattering angle (θ) that are specific to the inclusion of density fluctuations. It can be seen from the expressions for A and B that the contribution to scattering from the density term is maximum at backscatter ($\theta = \pi$) and disappears at angles close to forward scattering ($\theta = 0$). Finally, it should be noted that since $\alpha \gg a$ while $\beta \approx b$, the contribution to scattering from density fluctuations will be most significant under conditions in which salinity (and not temperature) microstructure dominates the scattering.

For homogeneous and isotropic random media fluctuations, the 3D wave number spectrum $\Phi(K)$ can be related to the 1D wave number spectrum, $\phi(K)$ (Tatarski, 1961, p.17, Eq. 1.27) by

$$\Phi(K) = \left[-\frac{1}{2\pi k} \frac{d\phi(k)}{dk} \right]_K. \quad (22)$$

Expressing σ_v in terms of 1D wave number spectra is useful as standard oceanographic measurements typically involve performing vertical temperature and conductivity profiles, resulting in 1D wave number spectra. Applying Eq. (22) to Eq. (20), σ_v is given by

$$\begin{aligned} \sigma_v &= -k^4 \left[\frac{1}{k} \frac{d}{dk} (A^2 \phi_T(k) + B^2 \phi_S(k) + 2AB\phi_{ST}(k)) \right]_K \\ &= -\frac{k^4}{K} \left(A^2 \frac{d\phi_T(K)}{dk} + B^2 \frac{d\phi_S(K)}{dk} + 2AB \frac{d\phi_{ST}(K)}{dk} \right), \end{aligned} \quad (23)$$

where the contribution to scattering from temperature fluctuations alone, σ_v^T, is given by the first term in this expression, scattering from salinity fluctuations alone, σ_v^S, is given by the second term in this expression, and the scattering from the temperature-salinity co-spectrum alone, σ_v^{ST}, is given by the last term. Using these expressions, it may ultimately be possible to measure the acoustic scattering from turbulent oceanic microstructure and invert for the wave number spectra of temperature and salinity, and their co-spectrum. Thus, using high-frequency acoustic scattering techniques, it may be possible to (1) overcome the technical difficulties associated with resolving the salinity spectrum with standard oceanographic measurement techniques, and (2) obtain an estimate for the yet unmeasured temperature-salinity co-spectrum.

III. PARAMETRIZATION OF SCATTERING IN TERMS OF MEASURABLE PHYSICAL PARAMETERS

The 1D wave number spectra of temperature, salinity, and the co-spectrum must be expressed in terms of measurable physical parameters in order to predict acoustic scattering from turbulent oceanic microstructure. In this section we derive expressions for σ_v^T and σ_v^S based on wave number spectra for temperature and salinity similar to those proposed by Batchelor (1959). In addition, two models for the co-spectrum of temperature and salinity are used to calculate σ_v^{ST}.

A. The temperature and salinity spectra

As suggested by Seim *et al.* (1995, 1999), we use an inertial-convective model for spatial wave numbers smaller than $k^* = (1/2^3)(\epsilon/\nu^3)^{1/4}$, and a viscous-convective model at higher wave numbers, where ν ($\mathrm{m^2 s^{-1}}$) is the molecular viscosity, and ϵ ($\mathrm{W\,kg^{-1}}$) is the dissipation rate of turbulent kinetic energy. Within the framework of these models, the temperature spectrum (Batchelor, 1959; Dillon and Caldwell, 1980) is given by

$$\begin{aligned} \phi_T^<(\tilde{k}) &= A^* \chi_T \epsilon^{-1/3} \tilde{k}^{-5/3} \quad \text{for } \tilde{k} \leqslant k^*, \\ \phi_T^>(\tilde{k}) &= \frac{\chi_T}{\zeta} \frac{G(\xi_T)}{\tilde{k}} \quad \text{for } \tilde{k} \geqslant k^*, \end{aligned} \quad (24)$$

where $\zeta = (\epsilon/\nu)^{1/2}/q$ is the strain rate. The spatial wave number is denoted by $\tilde{k}$ to distinguish it from the acoustic wave number k. Throughout the remainder of this paper, superscripts $^<$ and $^>$ indicate that the quantity being referred to is valid for wave numbers smaller than, and larger than, k^*, respectively. The salinity spectrum is given by

$$\begin{aligned} \phi_S^<(\tilde{k}) &= A^* \chi_S \epsilon^{-1/3} \tilde{k}^{-5/3} \quad \text{for } \tilde{k} \leqslant k^*, \\ \phi_S^>(\tilde{k}) &= \frac{\chi_S}{\zeta} \frac{G(\xi_S)}{\tilde{k}} \quad \text{for } \tilde{k} \geqslant k^*. \end{aligned} \quad (25)$$

The dimensionless function $G(\xi)$ (valid for either ξ_T or ξ_S) is defined by

$$G(\xi) \equiv e^{-\xi^2/2} - \xi \int_\xi^\infty e^{-x^2/2} dx, \quad (26)$$

where $\xi_T = \sqrt{2q}\tilde{k}/k_d$ and $\xi_S = \sqrt{2q}\tilde{k}/k_{ds}$ are dimensionless wave numbers, $k_d = (\epsilon/\nu D_T^2)^{1/4}$ is the diffusive cutoff wave number for temperature, $k_{ds} = (\epsilon/\nu D_S^2)^{1/4}$ is the diffusive cutoff wave number for salt, D_T ($\mathrm{m^2 s^{-1}}$) is the molecular diffusivity for temperature, and D_S ($\mathrm{m^2 s^{-1}}$) is the molecular diffusivity for salt. A^* and q are constants: $A^* = 0.925$ is chosen such that the spectra are equal at k^* (Dillon and Caldwell, 1980), and $q = 3.7$ (Oakey, 1982). The dissipation rates of salt and temperature variance are given by χ_S

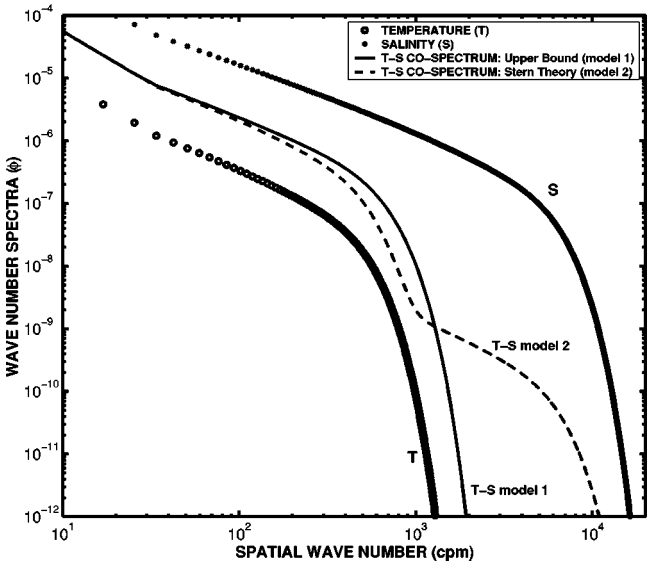


FIG. 2. Estimated wave number spectra (based on parameters for case 3 in Table I) as a function of spatial wave number for temperature (open circles), ϕ_T; salinity (crosses), ϕ_S; and the temperature-salinity co-spectrum, ϕ_{ST} (solid and dashed lines). Two models have been used for the co-spectrum: the first is an upper bound for co-spectrum (solid line), representing perfect positive correlation between temperature and salinity fluctuations, while the second model for the co-spectrum (dashed line) we have derived.

(psu^{2}s^{-1}) and χ_T (°C^{2}s^{-1}), respectively. Typical values of these parameters are given in Table I.

Since the parametrization of the temperature and salinity spectra are identical, the only difference between these spectra is the relevant scale set by k_d and k_{ds}, and the overall magnitude of the spectra, which is determined to a large part by χ_T and χ_S (Fig. 2). Since χ_S remains unmeasured, it is assumed that temperature and salinity have equal eddy diffusivities, from which it follows that the scalar dissipation rates are related by $\chi_S = \chi_T/\delta^2$ (Osborn and Cox, 1972; Gregg, 1984). The roll-off for the salinity spectrum occurs at higher spatial wave numbers ($\sim$ a factor of 10) than the roll-off for the temperature spectrum since the molecular diffusivity for salt is approximately two orders of magnitude smaller than the molecular diffusivity for temperature.

In order to evaluate σ_v it is necessary to calculate terms such as $d\phi_T/d\tilde{k}$ and $d\phi_S/d\tilde{k}$. For $\tilde{k} \leqslant k^*$,

$$\frac{d\phi_T^<(\tilde{k})}{d\tilde{k}} = -\frac{5}{3}\frac{\phi_T^<(\tilde{k})}{\tilde{k}} = -\frac{5}{3}\chi_T \epsilon^{-1/3}\tilde{k}^{-8/3},$$
$$\frac{d\phi_S^<(\tilde{k})}{d\tilde{k}} = -\frac{5}{3}\frac{\phi_S^<(\tilde{k})}{\tilde{k}} = -\frac{5}{3}\chi_S \epsilon^{-1/3}\tilde{k}^{-8/3}, \tag{27}$$

while for $\tilde{k} \geqslant k^*$,

$$\frac{d\phi_T^>(\tilde{k})}{d\tilde{k}} = -\frac{\phi_T^>(\tilde{k})}{\tilde{k}}\frac{e^{-\xi_T^2/2}}{G(\xi_T)} = -\frac{\chi_T}{\zeta}\frac{e^{-\xi_T^2/2}}{\tilde{k}^2},$$
$$\frac{d\phi_S^>(\tilde{k})}{d\tilde{k}} = -\frac{\phi_S^>(\tilde{k})}{\tilde{k}}\frac{e^{-\xi_S^2/2}}{G(\xi_S)} = -\frac{\chi_S}{\zeta}\frac{e^{-\xi_S^2/2}}{\tilde{k}^2}. \tag{28}$$

A quantity which will be useful later is the ratio of ϕ_S to ϕ_T:

$$\left(\frac{\phi_S^<(\tilde{k})}{\phi_T^<(\tilde{k})}\right)^{1/2} = \left(\frac{\chi_S}{\chi_T}\right)^{1/2} \equiv \frac{1}{\delta} \quad \text{for } \tilde{k} \leqslant k^*,$$
$$\left(\frac{\phi_S^>(\tilde{k})}{\phi_T^>(\tilde{k})}\right)^{1/2} = \left(\frac{\chi_S}{\chi_T}\frac{G(\xi_S)}{G(\xi_T)}\right)^{1/2} = \frac{R_G}{\delta} \quad \text{for } \tilde{k} \geqslant k^*, \tag{29}$$

where $R_G \equiv (G(\xi_S)/G(\xi_T))^{1/2}$. At low wave numbers $R_G \approx 1$, while above the diffusive cutoff wave number for heat, R_G increases rapidly, and the temperature spectrum is much smaller than the salinity spectrum, though both may be small.

It is now possible to evaluate σ_v^T and σ_v^S, recalling that the wave number spectra must be evaluated at the wave number, K, which is related to the acoustic wave number, k, by $K = 2k\sin(\theta/2)$:

$$\sigma_v^{T<} = -A^2\left(\frac{k^4}{K}\right)\left[\frac{d\phi_T^<(\tilde{k})}{d\tilde{k}}\right]_K = A^2\left(\frac{k^4}{K}\right)\left(\frac{5}{3}\frac{\phi_T^<(K)}{K}\right),$$
$$\sigma_v^{T>} = -A^2\left(\frac{k^4}{K}\right)\left[\frac{d\phi_T^>(\tilde{k})}{d\tilde{k}}\right]_K = A^2\left(\frac{k^4}{K}\right)\left(\frac{e^{-\xi_T^2/2}}{G(\xi_T)}\frac{\phi_T^>(K)}{K}\right), \tag{30}$$

and

$$\sigma_v^{S<} = -B^2\left(\frac{k^4}{K}\right)\left[\frac{d\phi_S^<(\tilde{k})}{d\tilde{k}}\right]_K = B^2\left(\frac{k^4}{K}\right)\left(\frac{5}{3}\frac{\phi_S^<(K)}{K}\right),$$
$$\sigma_v^{S>} = -B^2\left(\frac{k^4}{K}\right)\left[\frac{d\phi_S^>(\tilde{k})}{d\tilde{k}}\right]_K = B^2\left(\frac{k^4}{K}\right)\left(\frac{e^{-\xi_S^2/2}}{G(\xi_S)}\frac{\phi_S^>(K)}{K}\right). \tag{31}$$

Using Eqs. (30) and (31), the ratio of the temperature to salinity contribution to scattering is given by

$$\frac{\sigma_v^{T<}}{\sigma_v^{S<}} = \frac{A^2}{B^2}\frac{\phi_T^<(K)}{\phi_S^<(K)} = \delta^2\frac{A^2}{B^2} = R_{\rho c}^2 \quad \text{for } \tilde{k} \leqslant k^*,$$
$$\frac{\sigma_v^{T>}}{\sigma_v^{S>}} = \delta^2\frac{A^2}{B^2}e^{-(\xi_T^2-\xi_S^2)/2} = R_{\rho c}^2 e^{-(\xi_T^2-\xi_S^2)/2} \quad \text{for } \tilde{k} \geqslant k^*. \tag{32}$$

Here $R_{\rho c}$ is defined by

$$R_{\rho c} \equiv \delta \frac{A}{B} = \delta \frac{(a - \alpha \sin^2(\theta/2))}{(b + \beta \sin^2(\theta/2))}$$

$$= R_c \frac{(1 - (\alpha/a)\sin^2(\theta/2))}{(1 + (\beta/b)\sin^2(\theta/2))}. \quad (33)$$

The definition of $R_{\rho c}$ parallels the definitions for the vertical density, R_ρ, and sound speed, R_c, ratios [Eq. (17)]. $R_{\rho c}$ expresses the relative importance of temperature versus salinity in determining the vertical gradient for a combined density and sound speed expression. $R_{\rho c}$ depends on the multi-static scattering angle, but is independent of the acoustic wave number. For a particular set of (α, β, δ, a, b) values, $R_{\rho c}$ changes by approximately a factor 2 as a function of angle, decreasing from R_c at $\theta = 0°$ (forward scattering) to approximately $R_c/2$ at $\theta = 180°$ (backscattering). The reason for this is that for typical open ocean parameters, $\alpha \ll a$, and $(1 - \alpha/a) \approx 1$, while $\beta \approx b$, and $(1 + \beta/b) \approx 2$.

Finally, it is straightforward to show that the maximum value of σ_v^T occurs at a wave number corresponding to $k_d(2q)^{-1/2}$ (subject to the condition that $k^* < k_d$), while the maximum value of σ_v^S occurs at $k_{ds}(2q)^{-1/2}$. The ratio of the maximum value of σ_v^T to the maximum value in σ_v^S is given by

$$\frac{\max(\sigma_v^T)}{\max(\sigma_v^S)} = R_{\rho c}^2 \left(\frac{k_d}{k_{ds}} \right) = R_{\rho c}^2 \left(\frac{D_S}{D_T} \right)^{1/2} \approx \frac{R_{\rho c}^2}{10}. \quad (34)$$

B. The temperature and salinity co-spectrum

Currently, there are no data regarding the co-spectrum of temperature and salinity, and the existing theory is based on limiting cases, such as perfect correlation between temperature and salinity, or no correlation at all. We make predictions based on two models for the co-spectrum. The first involves an upper bound for the co-spectrum (Bendat and Piersol, 1986), representing a perfect correlation between temperature and salinity fluctuations. This upper bound for the co-spectrum has been used previously by other authors (Washburn *et al.*, 1996; Seim, 1999). The second model for the co-spectrum we have derived and it is based on the temperature-salinity co-variance theory of Stern (1968). It should be noted that a zero-correlation model can be implemented by neglecting the contribution to scattering from the co-spectrum term.

1. Model 1: Upper bound

The upper bound for the temperature and salinity co-spectrum is given by (Bendat and Piersol, 1986, p. 117, Eq. 5.11) $\phi_{ST}(\tilde{k}) = (\phi_T(\tilde{k})\phi_S(\tilde{k}))^{1/2}$. To evaluate σ_v^{ST} using this model for the co-spectrum it is necessary to first evaluate $d\phi_{ST}/d\tilde{k}$:

$$\frac{d\phi_{ST}}{d\tilde{k}} = \frac{1}{2}\left(\frac{\phi_S}{\phi_T} \right)^{1/2} \frac{d\phi_T}{d\tilde{k}} + \frac{1}{2}\left(\frac{\phi_S}{\phi_T} \right)^{-1/2} \frac{d\phi_S}{d\tilde{k}}, \quad (35)$$

where $d\phi_T/d\tilde{k}$ and $d\phi_S/d\tilde{k}$ are given by Eqs. (27) and (28). From Eq. (23), for $\tilde{k} < k^*$,

$$\sigma_v^{ST<} = -2AB\left(\frac{k^4}{K} \right) \frac{d\phi_{ST}^<}{dk}$$

$$= \frac{B}{A}\frac{1}{\delta}\sigma_v^{T<} + \frac{A}{B}\delta\sigma_v^{S<}$$

$$= \frac{1}{R_{\rho c}}\sigma_v^{T<} + R_{\rho c}\sigma_v^{S<}, \quad (36)$$

and, for wave numbers above k^*,

$$\sigma_v^{ST>} = -2AB\left(\frac{k^4}{K} \right) \frac{d\phi_{ST}^>}{dk}$$

$$= \frac{B}{A}\frac{R_G}{\delta}\sigma_v^{T>} + \frac{A}{B}\frac{\delta}{R_G}\sigma_v^{S>}$$

$$= \frac{R_G}{R_{\rho c}}\sigma_v^{T>} + \frac{R_{\rho c}}{R_G}\sigma_v^{S>}. \quad (37)$$

2. Model 2: Co-spectrum based on Stern's theory

We have developed a model for the co-spectrum that is based on Stern's theory (1968). Batchelor spectra for T and S are used in this derivation to evaluate the variances as a function of wave number, and we assume that this model is valid for all wave numbers of interest here. Our co-spectrum model is given by

$$\phi_{ST}(\tilde{k}) = \frac{\phi_T(\tilde{k})}{\delta} + \frac{\delta}{L_e}\phi_S(\tilde{k}), \quad (38)$$

where L_e ($= D_T/D_S$) is the diffusivity ratio, or Lewis number, and varies from ≈ 80 in warm water to ≈ 230 in cold water. It can be seen from Fig. 2 that at high spatial wave numbers our co-spectrum model predicts a higher correlation between temperature and salinity fluctuations than the upper bound model.

The scattering contribution from this co-spectrum model is given by

$$\sigma_v^{ST} = -2AB\left(\frac{k^4}{K} \right) \frac{d\phi_{ST}(K)}{d\tilde{k}}$$

$$= \frac{2}{\delta}\frac{B}{A}\sigma_v^T + 2\frac{\delta}{L_e}\frac{A}{B}\sigma_v^S$$

$$= \frac{2}{R_{\rho c}}\sigma_v^T + \frac{2R_{\rho c}}{L_e}\sigma_v^S. \quad (39)$$

Since the sign of $R_{\rho c}$ is determined by δ, and all terms relating to the co-spectrum contribution to scattering have $R_{\rho c}$ or $R_{\rho c}^{-1}$ as a prefactor, the effect of the co-spectrum on σ_v is determined by the sign of δ. If δ is positive, either model for the co-spectrum results in an increase in the magnitude of σ_v: $\sigma_v \geq \sigma_v^S + \sigma_v^T$. If δ is negative, $\sigma_v \leq \sigma_v^S + \sigma_v^T$.

IV. PREDICTIONS

To illustrate many of the points of this paper, values that are typical for an open ocean environment were chosen for the model input parameters (Table I). Modest values of ϵ and

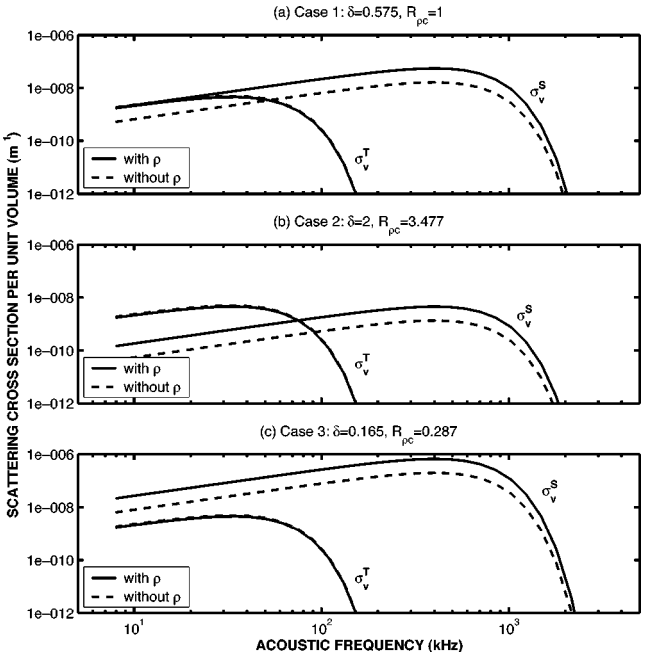


FIG. 3. The temperature, σ_v^T, and salinity, σ_v^S, contributions to the scattering cross section per unit volume as a function of acoustic frequency (at backscattering) for model input parameters given in Table I. The solid lines correspond to σ_v^T and σ_v^S calculated with the density fluctuations included, while the dashed lines correspond to σ_v^T and σ_v^S calculated without the inclusion of density fluctuations. The inclusion of density fluctuations does not significantly affect the value of σ_v^T at any frequency or angle. In contrast, inclusion of the density term significantly increases the value of σ_v^S over the entire frequency range investigated. The *relative* increase in scattering that occurs due to the inclusion of density fluctuations does not depend on the values of ϵ and χ_T. Thus the importance of including the density term can be assessed without precise knowledge of the dissipation rates.

χ_T were chosen. Changes in χ_T simply result in an overall increase or decrease in σ_v, and do not change the overall shape of σ_v. On the other hand, changes in ϵ affect both the overall magnitude as well as the wave number dependence of σ_v. Since the parameter $R_{\rho c}$ is important in determining the relative magnitudes of the temperature and salinity contributions to scattering, and thus is also important in determining the relative importance of the density versus sound speed contribution, predictions based on three values of $R_{\rho c}$, evaluated at $\theta=\pi$, are discussed in this section (Table I): (1) $R_{\rho c}=1$, (2) $R_{\rho c}=3.477(>1)$, and (3) $R_{\rho c}=0.287(<1)$. We assume a backscattering orientation for most of the predictions presented in this section, though the angular dependence is also investigated in Sec. IV D.

A. The relative contribution to scattering from density versus sound speed fluctuations

The importance of including density fluctuations in predicting scattering from turbulent oceanic microstructure can be assessed by comparing predictions of σ_v arising from the inclusion of both sound speed and density fluctuations [Eq. (20)] to predictions of σ_v arising from the inclusion of sound speed fluctuations alone [Eq. (21)] (Fig. 3). The expressions for σ_v are very similar in form: the coefficients A and B in Eq. (20) are simply replaced by a and b in Eq. (21). Since $a \gg \alpha$, and $A=a-\alpha\sin^2(\theta/2)\approx a$, the inclusion of the density term has a minimal effect on σ_v^T, at any angle. In contrast, since $b\approx\beta$, and $B=b+\beta\sin^2(\theta/2)$, the inclusion of the density term increases σ_v^S from b^2 to $B^2\approx 4b^2$ at backscattering. This corresponds to approximately a 6-dB increase in the scattering level. The magnitude of σ_v^{ST} increases from $2ab$ to $2AB\approx 4ab$ due to the inclusion of the density term, corresponding to a 3-dB increase. These *relative* increases do not depend strongly on the model input parameter values, the form of the spectra or co-spectrum, or on the acoustic wave number (at any given scattering angle). However, the effects of including density fluctuations are strongly dependent on the scattering angle, with the largest effects occurring at backscattering, while becoming negligible at angles close to forward scattering. From this point on, all predictions will include the contribution to scattering from density fluctuations.

B. The relative contribution to scattering from temperature versus salinity microstructure

The relative magnitude of σ_v^T and σ_v^S is strongly dependent on the spatial wave number (Fig. 3), which is related to the acoustic wave number through the condition that $K=2k\sin(\theta/2)$. For spatial wave numbers below k^*, the term $R_{\rho c}^2$ determines the relative magnitude of σ_v^T and σ_v^S [Eq. (32)]. Since $R_{\rho c}^2$ is independent of wave number, at any given scattering angle, σ_v^T and σ_v^S are simply offset by a constant amount: if $|R_{\rho c}|>1$, then σ_v^T dominates, while if $|R_{\rho c}|<1$, then σ_v^S dominates the scattering. For spatial wave numbers above k^*, the term $R_{\rho c}^2 e^{-(\xi_T^2-\xi_S^2)/2}$, which is a function of both angle and wave number, determines the relative magnitude of σ_v^T and σ_v^S [Eq. (32)]. Above the diffusive cutoff wave number for temperature, k_d, the exponential term decays rapidly and, for any reasonable value of $R_{\rho c}^2$, $\sigma_v^S \gg \sigma_v^T$ (though both may be small). For wave numbers between k^* and k_d, the exponential term does not yet deviate greatly from 1, and $R_{\rho c}^2$ is again critical in determining the relative magnitude of σ_v^T and σ_v^S. A more general way to estimate the importance of salinity versus temperature in determining scattering is to examine the maximum values attained by σ_v^S and σ_v^T [Eq. (34)]. The maximum value of σ_v^S is larger than the maximum value of σ_v^T when $R_{\rho c}<(D_T/D_S)^{1/4}\approx 3$, though it must be recalled that these maxima occur at different wave numbers since the diffusive cutoff wave numbers for heat and salt differ by an order of magnitude.

C. The relative contribution to scattering from the co-spectrum

The contribution to scattering from the upper bound co-spectrum (model 1) is given by Eq. (36) for spatial wave numbers below k^* and by Eq. (37) for spatial wave numbers above k^* (Fig. 4). Recalling that $\sigma_v^{T<}=R_{\rho c}^2\sigma_v^{S<}$ [Eq. (32)], the contribution to scattering from the upper bound co-spectrum (model 1) for wave numbers below k^* is thus given by $\sigma_v^{ST<}=\sigma_v^{T<}/R_{\rho c}+R_{\rho c}\sigma_v^{S<}=\sigma_v^T(2/R_{\rho c})$. The contribution to scattering from our co-spectrum [model 2: Eq. (39)] is $\sigma_v^{ST<}=\sigma_v^{T<}(1+L_e^{-1})(2/R_{\rho c})$ for wave numbers below k^*. Since $L_e \gg 1$, the contribution to scattering from σ_v^{ST} is approximately equal for both models and is determined by $2/|R_{\rho c}|$.

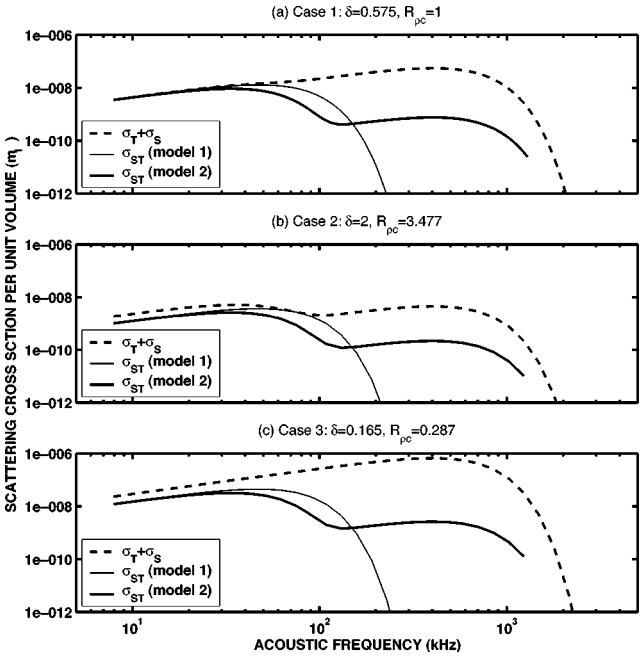


FIG. 4. The sum of the temperature and salinity contributions to the scattering cross section per unit volume $\sigma_v^T+\sigma_v^S$ (dashed line), the contribution to the scattering cross section per unit volume from the upper bound model for the co-spectrum σ_v^{ST} (model 1: thin line), and the contribution to the scattering cross section per unit volume from the co-spectrum model based on Stern's theory σ_v^{ST} (model 2: thick line). Predictions are plotted as a function of acoustic frequency, at backscattering, for model input parameters taken from Table I.

For wave numbers above k_d, $\sigma_v^S \gg \sigma_v^T$. Consequently, for the upper bound co-spectrum [model 1: Eq. (37)] $\sigma_v^{ST}=\sigma_v^S(R_{\rho c}/R_G)$, while for our co-spectrum [model 2: Eq. (39)] $\sigma_v^{ST}\approx\sigma_v^S(2R_{\rho c}/L_e)$. However, since R_G increases relatively rapidly, and $2R_{\rho c}/L_e \ll 1$ for typical values of $R_{\rho c}$ investigated here, the contribution to scattering from either co-spectrum, for wave numbers above k_d, is small. For the small range of wave numbers between approximately k^* and k_d, the wave number dependence of σ_v^{ST} is quite complicated, depending on a delicate balance set by the values of $R_{\rho c}$, R_G, L_e, and the relative magnitudes of σ_v^T and σ_v^S (Fig. 4).

Briefly synthesizing, the contribution from the co-spectrum does not, in general, significantly alter the scattering trends, particularly at wave numbers above k_d. The most significant changes are expected to occur over a small wave number range between k^* and k_d.

D. The angular dependence of scattering from temperature and salinity microstructure

The angular dependence of σ_v^T and σ_v^S (Fig. 5) is determined by both $R_{\rho c}$ and the acoustic wave number, k. However, the angular dependence is significantly more sensitive to changes in k than to changes in $R_{\rho c}$, since $R_{\rho c}$ only changes by a factor of 2 between forward and backscattering. For a fixed acoustic wave number, k, as θ changes from backscattering ($\theta=\pi$) to forward ($\theta=0$) scattering, the wave number K spans through the values from $K=2k$ to 0. There are three wave number regions that must be considered in order to understand the general angular trends in σ_v^S

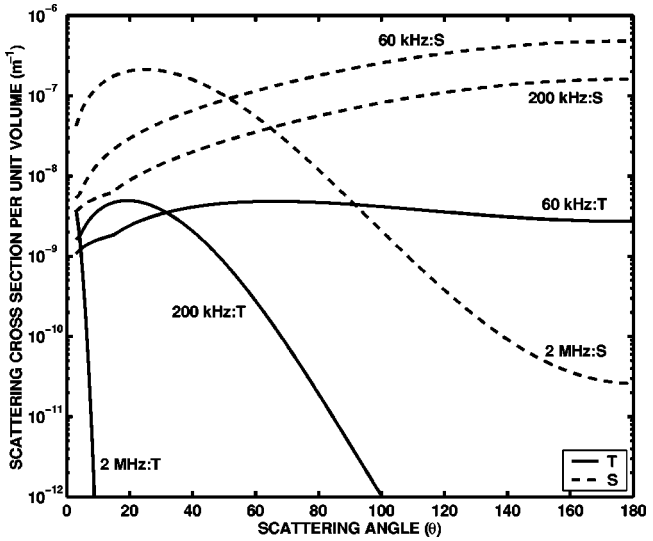


FIG. 5. The angular dependence of the temperature, σ_v^T, and salinity, σ_v^S, contributions to the scattering cross section per unit volume at (i) 60 kHz (region I: $k_d=769$ cpm, $k=251$ cpm, $K=502$ cpm at $\theta=\pi$), (ii) 200 kHz (region II: $k_d=769$ cpm, $k_{ds}=9207$ cpm, $k=838$ cpm, $K=1676$ cpm at $\theta=\pi$), and (iii) 2.3 MHz (region III: $k_{ds}=9207$ cpm, $k=9634$, $K=19268$ cpm at $\theta=\pi$), for model input parameters taken from case 3 in Table I.

and σ_v^T (Fig. 5): (i) $2k \leqslant k_d$, (ii) $k_d \leqslant 2k \leqslant k_{ds}$, and ($iii$) $2k \geqslant k_{ds}$. In region (i), the angular dependence of both σ_v^T and σ_v^S is relatively flat since $K=2k$ lies below the sharp roll-off in σ_v^T. In region (ii), σ_v^T depends very strongly on the scattering angle, while σ_v^S does not. At backscattering, $K=2k$ corresponds to wave numbers above the sharp roll-off in σ_v^T, and as the angle decreases towards forward scattering, the wave number K spans through the sharp roll-off, resulting in a strong angular dependence. In contrast, σ_v^S remains relatively flat since $K<k_{ds}$ for all angles. In region (iii), both σ_v^T and σ_v^S depend strongly on the multi-static scattering angle since $2k>k_{ds}>k_d$ and as θ changes from π to 0, K sweeps through the sharp roll-off in both the temperature and salinity. It is clear that the differences between the angular trends in σ_v^T and σ_v^S are largest in region (ii), where the wave number K at backscattering lies between the diffusive cutoff wave number for temperature and salinity. In fact, the vast differences predicted between the angular dependence of σ_v^S and σ_v^T in this region of wave number space suggest that multi-static measurements of acoustic scattering may provide a very fruitful technique for discriminating between temperature and salinity microstructure.

E. Acoustic scattering predictions based on high resolution microstructure data

In this section we make predictions of acoustic backscattering from oceanic microstructure based on data from which all the necessary model input parameters can be extracted. This data set involves microstructure data taken on the R/V NEW HORIZON in March 1991 on a cruise to the seamount Fieberling Guyot. A high resolution profiler (HRP) (Schmitt *et al.*, 1988) was deployed 95 times above and around the seamount, which rises from background depths of approximately 4000 m to about 500 m below the surface.

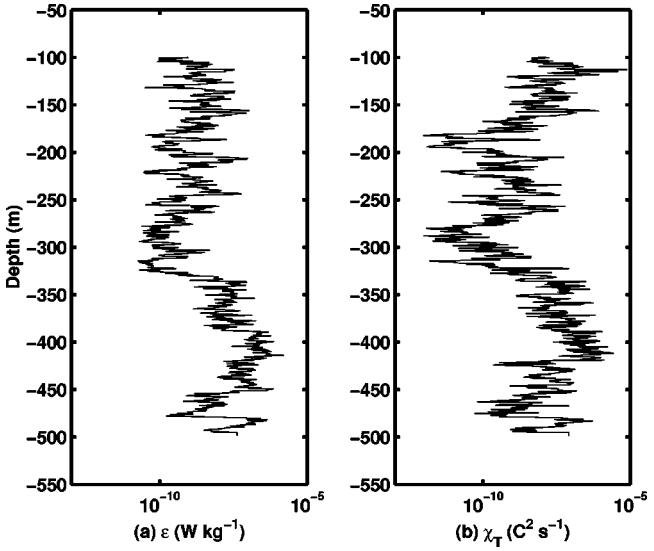


FIG. 6. Vertical profiles of (a) ϵ (W kg^{-1}) and (b) χ_T (°C^2 s^{-1}) performed with the HRP directly over the seamount summit Fieberling Guyot in 500 m of water. Dissipation rates just above the summit are significantly elevated. Profiles performed off the seamount shoulder (not shown here) show less structure and have uniformly lower values. The noise floor is typically around 10^{-11} (W kg^{-1}) for ϵ and 10^{-12} (°C^2 s^{-1}) for χ_T.

From the microstructure sensors on the HRP it is possible to obtain all the necessary model input parameters, at scales commensurate with the acoustic wavelengths of interest.

Acoustic scattering predictions were made for a large subset of the HRP profiles. We present results for a HRP profile performed directly above the seamount summit, since the microstructure measurements indicated that there were increased dissipation rates and velocity shear at the seamount summit, resulting in increased mixing and turbulence (Toole *et al.*, 1997) (Fig. 6). The acoustic scattering predictions reflect these observations, with elevated scattering levels predicted for both temperature and salinity microstructure over the seamount summit (Fig. 7). Over a broad frequency range, the predicted contribution to scattering from salinity microstructure just over the seamount summit is larger than the contribution from temperature microstructure. The results of our model also indicate that the contribution to scattering from the density term is at least as significant as the contribution from the sound speed term. There is a scattering layer between 350 and 450 m in which the predicted scattering levels from salinity microstructure (with approximately equal contributions from density and sound speed fluctuations) are similar in magnitude to typical scattering levels expected for zooplankton (Wiebe *et al.*, 1997). These results suggest that if the biological processes in the vicinity of the seamount could be accurately characterized, acoustic scattering techniques might provide a viable means to map areas of high turbulence.

V. SUMMARY AND CONCLUSIONS

In this paper we have extended the current theory of acoustic scattering from turbulent oceanic microstructure to include random fluctuations in density. Previously, it had been assumed that acoustic scattering from oceanic microstructure was due to sound speed fluctuations alone. We have predicted that the contribution to scattering from fluctuations in the density can be comparable to the contribution from sound speed fluctuations, under some circumstances. Depending on the scattering angle, the density contribution can increase the scattering levels by as much as 6 dB, resulting in peak scattering levels that, under certain conditions, could be comparable to levels typically observed for scattering from zooplankton. Neglecting to include the density term can consequently lead to a potentially significant underestimate of volume scattering strengths.

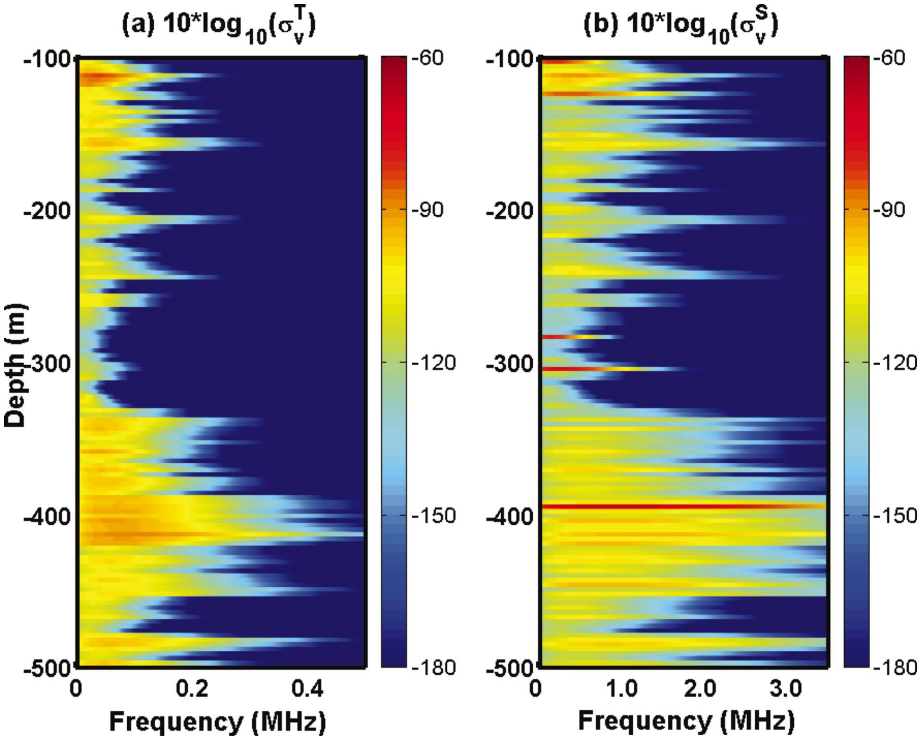


FIG. 7. Predicted contributions to backscattering from (a) temperature microstructure ($10\log_{10}\sigma_v^T$) and (b) salinity microstructure ($10\log_{10}\sigma_v^S$) as a function of depth and acoustic frequency based on HRP microstructure data obtained directly above the seamount Fieberling Guyot (see Fig. 6). The contribution to scattering from density fluctuations has been included in all scattering predictions. The contribution to scattering from the cospectrum (not shown) was found to be small. Note the different horizontal scales for temperature and salinity. There is a layer of elevated scattering just above the seamount summit, with levels that are comparable to those expected from typical zooplankton patches.

As with scattering from media variability in the atmosphere, the inclusion of the density term results in an expression for the scattering cross section per unit volume that is explicitly dependent on the multi-static scattering angle. We predict a very strong dependence on the multi-static scattering angle at certain acoustic wave numbers, and suggest that this dependence could be exploited to distinguish between the contribution to scattering from temperature and salinity microstructure.

The derivations presented here are based on far field weak scattering theory, for which the Born approximation is valid. One of the primary assumptions is that the medium variability is stationary and homogeneous, allowing a statistical description of the scattering in terms of the spatial Fourier transform of the correlation function of the medium variability, which is simply the 3D wave number spectrum of the medium variability. By relating the variability in the density and sound speed to fluctuations in temperature and salinity at the microstructure scale we have obtained an expression for the scattering cross section per unit volume in terms of the wave number spectra of temperature, salinity, and the co-spectrum of temperature and salinity. The assumption of isotropy also allows the 3D wave number spectra to be expressed in terms of the 1D wave number spectra, which are more representative of spectra derived from oceanographic measurements.

By assuming a 1D Batchelor spectrum for temperature and salinity, expressions for the scattering cross section per unit volume have been derived in terms of parameters such as the dissipation rates of turbulent kinetic energy, temperature and salt variance. We have found that the parameter $R_{\rho c}$ is critical in determining the relative contribution to scattering from temperature and salinity microstructure. $R_{\rho c}$ is defined in a manner similar to the density (R_ρ) and sound speed (R_c) ratios, but combines the effects of both vertical density and sound speed changes. $R_{\rho c}$ depends on the multi-static scattering angle, but not on the acoustic frequency. At angles close to forward scattering $R_{\rho c}$ is approximately equal to the vertical sound speed ratio R_c. However, at angles close to backscattering, $R_{\rho c}$ is approximately equal to $R_c/2$.

Two models for the co-spectrum of temperature and salinity have been used. The first expression represents an upper bound for the temperature and salinity co-spectrum. The second expression for the co-spectrum we have derived, and is based on the temperature-salinity covariance theory of Stern (1968). We have found that the contribution to scattering from either of the co-spectrum models tends to be smallest at spatial wave numbers above the diffusive cutoff wave number for heat. For wave numbers below this, the approximate contribution from the co-spectrum for either model is given by $2/R_{\rho c}$. If $R_{\rho c} \approx -1$, the contribution to scattering from the co-spectrum term can almost exactly cancel the contribution from the temperature and salinity terms combined, over a limited range of wave numbers. If $R_{\rho c} < 0$ (>0), the contribution to the scattering from the co-spectrum tends to reduce (increase) the magnitude of the scattering that results from the sum of the temperature and salinity contributions.

We have made scattering predictions based on generic model input parameter values that are typically for open ocean environments. In addition, scattering predictions based on high resolution microstructure measurements obtained in the vicinity of a seamount in the northeast subtropical Pacific ocean, Fieberling Guyot, have also been made. For these data, there are no free model input parameters outside of the initial model assumptions. Our results indicate that the layer of elevated turbulence above the seamount summit could give rise to significant scattering levels, comparable to those of typical zooplankton patches, particularly at higher wave numbers. The possibility that particulate scatterers, such as microbubbles, small zooplankton, or sand, may aggregate at the locations of energetic turbulence has not been considered in this analysis. Our predictions indicate that high-frequency acoustic scattering could be a viable technique, in combination with appropriate ground truthing, to map regions of elevated turbulent microstructure.

In conclusion, models such as the one presented here are important for the accurate interpretation of acoustic scattering data, though supporting physical and environmental information, gathered by any variety of techniques, will probably always be necessary for the unambiguous interpretation of high-frequency acoustic scattering data in terms of either physical or biological processes. However, before it is possible to fully capitalize on the acoustic scattering model presented here for scattering from turbulent oceanic microstructure, controlled field and laboratory testing and validation are necessary, in which the physical environment (temperature, conductivity, and fluid velocity) is characterized at least at the same resolution as the acoustic wave length, in addition to the adequate characterization of particulate scatterers that may be present in the areas of elevated turbulence.

ACKNOWLEDGMENTS

The authors would like to thank Harvey Seim (Marine Sciences Department, University of North Carolina Chapel Hill) for providing his code for calculating the backscattering cross section due to sound speed fluctuations. We would also like to thank Lou Goodman (Graduate School for Marine Science and Technology, University of Massachusetts Dartmouth) for invaluable discussions on the importance of density fluctuations in determining high-frequency acoustic scattering. This work was supported in part by ONR, NSF, and the Woods Hole Oceanographic Institution. This is Woods Hole Oceanographic Institution Contribution No. 10750.

Batchelor, G. K. (**1959**). ''Small-scale variation of convective quantities like temperature in a turbulent fluid,'' J. Fluid Mech. **5**, 113–133.

Bendat, J. S., and Piersol, A. G. (**1986**). *Random Data: Analysis and Measurements Procedures* (Wiley, New York).

Chernov, L. A. (**1960**). *Wave Propagation in a Random Medium* (McGraw-Hill, New York).

Chu, D., and Ye, Z. (**1999**). ''A phase-compensated distorted wave Born approximation representation of the bistatic scattering by weakly scattering objects: Application to zo-oplankton,'' J. Acoust. Soc. Am. **106**, 1732–1743.

Dillon, T. M., and Caldwell, D. R. (**1980**). ''The Batchelor spectrum and dissipation in the upper ocean,'' J. Geophys. Res. **85**, 1910–1916.

Farmer, D. M., and Smith, J. D. (**1979**). ''Tidal interaction of stratified flow with a sill in Knight inlet,'' Deep-Sea Res., Part A **27**, 239–254.

Farmer, D. M., and Armi, L. (**1999**). ''The generation and trapping of internal solitary waves over topography,'' Science **283**, 188–190.

Flatte, S. R., Dashen, R., Munk, W. H., Watson, K. M., and Zachariasen, F. (**1979**). *Sound Transmission through a Fluctuating Ocean* (Cambridge U. P., London), p. 5.
Goodman, L. (**1990**). ''Acoustic scattering from ocean microstructure,'' J. Geophys. Res. **95**, 11557–11573.
Goodman, L., and Kemp, K. A. (**1981**). ''Scattering from volume variability,'' J. Geophys. Res. **86**, 4083–4088.
Goodman, L., Oeschger, J., and Szargowicz, D. (**1992**). ''Ocean acoustics turbulence study: Acoustic scattering from a buoyant axisymmetric plume,'' J. Acoust. Soc. Am. **91**, 3212–3227.
Gregg, M. (**1984**). ''Entropy generation in the ocean by small-scale mixing,'' J. Phys. Oceanogr. **14**, 688–711.
Haury, L. R., Briscoe, M. G., and Orr, M. H. (**1979**). ''Tidally generated internal wave packets in Massachusetts Bay,'' Science **278**, 312–317.
Holliday, D. V., and Pieper, R. E. (**1995**). ''Bioacoustical oceanography at high frequencies,'' ICES J. Mar. Sci. **52**, 279–296.
Ishimaru, A. (**1978**). *Wave Propagation and Scattering in Random Media* (Academic, New York), Vol. 2.
Kraichnan, R. H. (**1953**). ''The scattering of sound in a turbulent medium,'' J. Acoust. Soc. Am. **25**, 1096–1104.
Lavery, A. C., Stanton, T. K., Chu, D., and McGehee, D. (**2002**). ''Three-dimensional modeling of acoustic backscattering from fluid-like zooplankton,'' J. Acoust. Soc. Am. **111**, 1197–1210.
Lhermitte, R., and Lemmin, U. (**1993**). ''Turbulent flow microstructures observed by sonar,'' Geophys. Res. Lett. **20**, 823–826.
McGehee, D. M., O'Driscoll, R. L., and Martin Traykovski, L. V. (**1998**). ''Effect of orientation on acoustic scattering from Antarctic krill at 120 kHz,'' Deep-Sea Res., Part II **45**, 1273–1294.
Medwin, H., and Clay, C. S. (**1998**). *Fundamentals of Acoustical Oceanography* (Academic, New York).
Menemenlis, D., and Farmer, D. M. (**1995**). ''Path-averaged measurements of turbulence beneath ice in the Arctic,'' J. Geophys. Res. **100**, 13655–13663.
Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (Princeton U. P., Princeton, NJ).
Munk, W. H., and Garrett, C. J. R. (**1973**). ''Internal wave breaking and microstructure,'' Boundary-Layer Meteorol. **4**, 37–45.
Oakey, N. S. (**1982**). ''Determination of the rate of dissipation of turbulent energy from simultaneous temperature and shear microstructure measurements,'' J. Phys. Oceanogr. **12**, 256–271.
Oeschger, J., and Goodman, L. (**1995**). ''Acoustic scattering from a thermally driven buoyant plume,'' J. Acoust. Soc. Am. **100**, 1421–1462.
Osborn, T., and Cox, C. S. (**1972**). ''Oceanic fine structure,'' Geophys. Fluid Dyn. **3**, 321–345.
Orr, M. H., Haury, L. R., Wiebe, P. H., and Briscoe, M. G. (**2000**). ''Backscatter of high-frequency (200 kHz) acoustic wave fields from ocean turbulence,'' J. Acoust. Soc. Am. **108**, 1595–1601.
Proni, J. R., and Apel, J. R. (**1975**). ''On the use of high-frequency acoustics for the study of internal waves and microstructure,'' J. Geophys. Res. **80**, 1147–1151.
Sandstrom, H., Elliott, J. A., and Cochrane, N. A. (**1989**). ''Observing groups of solitary internal waves and turbulence with BATFISH and Echo-Sounder,'' J. Phys. Oceanogr. **19**, 987–997.
Schmitt, R. W., Toole, J. M., Koehler, R. L., Mellinger, E. C., and Doherty, K. W. (**1988**). ''The development of a fine- and microstructure profiler,'' J. Atmos. Ocean. Technol. **5**, 484–500.
Seim, H. E. (**1999**). ''Acoustic backscatter from salinity microstructure,'' J. Atmos. Ocean. Technol. **16**, 1491–1498.
Seim, H. E., Gregg, M. C., and Miyamoto, R. T. (**1995**). ''Acoustic backscatter from turbulent microstructure,'' J. Atmos. Ocean. Technol. **12**, 367–380.
Stanton, T. K., Chu, D., and Wiebe, P. H. (**1998**). ''Sound scattering by several zooplankton groups. II. Scattering models,'' J. Acoust. Soc. Am. **103**, 236–253.
Stanton, T. K., Wiebe, P. H., Chu, D., and Goodman, L. (**1994**). ''Acoustic characterization and discrimination of marine zooplankton and turbulence,'' ICES J. Mar. Sci. **51**, 469–479.
Stern, M. E. (**1968**). ''T-S gradients on the micro-scale,'' Deep-Sea Res. Oceanogr. Abstr. **15**, 245–250.
Tatarski, V. I. (**1961**). *Wave Propagation in a Turbulent Medium* (McGraw-Hill, New York).
Thorpe, S. A., and Brubaker, J. M. (**1983**). ''Observation of sound reflection by temperature microstructure,'' Limnol. Oceanogr. **28**, 601–613.
Toole, J. M., Schmitt, R. W., Polzin, K. L., and Kunze, E. (**1997**). ''Near-boundary mixing above the flanks of a midlatitude seamount,'' J. Geophys. Res. **102**, 947–959.
Trevorrow, M. V. (**1998**). ''Observation of internal solitary waves near the Oregon coast using an inverted echo-sounder,'' J. Geophys. Res. **103**, 7671–7680.
Urick, R. J. (**1980**). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York).
Waag, R. C. (**1984**). ''A review of tissue characterization from ultrasonic scattering,'' IEEE Trans. Biomed. Eng. **88**, 2418–2436.
Waag, R. C., Astheimer, J. P., and Swarthout, G. W. (**1985**). ''A characterization of wave-front distortion for analysis of ultrasound diffraction measurements made through an inhomogeneous medium,'' IEEE Trans. Sonics Ultrason. **32**, 36–48.
Waag, R. C., Dalecki, D., and Christopher, P. E. (**1989**). ''Spectral power determinations of compressibility and density variations in model media and calf liver using ultrasound,'' J. Acoust. Soc. Am. **85**, 423–431.
Washburn, L., Duda, T. F., and Jacobs, D. C. (**1996**). ''Interpreting conductivity microstructure: Estimating the temperature variance dissipation rate,'' J. Atmos. Ocean. Technol. **13**, 1166–1188.
Wiebe, P. H., Stanton, T. K., Benfield, M. C., Mountain, D. G., and Greene, C. H. (**1997**). ''High-frequency acoustic volume backscattering in the Georges Bank coastal region and its interpretation using scattering models,'' IEEE J. Ocean. Eng. **22**(3), 445–464.

Changes in volume reverberation from deep to shallow water in the eastern Gulf of Mexico

Richard H. Love
BayouAcoustics, Pass Christian, Mississippi 39571-2111

Charles H. Thompson and Redwood W. Nero
Naval Research Laboratory, Stennis Space Center, Mississippi 39529-5004

(Received 29 January 2002; revised 18 May 2003; accepted 28 July 2003)

Scattering from fish is a primary cause of volume reverberation and, since fish populations change from deep to shallow water, the character of volume reverberation should also change. However, there are few data available to document expected changes. Therefore, an experiment was conducted in the eastern Gulf of Mexico to investigate possible changes in volume reverberation from deep to slope to shelf waters. Results showed that volume reverberation in outer shelf waters varied more rapidly with respect to both time and space than that in deeper waters. Day-time scattering was similar for deep, slope and shelf waters, total scattering strengths generally increased with frequency. Night-time scattering for the deep ocean and slope also increased with increasing frequency. Scattering modeling suggests that swimbladder-bearing fishes smaller than 10 cm were responsible for the observed volume reverberation. Night-time scattering at the outer shelf location was very different, with strong scattering peaks at low frequencies. Scattering modeling implicates 12–15 cm rough scad and round herring as potential causes of the low frequency peaks. Hence, experiment results confirmed that, as expected, volume reverberation over shelf waters is different and more variable than in deeper waters of the eastern Gulf of Mexico. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1610458]

PACS numbers: 43.30.Ft, 43.30.Sf [WMC] Pages: 2698–2708

I. INTRODUCTION

Volume reverberation at frequencies less than 10 kHz is caused primarily by resonance scattering from swimbladders of dispersed fish. Since fish populations naturally change from deep to shallow water, one would expect the character of volume reverberation to likewise change. However, because volume reverberation over shelf and slope waters has not been of interest until recently, there are few data available to document expected changes.

In most deep water regions of the world's oceans, the dominant volume scatterers are small mesopelagic fishes.[1] The geographic distributions of these fishes are governed by water mass properties and are usually quite uniform within a water mass.[2] Therefore, the characteristics of volume reverberation in these regions are normally stable within a water mass. In a few regions, such as the Norwegian Sea, larger, commercial-size fishes are the dominant scatterers.[3] These fish are more mobile than the mesopelagics and they are not uniformly distributed within a water mass. Thus, the characteristics of volume reverberation in these regions will tend to vary over a water mass.

Mesopelagic fishes are generally considered to be confined to the deep ocean but significant numbers have been captured over the outer shelf and slope in a number of areas, including the Gulf of Mexico.[4] Other fish communities found on the outer shelf and slope include pelagic, semi-demersal, and demersal species. Distributions of these fishes depend on bottom depths and types and oceanographic conditions. In shelf waters, bottom types and oceanographic conditions can change over relatively short distances. Oceanographic conditions also change over relatively short time periods. Thus, the characteristics of volume reverberation over shelves can be expected to be much more variable than in deep water.

During January 1993, the Naval Research Laboratory (NRL) conducted an engineering test of new volume reverberation instrumentation in the Gulf of Mexico. This test was designed to include a short experiment to examine volume reverberation in deep, slope and shelf waters of the eastern Gulf. The results of the experiment, which was the first time NRL obtained volume reverberation data in relatively shallow water, are the subject of this paper.

II. MEASUREMENTS

Volume reverberation measurements were made from 9 to 11 January 1993 aboard the *USNS Bartlett (T-AGOR-13)* at the three locations shown in Fig. 1. Station 1 was in water depths over 3000 m. Station 2 was on the continental slope in 420–450 m. Station 3 was near the edge of the continental shelf in 175–205 m. Day-time and night-time measurements were made at each station.

The volume reverberation measurement technique employed explosive sources and downward-looking receivers to measure volume scattering strengths versus depth over a wide frequency range. Volume scattering strength (S_V) is defined as

$$S_V = 10\log(s_V), \quad \text{where } s_V = (I_{\text{SCAT}}/I_{\text{INC}}). \tag{1}$$

I_{SCAT} is the intensity of sound scattered from a unit volume (1 m^3) of water at unit distance (1 m) from the volume and

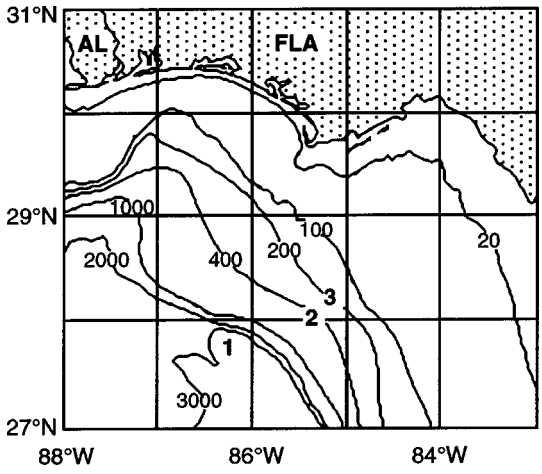


FIG. 1. Station locations.

I_{INC} is the intensity incident on the volume.[5] S_V is given in decibels.

Sources were 0.23 kg blocks of TNT electrically detonated 0.5 m below the surface. This shallow detonation depth allows the gas bubble created by the explosion to vent and eliminates the multiple sound pulses caused by bubble oscillations associated with detonations at greater depths.

The receiver NRL used prior to this experiment was an NRL USRD Model F55, which consists of a 32-element line hydrophone, 0.9 m long, mounted along the axis of a 1.8 m diameter 45° conical reflector. By grouping the upper 8, 16, or all 32 hydrophone elements, the F55's 3 dB beamwidth can be maintained between 10° and 20° between 2.5 and 20 kHz. At low frequencies, increasing beamwidth couples with decreasing sensitivity to make the F55 ineffective. Under conditions of high volume reverberation and low sea surface reverberation, valid F55 data have been collected down to 800 Hz.[3] In an effort to obtain usable data to lower frequencies, the NRL USRD Model F78, a larger version of the F55, was designed and built. The F78 consists of a 48-element line hydrophone, 1.4 m long, mounted along the axis of a 2.7 m diameter 45° conical reflector. The upper 8, 16, 32 or all 48 elements of the F78 can be grouped. Beam patterns for the first three groupings are similar to those of the F55. Beamwidths of the 48 element grouping are less than 20° down to 1.7 kHz. Both the F55 and F78 were deployed to a depth of 8 m, just below the ship's hull.

The primary purpose of the engineering test was to evaluate the F78 and compare its performance to that of the F55. Ideally, this would have been accomplished with both receivers deployed at the same time. However, simultaneous deployment was not possible, so measurements with the second receiver were conducted as quickly as possible after those with the first. Excluding the first set of day-time measurements at Station 1, where a number of system checks were conducted, and the day-time measurements at Station 3, where there was a 1 h delay, total time for a measurement sequence with both receivers was about 1 h, with about 0.5 h between the last shot with the first receiver and the first shot with the second receiver.

After night-time measurements with the F55 were completed at Station 2 on 10 January, it was decided that weather conditions had become too rough to deploy the F78. Measurements at Station 2 with both the F78 and the F55 were completed the following night, 11 January. Thus, there were two night-time sequences at Station 2.

In previous experiments using the F55, a typical measurement sequence consisted of six shots. For this experiment, the typical sequence consisted of three shots with each receiver. Received signals from each shot were amplified, high- and low-pass filtered at 400 Hz and 12 kHz, digitized at a 40 kHz sampling rate, and stored. Digitally stored data were subsequently filtered into 1/6-octave bands and amplitude versus time envelopes calculated for each band. Data samples collected just prior to source detonation were used to calculate average noise levels, which were used for noise level corrections to the data. Noise level corrected amplitudes were used to calculate S_V for each band,

$$S_V = 20\log V + 20\log t - 10\log E - 10\log(1-\cos(\beta/2)) + \alpha c t + 10\log\eta - \mathrm{FFVS} - \mathrm{Gain} - 171.1, \quad (2)$$

where V is the noise-corrected voltage amplitude, t is time in seconds after source detonation, E is the source energy per unit area in the 1/6-octave band in erg/cm^2 at 100 m, β is the 3 dB beamwidth of the receiver, α is absorption in dB/m, c is sound speed in m/s, the η term accounts for noncoincident source/receiver geometry, FFVS is the free field voltage sensitivity of the receiver in dB re 1 v/μPa, Gain is the amplification of the recorded signal, and 171.1 is a constant that includes the density of seawater and unit conversions for acoustic pressure and source energy. Equation (2) is derived from a method published by Machlup and Hersey.[6]

This initial S_V calculation uses ideal beam patterns, which can cause scattering strengths below a strong layer to be artificially raised due to off-axis signals from depths within the layer. Therefore, entire beam patterns were used in an iterative process to remove off-axis contributions from the data.[7] Images showing final S_V in dB on a color scale as a function of frequency and depth were then produced for each shot.

Depths where significant scattering occurred were determined by visually examining the S_V images. Integration over those depths produced a layer scattering strength (S_L) for each frequency band:

$$S_L = 10\log\int_{z_t}^{z_b} s_v\, dz, \quad (3)$$

where z is depth in meters, and z_t and z_b are the depths of the top and bottom of the layer.[5] When more than one layer existed at a location, integration over all layers produced a total layer scattering strength (S_T) for each frequency band. Curves of S_L and S_T versus frequency were then produced.

III. RESULTS

The S_L and S_T curves obtained at each station constitute the results of the experiment. Various combinations of these curves are presented in this section.

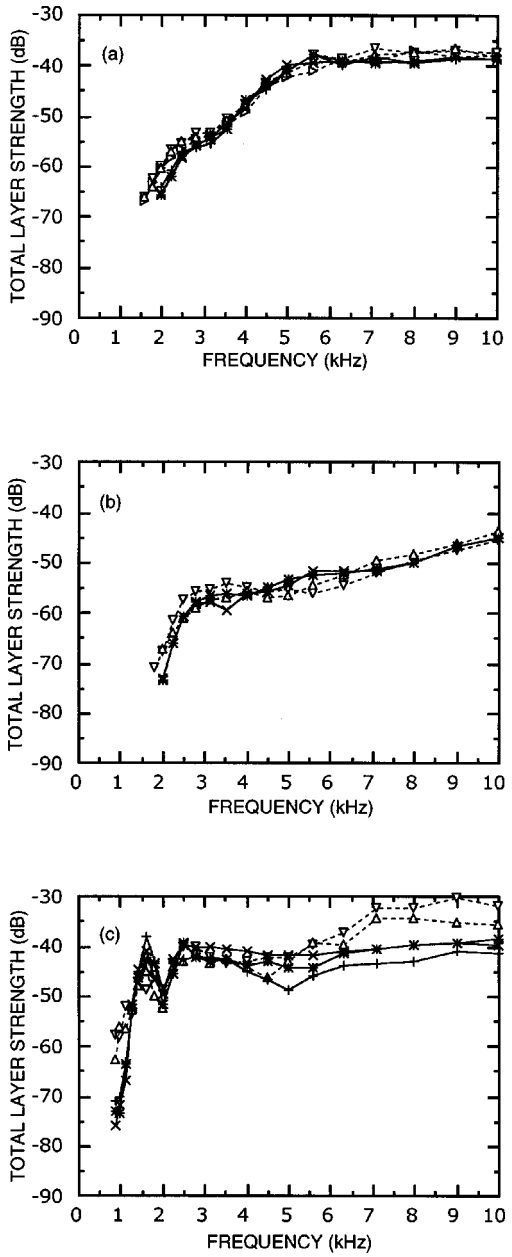


FIG. 2. Night-time total layer strengths for individual shots at (a) Station 1, (b) Station 2 on January 11, (c) Station 3. Solid lines with crosses denote F78 shots and dashed lines with triangles denote F55 shots.

A. Variability

Figure 2 shows night-time S_T curves at each station. Figure 2(a) indicates that shot-to-shot variability in S_T was small for both receivers at Station 1. To quantify this variability, the absolute value of the differences in S_T for any two shots was determined for each frequency and pair of shots and an average absolute difference calculated. The average absolute difference in S_T between any two shots was 0.8 dB for both receivers. Differences between the F78 and F55 were also small. The average absolute difference between the averages for each receiver was 1.6 dB. If the two frequency bands below 2.5 kHz are neglected, the average difference between the F78 and F55 was only 1.1 dB. Average differences in S_T were also small for the day-time measurements at Station 1. Thus, results obtained using the F78 are essentially equivalent to those obtained using the F55. This is particularly true at frequencies above 2.5 kHz, where their designs are virtually identical. Hence, any differences in results obtained with the F78 and F55 at other stations can be attributed to something other than the receivers.

Variability in S_T during a measurement sequence at Station 2 was about the same as at Station 1. Figure 2(b) shows results for the night of 11 January. Average absolute differences in S_T between shots were 0.6 dB for the F78 and 1.9 for the F55. The average absolute difference between the F78 and F55 was 1.7 dB; above 2.5 kHz, it was 1.2 dB. Differences during the day and during the night of 10 January were also small. However, there were clear differences between the two night-time sequences.

Variability in S_T during a measurement sequence at Station 3 was higher than at the deeper stations [Fig. 2(c)]. Average absolute differences in S_T between shots were just over 2 dB for both receivers. The average absolute difference between the F78 and F55 was almost 4 dB; for frequencies above 5 kHz, it was above 6 dB. Day-time variability was also higher at Station 3 than at the deeper stations.

Based on the shot-to-shot differences in S_T of about 2 dB or less, reverberation was considered to be stable over the 5 to 15 minutes of a measurement with one receiver at all three stations. Therefore, shots with a given receiver were averaged for each sequence. At Stations 1 and 2, reverberation was considered to be stable over periods of 1 h or more and shots with both receivers were averaged. Because results for the two night sequences at Station 2 were different, they were not averaged. At Station 3, there was noticeable variability over periods of 1 h or less and results from the two receivers were not averaged.

B. Averages

Average depths, layer strengths of the individual layers, and total layer strengths for Stations 1, 2, and 3 are shown in Figs. 3–5, respectively. Two to four significant layers were discerned for each station and time of day.

At Station 1 during the day, the strongest reverberation was from depths between 250 and 800 m. During the night, scatterers in the upper 200 m accounted for almost all the reverberation. Overall, reverberation was much higher during the night than during the day at all frequencies. Reverberation from the layer at 500 to 800 m changed very little from day to night.

At Station 2 during the day, the strongest reverberation at frequencies below 8 kHz was within 70 m of the bottom. Night-time reverberation in the upper 150 m at frequencies above 5 kHz was significantly stronger on 10 January than it was on 11 January. Night-time reverberation below 150 m was similar on the two nights. Night-time total layer strengths were somewhat to much greater than day-time strengths.

At Station 3, day-time reverberation in about 175 m of water was weak and predominantly from the upper 80 m. There were large changes in scattering strengths around 5 kHz over the 1.75 hour extent of the measurements. Night-time reverberation in about 200 m of water was predominantly from the upper 100 m at frequencies below 5 kHz.

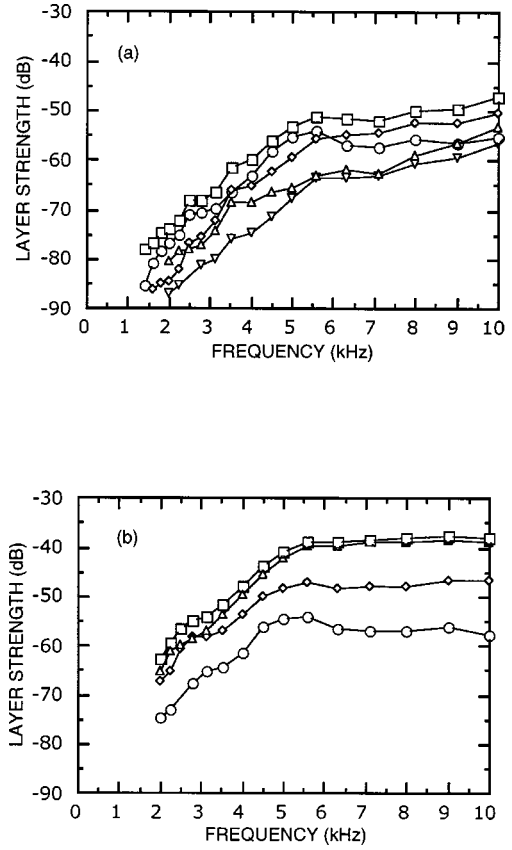


FIG. 3. Layer strengths at Station 1: (a) Day-time, square, S_T; triangle, S_L for 80–250 m; diamond, S_L for 250–500 m; circle, S_L for 500–800 m; inverted triangle, S_L for 800–1000 m. (b) Night-time, square, S_T; triangle, S_L for 40–200 m; diamond, S_L for 200–500 m; circle, S_L for 500–800 m.

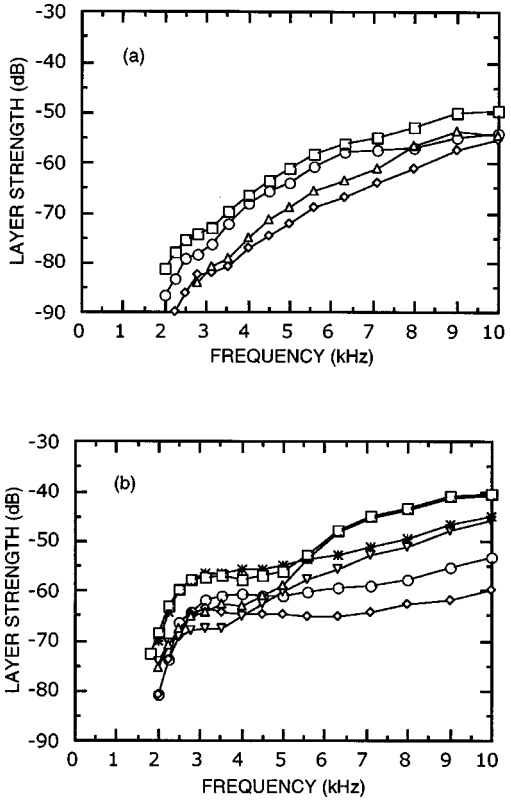


FIG. 4. Layer strengths at Station 2: (a) Day-time, square, S_T; triangle, S_L for 80–200 m; diamond, S_L for 270–320 m; circle, S_L for 380–450 m. (b) Night-time, square, S_T for January 10; asterisk, S_T for January 11; triangle, S_L for 50–150 m for January 10; inverted triangle, S_L for 80–150 m for January 11; diamond, S_L for 280–320 m; circle, S_L for 360–420 m.

Above 5 kHz, there was substantial reverberation from all depths, with significant variations in scattering strengths below 100 m over the 0.5 hour extent of the measurements. The most striking feature of the reverberation at this station is the large increase in scattering strengths from day to night. Night-time total layer strengths were 15–30 dB greater than day-time strengths.

Figure 6 shows day-time and night-time total layer strengths for all three stations. The temporal variations seen at Stations 2 and 3 have been averaged in this figure. The day-time curves are fairly similar at all three stations and scattering strengths at most frequencies increase with increasing water depth. The night-time curves all differ, with the biggest difference being the high scattering strengths at low frequencies at Station 3. Figure 6 shows that there can be significant changes in volume reverberation moving from deep water onto the continental shelf.

IV. SWIMBLADDER SCATTERING MODELING

In the Introduction, it was stated that one should expect the character of volume reverberation to change from deep to shelf waters and expect volume reverberation over the shelf to be more variable on the shelf than in deep water. The acoustic results fulfill these expectations. It was also stated that these differences were due to differences in fish populations in deep and shelf waters. In this section, a swimbladder scattering model is employed to examine possible differences in fish populations.

A. Swimbladder model

Swimbladders are essentially air bubbles that generally resemble prolate spheroids with major-to-minor axis ratios up to about 10. Omnidirectional scattering from the swimbladder dominates scattering from a fish at frequencies less than to somewhat above the swimbladder's resonance frequency. The swimbladder scattering model employed in this analysis assumes an air-filled sphere inside a viscous spherical shell.[8] The spherical assumption is reasonable; scattering functions of spheroids with aspect ratios of 10 are not appreciably different from those of spheres except for a small increase in resonance frequency.[9,10] The resonance frequency (f_0) is dependent on swimbladder radius (r) and depth,

$$f_0^2 = (3\gamma P)/(4\pi^2 r^2 \rho), \tag{4}$$

where P is the ambient pressure, γ is the ratio of specific heats of air (γ=1.4), and ρ is the density of fish flesh (ρ =1050 kg/m^3). The present analysis includes a 10% increase in resonance frequencies to account for the spheroidal nature of swimbladders. The model calculates the acoustic cross-section (σ) of a swimbladder fish as a function of acoustic frequency (f):

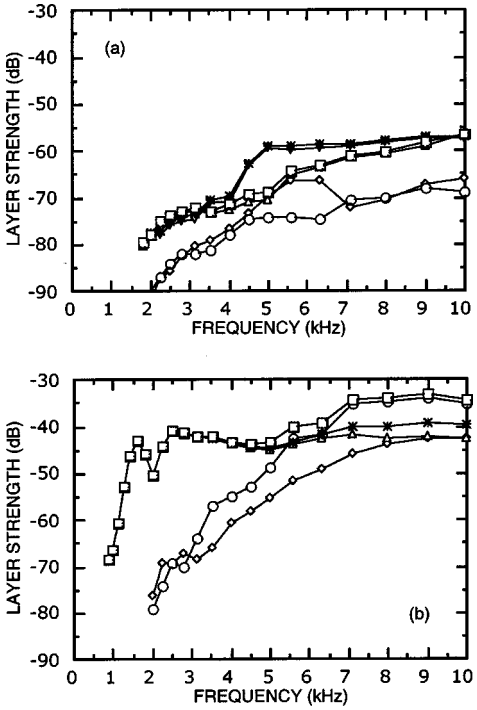


FIG. 5. Layer strengths at Station 3: (a) Day-time, square, S_T at 1330; asterisk, S_T at 1500; triangle, S_L for 50–80 m at 1330; inverted triangle, S_L for 50–80 m at 1500; circle, S_L for 80–175 m at 1330; diamond, S_L for 80–175 m at 1500. (b) Night-time, square, S_T at 2030; asterisk, S_T at 2100; triangle, S_L for 50–100 m; circle, S_L for 100–200 m at 2030; diamond, S_L for 100–205 m at 2100.

$$\sigma = \frac{4\pi r^2}{[(f_0^2/f^2)-1]^2+[(2\pi r f/c)+(\xi/\pi r^2 \rho f)]^2}, \tag{5}$$

where ξ is the viscosity of fish flesh.[8] The second bracketed term in the denominator is a damping term. The cross sections of a distribution of fish in a layer are summed to produce a layer strength

$$S_L = 10 \log \sum_{i=1}^{N} \sigma_i + 10 \log(\Phi/N), \tag{6}$$

where σ_i is the acoustic cross section of the ith fish and Φ is the number of fish within a 1 m^2 vertical column that extends over the depth of the layer. The modeled layer strength is

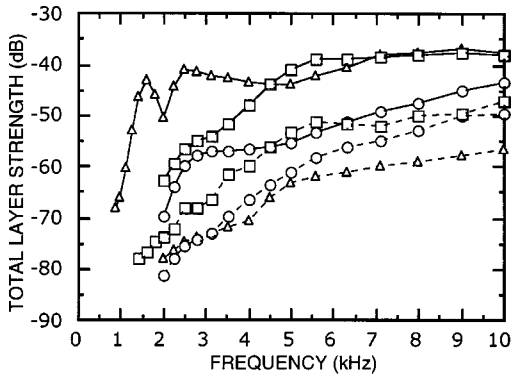


FIG. 6. Average total layer strengths at all stations. Solid lines, night-time; dashed lines, day-time. Squares, Station 1; circles, Station 2; triangles, Station 3.

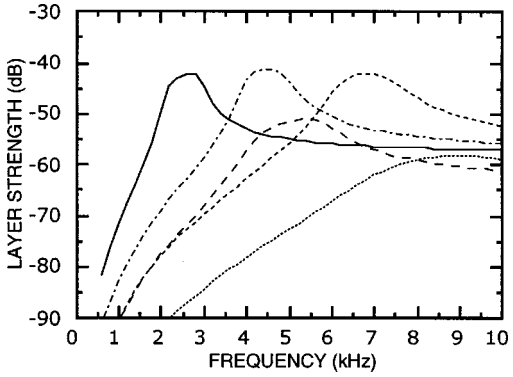


FIG. 7. Swimbladder model results for various swimbladder sizes and depths with $\Phi = 0.1/\text{m}^2$. Dotted line, $r = 0.12$ cm at 50–100 m; long dashes, $r = 0.20$ cm at 50–100 m; solid line, $r = 0.40$ cm at 50–100 m; dotted–dashed line, $r = 0.40$ cm at 200–300 m; short dashed line, $r = 0.40$ cm at 500–700 m.

then compared to acoustic data. Three parameters, r, Φ, and z, can be varied to provide a "fit" of the model to the data. The shape of the model curve is determined by r and z and its level is determined by Φ. However, the depth range, z, is determined from the acoustic measurements and is considered fixed, so that only the swimbladder size distribution is varied.

Figure 7 shows the results of varying z while holding r and Φ constant, and the results of varying r while holding z and Φ constant. In all cases, $\Phi = 0.1$ fish/m^2. Since most of the scatterers were expected to be small mesopelagics, a value of $\xi = 10$ Pa s was deemed reasonable.[8] Each curve has a resonance peak, with a region of Rayleigh scattering at frequencies below resonance and a region of geometric scattering above. Keeping swimbladder size constant while increasing depth increases the resonance frequency but has essentially no effect on the level of the resonance peak. Decreasing the size of the swimbladder while keeping the fish at the same depth not only increases the resonance frequency but also dramatically reduces the level of the resonance peak. The difference in peak levels for $r = 0.4$ cm and $r = 0.12$ cm is about 16 dB. Equation (2) indicates that about 10 dB $[20 \log(0.4/0.12) = 10]$ of that difference can be attributed to size difference; the other 6 dB is caused by greater damping for the smaller swimbladders. This 16 dB difference implies that 40 fish with $r = 0.12$ cm would be required to produce the same peak level as 1 fish with $r = 0.4$ cm.

It should be noted that scattering from the bodies of fish without swimbladders and other creatures such as squids can produce significant reverberation if they are large and/or abundant enough. For these animals, there is a Rayleigh scattering region at low frequencies, a geometric scattering region at high frequencies, and a transition region between. The S_L curves in Figs. 3–6 that increase with increasing frequency could be caused by these animals. However, unlike scattering from swimbladders, scattering from these animals does not change with depth, so that reverberation caused by them should not vary from day to night. Since the S_L curves at each station changed to some degree from day to night as the layers rose in the water column, it was as-

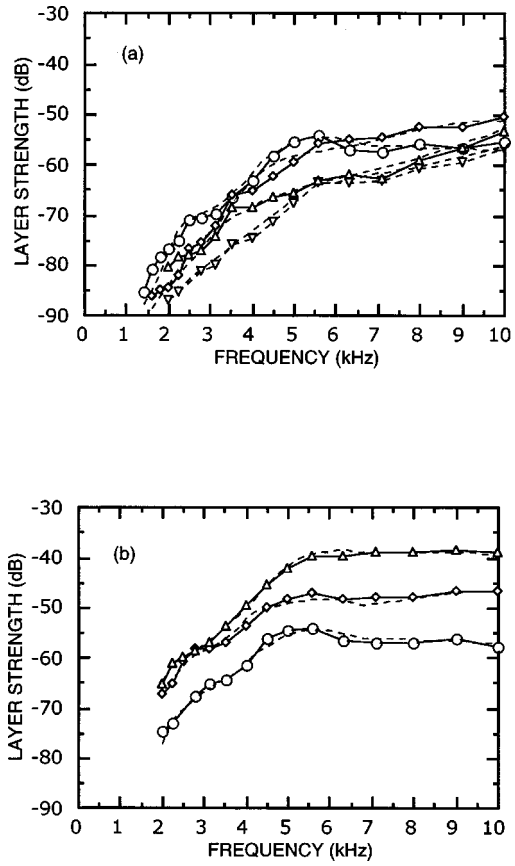


FIG. 8. Data/model S_L comparisons for Station 1 layers. Solid lines with symbols denote data and dashed lines denote model results. (a) Day-time, triangle, 80–250 m; diamond, 250–500 m; circle, 500–800 m; inverted triangle, 800–1000 m. (b) Night-time, triangle, 40–200 m; diamond, 200–500 m; circle, 500–800 m.

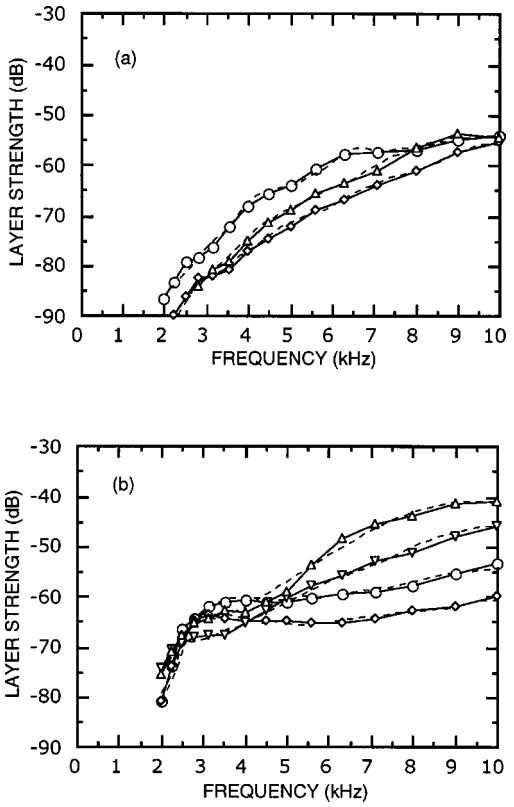


FIG. 9. Data/model S_L comparisons for Station 2 layers. Solid lines with symbols denote data and dashed lines denote model results. (a) Day-time, triangle, 80–200 m; diamond, 270–320 m; circle, 380–450 m. (b) Night-time, triangle, 50–150 m for January 10; inverted triangle, 80–150 m for January 11; diamond, 280–320 m; circle, 360–420 m.

sumed that animals without swimbladders were not important scatterers at any station.

B. Data-model comparison

The viscous spherical shell swimbladder scattering model has been used successfully in several widely different ocean areas.[3,11–13] In each of these instances, there was from fair to very good information on a few candidate fish species. Thus, swimbladder radii were determined from knowledge of fish sizes and fish size to swimbladder size relationships. In the Gulf of Mexico, numerous species could be responsible for scattering at 10 kHz. We have no *a priori* knowledge of fish sizes and, therefore, have employed the swimbladder model in an iterative mode, varying both r and Φ until reasonable fits to the data were obtained. Iterations included efforts to develop similar day and night size distributions for a given station.

Figures 8, 9, and 10 show data-model comparisons for the three stations. Tables I–III give the size distributions that produced the model curves. The figures show that, overall, the iteratiave process can produce a good fit of the model to the data. Minor differences occur where the model cannot change as rapidly with frequency as the data. These differences do not significantly affect the interpretation of the model results. The iterative process does not produce a unique solution. Size distributions different from those in the tables could produce equally reasonable fits to the data, but the overall character of the distributions would have to be the same and would not result in any changes in the interpretation of the results.

At Station 1 at night, a large number of small swimbladders is required to match most of the S_L curve for the shallow layer. Many fewer larger swimbladders are needed to match the middle layer. Very few somewhat larger swimbladders are needed to match the deep layer. For each layer, the relatively few large swimbladders listed in (b) of Table I are needed to produce the small bumps in the S_L curves below 3 or 4 kHz.

During the day at Station 1, the shallow layer still requires a relatively large number of small swimbladders ($r \leqslant 0.15$ cm) to match the high frequency ($f \geqslant 7$ kHz) portion of the S_L curve; very few slightly larger swimbladders are needed to match the curve between 4 and 7 kHz. Fewer larger swimbladders are needed to match most of the mid-layer. There was essentially no change in scattering from the deep layer from night to day; therefore, night and day swimbladder size distributions are essentially identical. The very deep layer effectively absorbs the large number of small swimbladders required for the shallow night-time layer; the curve was matched with a large number of small swimbladders and a few larger ones. The large number of very small swimbladders ($r \geqslant 0.15$ cm) resonate above 20 kHz and, therefore, have no effect. In fact, if having relatively equal overall night and day size distributions was not an objective,

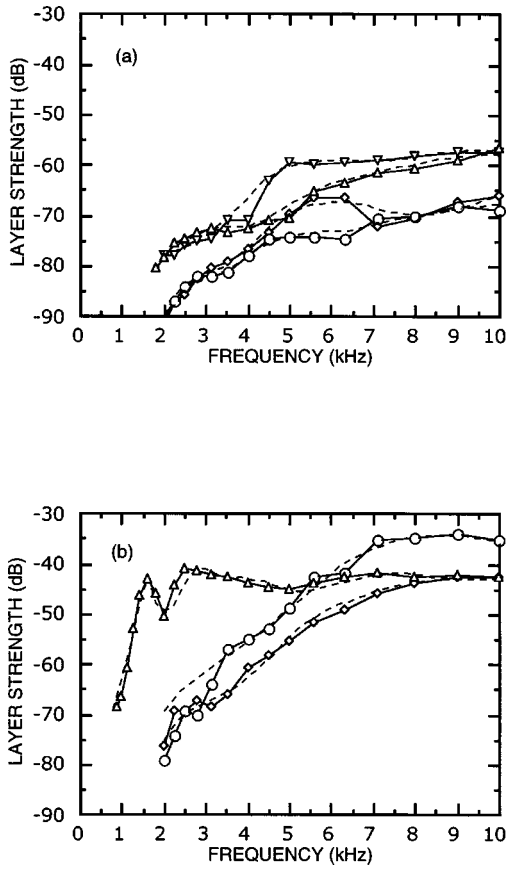


FIG. 10. Data/model S_L comparisons for Station 3 layers. Solid lines with symbols denote data and dashed lines denote model results. (a) Day-time, triangle, 50–80 m at 1330; inverted triangle, 50–80 m at 1500; circle, 80–175 m at 1330; diamond, 80–175 m at 1500. (b) Night-time, triangle, 50–100 m; circle, 100–200 m at 2030; diamond, 100–205 m at 2100.

the S_L curve for the very deep layer could have been matched with a very few ($\Phi=0.02$) larger ($0.3\leqslant r\leqslant 0.6$ cm) swimbladders. As at night, the relatively few largest swimbladders listed in (a) of Table I for each layer are needed to produce the small bumps in the S_L curves below 3 or 4 kHz.

The primary purpose of a swimbladder is to enable a fish to be neutrally buoyant (weightless in water). To achieve neutral buoyancy requires that a fish's swimbladder volume be about 5% of its total volume.[14] For "normally-shaped" fishes, this results in fish length (L) to swimbladder radius ratios of between 20 and 25. Mesopelagic swimbladders are usually less than 5% of fish volume and L/r is between 25 and 30 for most mesopelagics.[15,16] One exception is the deep-bodied hatchetfishes, for whom L/r is closer to 15.[16]

The many small fish in the night-time shallow layer were probably 2.5 to 6 cm long. Most of these fish migrated to day-time depths below 800 m. Between 200 and 500 m, the fish were primarily 4 to 8 cm both day and night. Hatchetfish are known to remain at the same depths day and night and have been caught in the eastern Gulf of Mexico at depths below 400 m.[17,18] Therefore, the 500 to 800 m layer most likely consisted of 3 to 9 cm long hatchetfish.

At Station 2 at night, scattering at 3 to 4 kHz cannot be considered as no more than a bump in the Rayleigh scattering region of the high frequency scatterers, as it was at Station 1. This is not because night-time scattering at 3 to 4 kHz was higher at Station 2 than at Station 1, but because scattering above 5 kHz was lower at Station 2 than at Station 1. The night-time shallow layers at Station 2 differed for the two nights. On 10 January, the strongest scattering was concentrated between 60 and 90 m; on 11 January, it was between 100 and 120 m. Matching the S_L curve for 10 January requires many small swimbladders ($r\leqslant 0.15$ cm) for frequencies above 4 kHz and very few larger ones below 4 kHz. For 11 January the mix is a little different; fewer, slightly larger swimbladders are needed. The small peak around 3 kHz in the S_L curve for the mid-layer is produced by very few large swimbladders ($r\geqslant 0.41$ cm); a small number of smaller swimbladders ($0.16\leqslant r\leqslant 0.4$ cm) causes the scattering above 4 kHz. The S_L curve for the deep layer is similar to that of the mid-layer. A very few large swimbladders ($r\geqslant 0.41$ cm) are required for the broad peak around 4 kHz and many more smaller swimbladders ($0.16\leqslant r\leqslant 0.4$ cm) are needed for higher frequencies.

The S_L curve for the day-time shallow layer at Station 2 can be matched by a few small swimbladders. The S_L curves for the mid and deep layers require fewer, slightly larger, swimbladders. The large number of small swimbladders in the night-time shallow layer could have migrated to either the day time mid or deep layers; their contribution to either is insignificant because their resonance is above 15 kHz. They have been arbitrarily placed in the deep layer. For the mid and deep layers, the largest swimbladders are needed only to produce the small bumps in the S_L curves around 2.5 kHz.

The natures of the S_L curves at Stations 1 and 2 are similar and swimbladder size distributions are generally comparable. Hence, it is probable that mesopelagic fishes were the cause of scattering at Station 2 also. The large number of fish required to cause scattering at the higher frequencies of the night-time shallow layer were probably 2.5 to 4.5 cm long. Most of these fish migrated to the day time mid or deep layers. The lower frequency scatterers in the night-time shallow layer were primarily between 4 and 12 cm long. Significant fish scatterers in the mid and deep day-time layers were between 4 and 12 cm long, while the mid and deep night-time layers had fish between 4 and 24 cm long. These larger fish either came off bottom at night or there was geographic variability that may or may not have been related to depth. (The day site was about 450 m deep and the night sites were about 420 m.)

Many fishes have closed swimbladders which they can actively maintain at constant volume at different depths, thereby remaining neutrally buoyant during the day and night. Other fishes have swimbladders that are open to the stomach; these passively expand or compress with changing depth. Since many mesopelagics maintain constant volume,[14] the initial assumption was that all the fish at Stations 1 and 2 did so. Because night-time fish size distributions are similar to day distributions at both stations, it appears that most, if not all, of the larger fishes did maintain constant volume. If the large numbers of small fish comprising the night-time shallow layers let their swimbladders compress, their resonance frequencies would be even further above 10 kHz than under the constant volume assumption. Since, under either scenario, day-time scattering from these small fish does not

TABLE I. (a) Swimbladder radii used to model Station 1 day-time layers. (b) Swimbladder radii used to model Station 1 night-time layers.

(a)

Radius range cm	Number (Φ) Shallow 80–250 m	Mid 250–500 m	Deep 500–800 m	Very deep 800–1000 m	Total
0.1–0.15	1.4	0	0	5.6	7
0.16–0.2	0.0075	0.065	0	2.6	2.6725
0.21–0.25	0.0017	0.033	0.002	0.065	0.1017
0.26–0.3	0.0005	0.006	0.005	0.002	0.0135
0.31–0.4	0.000 15	0.003	0.004	0.0005	0.007 65
0.41–0.5	0.000 01	0.0008	0.004	0.0004	0.005 21
0.51–0.6	0	0	0.001	0.0002	0.0012
0.61–0.8	0	0	0.000 07	0.000 013	0.000 083
0.81–1.0	0	0	0.000 015	0	0.000 015
1.01–1.2	0	0	0.000 01	0	0.000 01
Total	1.409 86	0.1078	0.016 095	8.268 113	9.801 868

(b)

Radius range cm	Number (Φ) Shallow 60–100 m	Mid 200–500 m	Deep 500–800 m	Total
0.1–0.15	7	0	0	7
0.16–0.2	2.4	0.24	0	2.64
0.21–0.25	0.0017	0.1	0.002	0.1039
0.26–0.3	0.0005	0.016	0.005	0.0215
0.31–0.4	0.0006	0.03	0.004	0.0346
0.41–0.5	0.0003	0.0012	0.004	0.0055
0.51–0.6	0	0.0006	0.001	0.0016
0.61–0.8	0	0.000 35	0.0001	0.000 45
0.81–1.0	0	0	0.000 035	0.000 035
1.01–1.2	0	0	0.000 01	0.000 01
Total	9.4031	0.388 15	0.016 145	9.807 395

influence the results below 10 kHz, their swimbladder behavior is not a factor.

The S_L curve for the night-time shallow layer at Station 3 was very different from that of any other layer. This layer was the only one that had scattering concentrated over different depth ranges at different frequencies. The sharp peak at 1.6 kHz was between 50 and 70 m. Modeling indicates that very few swimbladders of a single relatively large size ($r=0.58$ cm) are needed to produce this peak. The broad peak at about 2.5 to 4 kHz was between 60 and 90 m. This peak requires a few swimbladders of a broad size range ($0.16 \leqslant r \leqslant 0.5$ cm). Higher frequency scattering was diffused over the whole layer. A large number of small swimbladders ($r \leqslant 0.15$ cm) is required to produce the S_L curve above 6 kHz. The scatterers comprising the night-time deep layer changed over about 30 minutes. At 2100, a relatively large number of small swimbladders ($r \leqslant 0.25$ cm) produces most of the S_L curve, with very few larger swimbladders needed to produce the small bump between 2 and 3 kHz. At 2030, a very large number of small swimbladders ($r \leqslant 0.2$ cm) produces the complete S_L curve, contributions from larger swimbladders are insignificant. In fact, Rayleigh scattering from the small swimbladders does not decrease fast enough with decreasing frequency to be able to match the data below 3.5 kHz. This difference between model and data below 3.5 kHz is probably due to errors in the data. Using a downward-looking measurement technique, errors in the calculation of S_V are inevitable for a weak layer below a much stronger layer. This is generally not a problem, since depths of strong scattering are normally those of interest.

The high and low frequency portions of the 1330 and 1500 S_L curves for the day-time layers at Station 3 are about the same; differences occur between about 4.5 and 6.5 kHz. For the shallow layer, the high frequency portion of each curve requires a few very small swimbladders ($r \leqslant 0.15$ cm) and the low frequency portion requires a very few larger swimbladders ($r \geqslant 0.21$ cm). From 4.5 to 6.5 kHz for the shallow layer, matching the 1500 curve requires significantly more small swimbladders ($0.16 \leqslant r \leqslant 0.2$ cm) than the 1330 curve. For the deep layer, the high frequency portion of each curve requires a few small swimbladders ($r \leqslant 0.2$ cm) and the low frequency portion requires very few larger swimbladders ($r \geqslant 0.26$ cm). From 5 to 6.5 kHz for the deep layer, matching the 1500 curve again requires significantly more small swimbladders ($0.21 \leqslant r \leqslant 0.25$ cm) than the 1330 curve.

The fish that produced the 1.6 kHz peak in the S_L curve for the night-time shallow layer at Station 3 were probably 12 to 15 cm long, assuming that their swimbladder volumes were about 5% of fish volume. Trawling conducted by the National Marine Fisheries Service (NMFS) shows that two species in this size range, rough scad (*Trachurus lathami*)

TABLE II. (a) Swimbladder radii used to model Station 2 day-time layers. (b) Swimbladder radii used to model Station 2 night-time layers.

(a)

Radius range cm	Number (Φ) Shallow 80–200 m	Mid 270–320 m	Deep 380–450 m	Total
0.1–0.15	0.15	0	6.05	6.2
0.16–0.2	0.001	0.024	0.026	0.051
0.21–0.25	0.0005	0.0025	0.012	0.015
0.26–0.3	0	0.0005	0.004	0.0045
0.31–0.4	0	0.000 11	0.0033	0.003 41
0.41–0.5	0	0.000 008	0.0002	0.000 208
0.51–0.6	0	0.000 0015	0.000 06	0.000 0615
0.61–0.8	0	0.000 001	0.000 004	0.000 005
0.81–1.0	0	0	0	0
1.01–1.2	0	0	0	0
Total	0.1515	0.027 1205	6.095 564	6.274 1845

(b)

Radius range cm	Number (Φ) Shallow 60–90 m 10th	Shallow 100–120 m 11th	Mid 280–320 m	Deep 360–420 m	Total 10th	Total 11th
0.1–0.15	6.2	1.2	0	0	6.2	1.2
0.16–0.2	0.001	0.025	0.007	0.03	0.038	0.062
0.21–0.25	0.0001	0.0017	0.0017	0.011	0.0128	0.0144
0.26–0.3	0.0001	0.0004	0.0007	0.003	0.0038	0.0041
0.31–0.4	0.000 35	0.000 24	0.000 45	0.002	0.0028	0.002 69
0.41–0.5	0	0.000 07	0.0002	0.0006	0.0008	0.000 87
0.51–0.6	0	0.000 02	0.000 16	0.0004	0.000 56	0.000 58
0.61–0.8	0	0	0.000 13	0.000 25	0.000 38	0.000 38
0.81–1.0	0	0	0	0	0	0
1.01–1.2	0	0	0	0	0	0
Total	6.201 55	1.227 43	0.010 34	0.047 25	6.259 14	1.285 02

and round herring (*Etrumeus teres*), are common in the vicinity of Station 3.[4] Both species are winter residents of the outer continental shelf. Rough scad have closed swimbladders and herrings have open swimbladders. Thus, rough scad swimbladders would be expected to be 5% of fish volume and should resonate around 1.6 kHz. The picture for fish with open swimbladders is unclear. One hypothesis is that such fish are neutrally buoyant at night. If this is so, round herring should also resonate around 1.6 kHz. Another hypothesis is that herring rise to the surface and slightly overinflate their swimbladders, but the swimbladders are then compressed at both night-time and day-time depths. In the latter case, round herring swimbladders would resonate at higher frequencies. In fact, if 12 to 15 cm round herring swimbladders are 5% of fish volume at the surface, they would compress to resonate between 2.5 and 4 kHz at 60 to 90 m. If the fish causing the broad peak at 2.5 to 4 kHz were neutrally buoyant, they would have been 5 to 10 cm long. NMFS trawl data do not indicate any fish of this size range in the vicinity of Station 3. Thus, based on the NMFS trawls, the supposition is that rough scad caused the peak at 1.6 kHz and round herring caused the broad peak at 2.5 to 4 kHz. NMFS trawls also caught large numbers of pearlsides (*Maurolicus muelleri*) in the vicinity of Station 3.[4] These fish are about 2 to 6 cm long, which is the size range required for neutrally buoyant fish to produce the high frequency portion of the S_L curve for the shallow layer and the complete S_L curve for the deep layer. Thus, the supposition is that pearlsides, which are common permanent residents of the outer shelf, were significant scatterers throughout the water column at Station 3 at night.

The fish required to produce the day-time S_L curves at Station 3 are the same sizes as those that produced the night-time curves. However, many more fish of every size are required at night. Since night layers and day layers were at about the same depths, night fish cannot be hidden in day layers. One possible scenario is that the fish were on the bottom during the day. However, scad, herring, and pearlsides are not day-time bottom dwellers. Another possibility is geographic and/or depth differences between the day and night measurement locations, which were 2 nm apart, with the day measurement in about 175 m depth and the night in about 200 m. Recent studies in the northern Gulf of Mexico have shown that fish on the outer shelf can be quite patchy and restricted to narrow depth ranges.[19] Also, pearlsides are much more abundant at 200 m than they are at 100 m.[4] Thus, patchiness and/or depth change could account for day/night differences in the number of fish. One possibility is that a few pearlsides are the primary day-time scatterers throughout the water column and very few round herring near the bottom are responsible for the peak around 6 kHz in the S_L curve for the deep layer at 1500. If an absence of the larger scatterers was due to patchiness, then either or both species

TABLE III. (a) Swimbladder radii used to model Station 3 day-time layers. (b) Swimbladder radii used to model Station 3 night-time layers.

(a)
Number (Φ)

Radius range cm	Shallow 50–80 m (1330)	Shallow 50–80 m (1500)	Deep 80–175 m (1330)	Deep 80–175 m (1500)	Total (1330)	Total (1500)
0.1–0.15	0.19	0.21	0.023	0.03	0.213	0.24
0.16–0.2	0.004	0.023	0.0007	0.0007	0.0047	0.0237
0.21–0.25	0.000 13	0.000 18	0.0002	0.0011	0.000 33	0.001 28
0.26–0.3	0.000 12	0.000 07	0.000 04	0.000 04	0.000 16	0.000 11
0.31–0.4	0.000 045	0.000 02	0.000 007	0.000 007	0.000 052	0.000 027
0.41–0.5	0.000 012	0.000 01	0.000 004	0.000 004	0.000 016	0.000 014
0.51–0.6	0.000 001	0.000 001	0	0	0.000 001	0.000 001
0.61–0.8	0	0	0	0	0	0
0.81–1.0	0	0	0	0	0	0
1.01–1.2	0	0	0	0	0	0
Total	0.194 308	0.233 281	0.023 951	0.031 851	0.218 259	0.265 132

(b)
Number (Φ)

Radius range cm	Shallow 50–100 m	Deep 100–200 m (2030)	Deep 100–200 m (2100)	Total (2030)	Total (2100)
0.1–0.15	5	3	0.8	8	5.8
0.16–0.2	0.01	6	0.7	6.01	0.71
0.21–0.25	0.14	0.05	0.05	0.19	0.19
0.26–0.3	0.11	0.000 01	0.000 01	0.110 01	0.110 01
0.31–0.4	0.13	0.000 01	0.000 01	0.130 01	0.130 01
0.41–0.5	0.01	0.000 05	0.000 05	0.010 05	0.010 05
0.51–0.6	0.02	0.000 01	0.000 01	0.020 01	0.020 01
0.61–0.8	0	0	0	0	0
0.81–1.0	0	0	0	0	0
1.01–1.2	0	0	0	0	0
Total	5.42	9.050 08	1.550 08	14.470 08	6.970 08

could have been at a slightly different day-time measurement location and caused much higher scattering at the lower frequencies.

V. CONCLUSIONS

Measurements in deep water and over the slope and outer shelf in the eastern Gulf of Mexico demonstrate that the character of volume reverberation changes from deep to shallow water. Volume reverberation in shallower waters varies more rapidly with respect to both time and space than that in the deep ocean. Day-time total layer strength curves were similar for all three locations, generally increasing with frequency from 2 to 10 kHz. Scattering modeling indicates that day-time reverberation at all three locations was caused by small swimbladder-bearing fishes, the vast majority of which were smaller than 10 cm. Night-time total layer strength curves for the deep ocean and slope also increased with increasing frequency. Essentially the same fish were responsible for both day-time and night-time reverberation at these two locations. These fish were almost certainly mesopelagic species, the dominant scatterers in most deep ocean areas. Night-time scattering at the outer shelf location was very different. The night-time total layer strength curve at this location had sharp peaks at 1.6 and around 2.5 kHz, with comparable scattering strengths at high frequencies. Results of extensive fisheries trawling in the eastern Gulf implicate rough scad, round herring, and pearlsides as possible species responsible for scattering at 1.6 kHz, 2.5 kHz, and higher frequencies, respectively. Pearlsides could also have dominated the day-time scattering at the outer shelf location. Much of the difference between day-time and night-time scattering seen on the outer shelf was probably due to patchiness in the distributions of the larger scatterers; a slightly different day-time location might have had significantly higher scattering.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. Assistance at sea was provided by the Captain and crew of *USNS Bartlett*.

[1] *Proceedings of an International Symposium on Biological Sound Scattering in the Ocean*, Rep. 005, edited by G. B. Farquhar (Maury Center for Ocean Science, Washington, DC, 1970).
[2] *Oceanic Sound Scattering Prediction*, edited by N. R. Andersen and B. J. Zahuranec (Plenum, New York, 1977).
[3] R. H. Love, "A comparison of volume scattering strength data with model calculations based on quasisynoptically collected fishery data," J. Acoust. Soc. Am. **94**, 2255–2268 (1993).
[4] J. S. Link, T. A. Henwood, and C. T. Gledhill, "Small pelagics in the Gulf of Mexico: a description of the abundance, geographic distribution, size,

and depth distibution of major pelagic species,'' American Fisheries Society Symposium Proceedings (in press).
[5] R. J. Urick, *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York, 1983).
[6] S. Machlup and J. B. Hersey, ''Analysis of sound-scattering observations from non-uniform distributions of scatterers in the ocean,'' Deep-Sea Res. **3**, 1–22 (1955).
[7] O. Z. Bluy, ''Transducer sidelobe corrections for deep scattering layer data,'' Deep-Sea Res. **33**, 397–399 (1970).
[8] R. H. Love, ''Resonant acoustic scattering by swimbladder-bearing fish,'' J. Acoust. Soc. Am. **64**, 571–580 (1978).
[9] D. E. Weston, ''Sound propagation in the presence of bladder fish,'' in *Underwater Acoustics*, edited by V. M. Albers (Plenum, New York, 1967), Vol. 2, Chap. 5, pp. 55–88.
[10] C. Feuillade and M. F. Werby, ''Resonances of deformed gas bubbles in liquids,'' J. Acoust. Soc. Am. **96**, 3684–3692 (1994).
[11] R. W. Nero and M. E. Huster, ''Low-frequency acoustic imaging of Pacific salmon on the high seas,'' Can. J. Fish. Aquat. Sci. **53**, 2513–2523 (1996).
[12] R. W. Nero, C. H. Thompson, and R. H. Love, ''Abyssopelagic grenadiers: the probable cause of low frequency scattering at great depths off the Oregon and California coasts,'' Deep-Sea Res. **44**, 627–645 (1997).
[13] R. W. Nero, C. H. Thompson, and R. H. Love, ''Low frequency acoustic measurements of Pacific hake, *Merluccius productus*, off the west coast of the United States,'' Fish. Bull. **96**, 329–343 (1998).
[14] N. B. Marshall, *The Life of Fishes* (World Publishing, Cleveland, 1966), Chap. 5, pp. 67–80.
[15] A. L. Brooks, ''A study of the swimbladders of selected mesopelagic fish species,'' in *Oceanic Sound Scattering Prediction*, edited by N. R. Andersen and B. J. Zahuranec (Plenum, New York, 1977), pp. 565–590.
[16] R. H. Gibbs, Jr., R. H. Goodyear, R. C. Kleckner, C. F. E. Roper, M. J. Sweeney, B. J. Zahuranec, and W. L. Pugh, ''Mediterranean Biological Studies, Final Report,'' Smithsonian Inst., Washington, DC, 1972.
[17] R. H. Gibbs and C. F. E. Roper, ''Ocean Acre: Preliminary report on vertical distribution of fishes and cephalopods,'' in *Proceedings of an International Symposium on Biological Sound Scattering in the Ocean*, Rep. 005, edited by G. B. Farquhar (Maury Center for Ocean Science, Washington, DC, 1970), pp. 119–133.
[18] R. C. Baird and D. F. Wilson, ''Sound scattering and oceanic midwater fishes,'' in *Oceanic Sound Scattering Prediction*, edited by N. R. Andersen and B. J. Zahuranec (Plenum, New York, 1977), pp. 549–563.
[19] R. W. Nero, ''Herring hydroglyphics in littoral waters of the northern Gulf of Mexico,'' J. Acoust. Soc. Am. **108**, 2489 (2000).

Time-dependent seafloor acoustic backscatter (10–100 kHz)[a)]

Daniel D. Sternlicht[b)] and Christian P. de Moustier[c)]
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0205

(Received 1 November 1999; revised 13 June 2003; accepted 26 June 2003)

A time-dependent model of the acoustic intensity backscattered by the seafloor is described and compared with data from a calibrated, vertically oriented, echo-sounder operating at 33 and 93 kHz. The model incorporates the characteristics of the echo-sounder and transmitted pulse, and the water column spreading and absorption losses. Scattering from the water–sediment interface is predicted using Helmholtz–Kirchhoff theory, parametrized by the mean grain size, the coherent reflection coefficient, and the strength and exponent of a power-law roughness spectrum. The composite roughness approach of Jackson *et al.* [J. Acoust. Soc. Am. **79**, 1410–1422 (1986)], modified for the finite duration of the transmitted signal, is used to predict backscatter from subbottom inhomogeneities. It depends on the sediment's volume scattering and attenuation coefficients, as well as the interface characteristics governing sound transmission into the sediment. Estimation of model parameters (mean grain size, roughness spectrum strength and exponent, volume scattering coefficient) reveals ambiguous ranges for the two spectral components. Analyses of model outputs and of physical measurements reported in the literature yield practical constraints on roughness spectrum parameter settings appropriate for echo-envelope-based sediment classification procedures. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1608018]

PACS numbers: 43.30.Gv, 43.30.Hw, 43.30.Ft, 43.30.Zk [DLB] Pages: 2709–2725

I. INTRODUCTION

In the typical bottom echo-sounding geometry, a sound pulse is transmitted by a sonar system and the time of arrival of the echo provides a measure of the altitude of the sonar above the bottom. However, the shape and duration of the echo are often very different from the original pulse and these distortions contain information about the seafloor acoustic backscattering process as well as the relief and geoacoustic properties of the bottom. By comparing a time-dependent physical model of the acoustic intensity backscattered by the bottom with data from a calibrated echo-sounder operating at 33 and 93 kHz, we intend to infer physical characteristics of the bottom such as the roughness of the water–sediment interface and the mean grain size of the sediment, and geoacoustic properties such as the sediment's volume scattering coefficient and its acoustic impedance.

The bottom echo intensity envelope model described in this paper is a temporal implementation of the SONAR equation,[1,2] based on acoustic backscatter models described in Jackson *et al.*,[3] and analytical tools developed by de Moustier and Alexandrou[4] for modeling seafloor echoes measured with multibeam seafloor mapping sonars. No new theories are presented; rather, a specific combination of physical acoustic backscatter models, geoacoustic parameters, and echo processing techniques is used to estimate the shape and intensity of the average bottom echo envelope measured with simple calibrated echo-sounders. Similar temporal models described by others include: (a) time-dependent representations of surface and volume scattering originally formulated in ''Physics of Sound in the Sea;''[1] (b) temporal seafloor scattering models for swath-mapping sonars developed in Morgera,[5] Morgera and Sankar,[6] and de Moustier and Alexandrou;[4] (c) comprehensive treatment of the time dependence of signals scattered by rough surfaces given by Berry,[7,8] Berry and Blackwell,[9] and Haines and Langston;[10] (d) and, most specifically, average echo envelope models by Nesbitt,[11] Jackson and Nesbitt,[12] de Moustier and Alexandrou,[4] and Pouliquen and Lurton,[13,14] based on combining energy backscattered from the sediment surface and subbottom. Our approach differs from those presented in Refs. 4, 11–14 in its attempt to model and match absolute sound-pressure echo levels measured with a fully calibrated sonar system. Most of the model parameters used are common to Refs. 1–14, with some variations due to choice of physical scattering models detailed herein; however, we propose a specific combination of these parameters in a numerical implementation that incorporates the digitized transmitted waveform, and formulations for reducing the number of geometric dimensions during synthesis of the average backscatter echo envelope. The attempt to obtain absolute backscatter levels related to measurable geoacoustic parameters sets this work apart from mostly phenomenological sediment characterization approaches used in commercial systems,[15,16] which rely on nonparametric classifiers to separate the various substrates, and require independent identification of the substrate by video or core sampling.

At the acoustic wavelengths of interest here (respectively, 4.5 and 1.6 cm at 33 and 93 kHz) the generalized Rayleigh criterion for diffuse scattering of a monochromatic

[a)]Parts of this manuscript were presented at the 137th meeting of the ASA [de Moustier and Sternlicht, J. Acoust. Soc. Am. **105**, 1080 (1999)].

[b)]Current address: Dynamics Technology Inc., 21311 Hawthorne Blvd., Suite 300, Torrance, CA 90503. Electronic mail: dsternlicht@dynatec.com
[c)]Current address: Center for Coastal and Ocean Mapping, University of New Hampshire, 24 Colovos Road, Durham, NH 03824. Electronic mail: cpm@ccom.unh.edu

sound wave, with wavelength λ_a and wave number $k_a = 2\pi/\lambda_a$, on a surface with rms height deviation σ about a mean plane

$$2k_a\sigma \gg 1, \tag{1}$$

implies that the rms height deviations σ must exceed 1 cm. This is satisfied in the data presented here for sandy and silty substrates in San Diego Bay, and we can use a modeling approach based entirely on incoherent scattering. Also, at these frequencies, penetration of the bottom is limited to the first few meters; therefore, assumptions of bottom homogeneity are more likely to be valid. Based on the assumptions that interface scattering dominates the return at normal incidence and volume scattering dominates at oblique incidence, temporal separation of these components is achieved by informed selection of transducer beamwidth and orientation.

The model incorporates the characteristics and geometry of the echo-sounder: the transducer's beam pattern, its altitude, its tilt with respect to vertical, the characteristics of the transmitted sound pulse, and the roll and pitch angles of the platform to which the transducer is mounted; and environmental factors: spherical spreading and absorption losses as the signal propagates through the water column, backscattering of the signal at the water–sediment interface, and by inhomogeneities in the sediment volume.

The mean grain size, defined as ($M_\phi = -\log_2 D_g$) where D_g represents the mean particle diameter in millimeters, is an important component of the sediment's compressional sound speed, saturated bulk density, and attenuation constant, and is of particular interest as it roughly correlates to these parameters through a set of experimentally derived linear regression equations developed by Hamilton[17] and refined by others.[18,19]

Following Jackson *et al.*,[3] Nesbitt,[11] Nesbitt and Jackson,[12] de Moustier and Alexandrou,[4] and Pouliquen and Lurton,[13] the acoustic backscattering at the water–sediment interface and in the sediment volume are modeled and computed independently, then summed to estimate the overall echo intensity measured by the echo-sounder (Fig. 1). Formally, the time-dependent intensity of the bottom backscattered acoustic signal $I(t)$ measured at the transducer's face equals the sum of the intensity $I_i(t)$ backscattered at the water–sediment interface and the intensity $I_v(t)$ backscattered from the sediment volume

$$I(t) = I_i(t) + I_v(t). \tag{2}$$

This paper provides a detailed description of the model and of its sensitivity to changes in its parameters. Data examples are presented with the echo alignment and ensemble averaging processes required to compare measured and modeled echo envelopes. Then, a metric is developed to quantify the closest fit between model and data from which seafloor geoacoustic parameters can be inferred, and the potential for ambiguous results is discussed. The implementation of this model for automatic geoacoustic parameter estimation is the subject of another paper by the authors.[20]

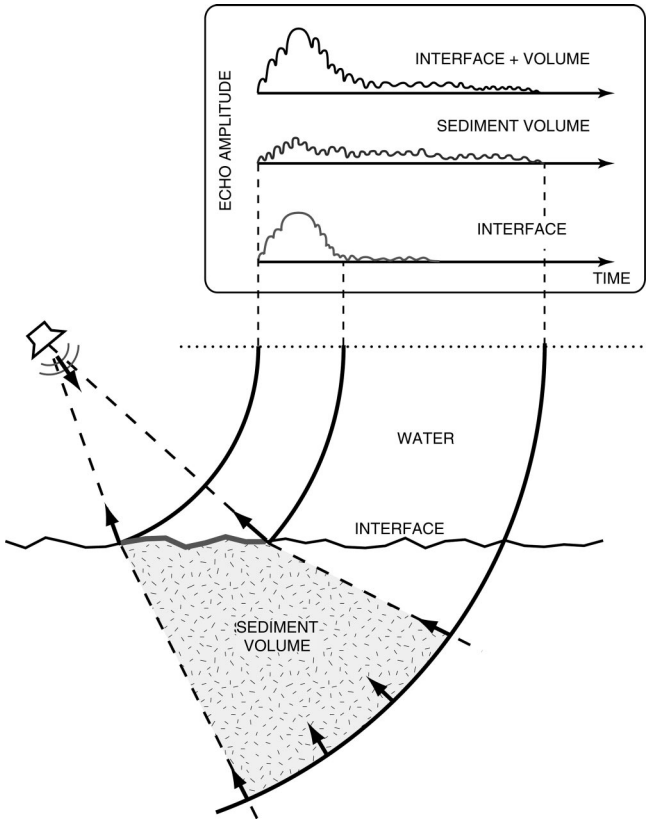


FIG. 1. Separation of acoustic backscatter contributions due to interface roughness and subbottom volume inhomogeneities.

II. MODEL

This section describes the components of Eq. (2), which define the temporal model of high-frequency acoustic seafloor backscatter.

A. Sediment interface characteristics

The analytical form of the signal component backscattered by the water–sediment interface is based on a solution of the Helmholtz diffraction integral for monochromatic sound waves. It uses the Kirchhoff approximation to express the pressure field at planes that are locally tangent to the interface, and the reflection coefficient, $\mathcal{R}(\theta_i)$ at angle of incidence θ_i, is assumed constant at each point on the interface and equal to the normal incidence reflection coefficient $\mathcal{R}_\perp = \mathcal{R}(0)$. The reflection coefficient is a function of the ratios of sediment saturated bulk density over water mass density (ρ) and sediment sound speed over water sound speed (ν)

$$\mathcal{R}(\theta_i) = \frac{\rho\nu\cos(\theta_i) - [1 - (\nu\sin(\theta_i))^2]^{1/2}}{\rho\nu\cos(\theta_i) + [1 - (\nu\sin(\theta_i))^2]^{1/2}}. \tag{3}$$

Given that we are concerned with acoustic backscatter measurements near normal incidence, shadowing and multiple scattering effects are neglected.

The relief statistics of the sediment interface are assumed to be isotropic and described by Gaussian-distributed height deviations (ζ) about a mean plane, with rms height σ, and by a 2D energy density spectrum. Since ζ is usually not stationary, and since the choice of a reference surface is

somewhat arbitrary, it is useful to consider the height difference between points on the surface separated by a fixed distance (horizontal vector $\mathbf{r}$). This height difference is a locally stationary random variable whose mean-squared value is the structure function $D(\mathbf{r})$

$$D(\mathbf{r}) = \langle [\zeta(\mathbf{r}+\mathbf{r}_o) - \zeta(\mathbf{r}_o)]^2 \rangle. \tag{4}$$

As described in Appendix B of Ref. 21, the relief energy density spectrum $W(\mathbf{k})$ for the 2D spatial wave vector $\mathbf{k}$ of magnitude k, and the structure function are related by

$$D(\mathbf{r}) = 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(1-\cos(\mathbf{k}\cdot\mathbf{r}))W(\mathbf{k})d^2k. \tag{5}$$

Power-law relief spectra have been measured over a wide range of spatial wavelengths, from geographic scales (meters–kilometers)[22–24] to centimeter scales commensurate with high-frequency acoustic wavelengths.[25–28] The relief energy density spectrum $W(\mathbf{k})$ is formulated as[3,19]

$$W(\mathbf{k}) = w_2 k^{-\gamma}, \tag{6}$$

where the spectral strength w_2 has units of length to the power $(4-\gamma)$, whereas the spectral exponent γ is unitless. For a reference length h_o, $W(\mathbf{k}) = w_2(h_o k)^{-\gamma}$. This removes dimensionality from the power term, and gives w_2 the dimensionality of h_o^4. In this work $h_o = 1$ cm, and w_2 is expressed in cm^4.

Spectral analyses of centimeter-scale topography have yielded 2D spectral exponents in the range $2<\gamma<4$ (Refs. 25–28). In the absence of measurements, a mean value of $\gamma = 3.25$ can be assumed for most bottom types.[19] In practice, γ and w_2 are determined by straight-line regression fit to the 2D spectra plotted on log–log scale, over spatial wave numbers spanning roughly an order of magnitude above and below the acoustic wave number.

Integration of Eq. (5), after substituting in Eq. (6), yields a simple expression for the structure function[3]

$$D(r) = C_\zeta^2 r^{2\alpha}, \tag{7}$$

with the structure constant C_ζ given by

$$C_\zeta^2 = [2\pi w_2 \Gamma(2-\alpha)2^{-2\alpha}]/[\alpha(1-\alpha)\Gamma(1+\alpha)], \tag{8}$$

where Γ is the gamma function, and $\alpha = (\gamma/2)-1$ is constrained between 0 and 1.

Following Jackson *et al.*,[3] the bandlimited power-law relief spectrum is separated into large- and small-scale roughness components with the boundary defined by a cutoff spatial wave number k_c

$$k_c = \left(\frac{\gamma-2}{8\pi w_2 k_a^2 \cos^2\theta_i}\right)^{1/(2-\gamma)}. \tag{9}$$

Within the bandlimited large-scale relief, we shall use also the mean-square slope ς^2 to compute sound energy transmission into the sediment, and the mean-square curvature R_c^{-2} for the Kirchhoff criterion

$$\varsigma^2 = 2\pi\int_0^{K_c} W(\mathbf{k})k^3 dk = \frac{2\pi w_2 k_c^{4-\gamma}}{4-\gamma}, \tag{10}$$

$$R_c^{-2} = 2\pi\int_0^{K_c} W(\mathbf{k})k^5 dk = \frac{2\pi w_2 k_c^{6-\gamma}}{6-\gamma}. \tag{11}$$

B. Sediment volume characteristics

At 10–100 kHz, penetration of water-saturated sediments is limited to the first several meters, and the bottom can be modeled as a lossy fluid with acoustic scattering due to fluctuations in the density and refraction indices of the medium. In addition, the intensity of a plane monochromatic sound wave entering the sediment is attenuated with distance. The corresponding attenuation coefficient α_b in dB/m is usually frequency dependent

$$\alpha_b = \kappa_p f_a^m, \tag{12}$$

where κ_p is an attenuation constant expressed in dB/m/kHz, and f_a is the acoustic frequency in kHz. For unconsolidated ocean sediments of the type presented here, we follow Hamilton's observations of $m=1$ and employ his regression equations[29] relating κ_p to the mean grain size (M_ϕ).

Ivakin and Lysanov[30] described a sediment volume backscatter model which includes the dissipative effects of a lossy medium and the transmission characteristics of a randomly rough interface. They used bottom slope variations to compute the energy transmitted through the interface. Jackson *et al.*[3] combined these ideas with Stockhausen's[31] scheme for subbottom acoustic scattering below a flat surface, by integrating transmission terms over a Gaussian distribution of incidence angles whose mean value was determined by the rms slope (ς). This approach is a variant of the composite roughness scattering model described by McDaniel and Gorman.[32] Here, we modify the composite roughness treatment to account for the finite duration of the transmitted signal.

C. Angle-dependent seafloor acoustic backscatter

Following the results of Jackson *et al.*,[3] the angular-dependent backscatter coefficient (s_i) at the water–sediment interface is given by

$$s_i(\theta_i) = \begin{cases} \mathcal{R}_\perp^2[8\pi\cos^2(\theta_i)\sin^2(\theta_i)]^{-1}\displaystyle\int_0^\infty \exp(-qu^{2\alpha}) \\ \quad \times J_o(u)u\,du & \theta_i > 0 \\ \mathcal{R}_\perp^2[8\pi\alpha]^{-1}C_\zeta^{-2/\alpha}(2k_a^2)^{(\alpha-1)/\alpha}\Gamma(1/\alpha) & \theta_i = 0, \end{cases} \tag{13}$$

with

$$q = \cos^2(\theta_i)\sin^{-2\alpha}(\theta_i)C_\zeta^2 2^{1-2\alpha}k_a^{2(1-\alpha)}. \tag{14}$$

In this formulation, J_o is the zeroth-order Bessel function of the first kind, and the normal incidence term is as derived in Ref. 4.

In the sediment volume, we use the unitless acoustic backscatter cross section per unit area per unit solid angle $s_{vl}(\theta_i)$, defined by

$$s_{vl}(\theta_i) = \sigma_2 \mathcal{V}_l(\theta_i)/\mathcal{A}, \quad \text{with } \mathcal{A} \equiv 2/(10\log_{10} e), \tag{15}$$

where $\mathcal{V}_l(\theta_i)$ accounts for the two-way transmission losses at the water–sediment interface with large-scale roughness hav-

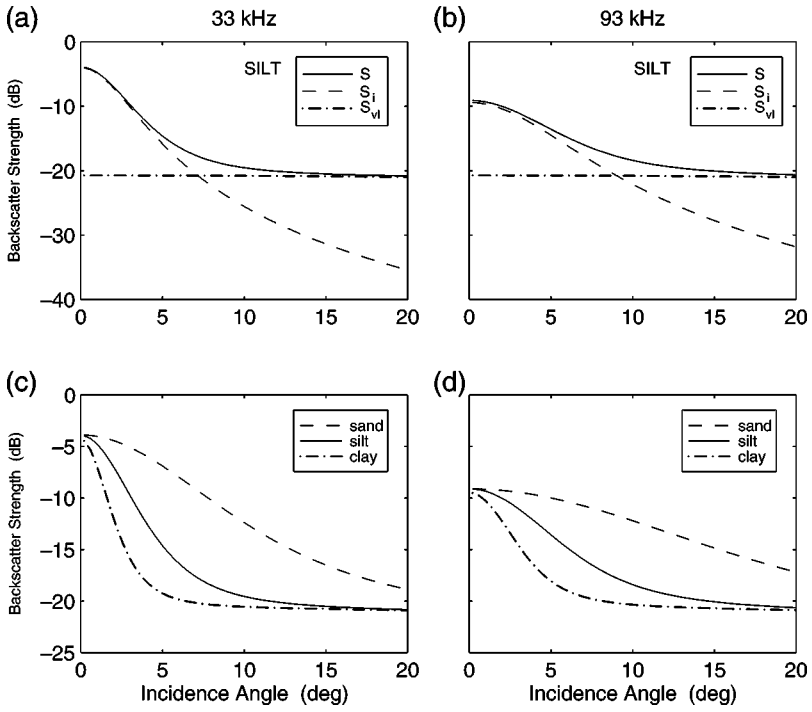


FIG. 2. Angular dependent seafloor acoustic backscatter at 33 and 93 kHz. Top row: total backscatter strength $S=10\log_{10}(s_i+s_{vl})$ over silt, and component backscatter strengths $S_i=10\log_{10}(s_i)$ for the interface term [Eq. (13)], and $S_{vl}=10\log_{10}(s_{vl})$ for the volume term [Eq. (15)]. Bottom row: total backscatter strength S over sand, silt, and clay.

ing small rms slope. This expression is identical to the large-scale volume scattering cross section described in Jackson *et al.*,[3] where $\sigma_2=\sigma_v/\alpha_b$ is termed the volume parameter, and σ_v is the volume scattering coefficient.

Stockhausen[31] gave an angle-dependent expression for two-way transmission through a flat water–sediment interface in terms of the plane-wave coherent reflection coefficient $\mathcal{R}$ and the sound-speed ratio ν at the interface

$$\mathcal{V}_f(\theta_i)=[1-\mathcal{R}^2(\theta_i)]^2\cos^2(\theta_i)[1-(\nu\sin(\theta_i))^2]^{-1/2}. \tag{16}$$

However, at 10–100 kHz very few water–sediment interfaces are likely to appear flat, and the effects of interface roughness must be considered in the two-way transmission through the interface. This is done by considering the distribution of incidence angles expected for large-scale roughness with local slope ϑ, such that the angle of incidence with respect to the local surface is $\theta_i-\vartheta$.

If we assume that the slopes are Gaussian distributed about the horizontal plane, with small rms slopes ($\varsigma<0.1$), then the transmission term for large-scale interface roughness, $\mathcal{V}_l$, is the average of the flat surface coherent reflection coefficients at each planar slope facet of the rough surface

$$\mathcal{V}_l(\theta_i)=\frac{1}{\sqrt{\pi}\varsigma}\int_{-(\pi/2-\theta_i)}^{\infty}\mathcal{V}_f(\theta_i-\vartheta)\exp\left(-\frac{\vartheta^2}{\varsigma^2}\right)d\vartheta. \tag{17}$$

Angle-dependent backscatter curves ($S(\theta_i)$) for sand ($M_\phi=2$), silt ($M_\phi=4$), and clay ($M_\phi=7$) substrates are computed for acoustic frequencies of 33 and 93 kHz, and plotted in Fig. 2, where generic values for the sediment geoacoustic parameters are correlated to mean grain size by the relationships described in Appendix A. These plots illustrate the dominance of interface scattering around normal incidence, giving way to volume scattering for angles exceeding 5–10 degrees. However, sand substrates exhibit a more gentle curve slope, whereas clay substrates exhibit large interface scattering strengths about normal incidence, falling off sharply as θ_i increases. As large particles are more likely than fines to settle in high-energy hydrodynamic environments, the relief energy density spectrum of coarse-grain sediments (small M_ϕ) typically has more energy (higher spectral strength w_2) than that of fine-grain sediments. An increase in w_2 causes a commensurate increase in the cutoff spatial wave number k_c [Eq. (9)]; hence, higher spatial frequencies are included in the theoretical large-scale surface roughness. It follows that the estimated rms slope (ς) of the interface increases, yielding more backscattered energy at higher angles and a backscatter curve with a gentler slope. While increasing roughness, relative to the acoustic wavelength, reduces the normal incidence component, this trend is typically offset by a commensurate rise in the impedance contrast ($\rho\nu$), resulting in $S(0)$ values which are similar for each of the three substrates.

D. Time-dependent sediment interface backscatter

Consider a monostatic transducer at altitude (a) which emits a pulse of duration τ_p seconds, with an intensity time series $I_x(t)$, $0\leqslant t\leqslant\tau_p$. The energy in the pulse propagates as a spherical shell with sound speed ν_w. The intersection of this shell with the bottom initially takes the shape of a disk, changing to that of an annulus. For a level surface, the pulse impacts nadir at time a/ν_w, and the annulus diverges from this point of origin. The bottom projection at some time $t>a/\nu_w+\tau_p$ is illustrated in Fig. 3.

The length of the pulse in water is $\nu_w\tau_p$, and its leading and trailing edges make angles θ_2 and θ_1, respectively, with nadir. The average echo intensity envelope $I_i(t)$ of the backscattered signal begins at the time of *bottom detect*, $t_{bd}=2a/\nu_w$. The backscattered energy in the scenario of Fig. 3

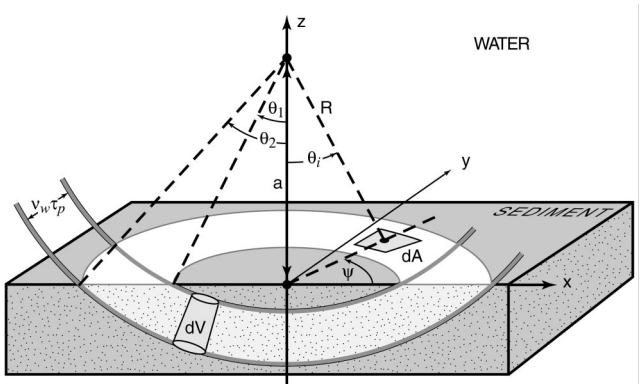


FIG. 3. Geometric representation of the elemental area dA and volume dV used to compute the time-dependent echo intensity envelope.

reaches the receiver at $t=t_{bd}/\cos\theta_2$. The elemental area dA is located at incidence angle θ_i, azimuthal angle ψ, and range $R=a/\cos(\theta_i)$ such that

$$dA=R^2\sin(\theta_i)\,d\theta_i\,d\psi. \tag{18}$$

Similar to sonar equation formulations described in Refs. 4, 11, 13, and 33, the total acoustic intensity field backscattered by the interface and received at the transducer is evaluated over the angular sectors $\theta_1\leqslant\theta_i\leqslant\theta_2$ and $0\leqslant\psi<2\pi$ by convolving the transmitted waveform $I_x(t)$ with the bottom backscattering coefficient $s_i(\theta_i)$ bound by the transducer beam pattern $b(\theta_i,\psi)$

$$I_i(t)=\int_{\psi=0}^{2\pi}\int_{\theta_i=\theta_1(t)}^{\theta_2(t)} I_x\left(t-\frac{2R}{\nu_w}\right)\frac{s_i(\theta_i)b^4(\theta_i,\psi)}{R^4 10^{\alpha_w R/5}}\,dA, \tag{19}$$

where

$$\theta_1(t)=\begin{cases}\cos^{-1}\left(\dfrac{2a}{\nu_w(t-\tau_p)}\right) & t-\tau_p\geqslant t_{bd}\\ 0 & t-\tau_p<t_{bd}\end{cases}$$

and (20)

$$\theta_2(t)=\begin{cases}\cos^{-1}\left(\dfrac{2a}{\nu_w t}\right) & t\geqslant t_{bd}\\ 0 & t<t_{bd},\end{cases}$$

and terms in the denominator of Eq. (19) account for spherical spreading and absorption in the water column during the round-trip travel of the pulse between the transducer and the seafloor. α_w is the frequency-dependent absorption coefficient in water. In the ideal case of a perfectly rectangular transmit pulse, the transmitted intensity $I_x(t-[2R/\nu_w])$ may be replaced by the constant I_x and moved outside the integral.

For sediment interface scattering we use the Kirchhoff approximation with the power-law roughness spectrum to calculate the bottom backscattering coefficient $s_i(\theta_i)$.[3,4,11] This approach has a frequency dependence based on power-law seafloor roughness spectra. This is in contrast to the geometrical optics approach which relies on the high-frequency limit of the Kirchhoff approximation, as described by Beckmann and Spizzichino[34] and Brekohvskikh and Lysanov,[35] and used in Pouliquen and Lurton[13] as a frequency-independent interface scattering component. This high-frequency limit does not exist in the power-law spectrum formulation chosen here. Although each approach has been used successfully in prior work,[7,11,13] we chose the frequency-dependent path because the power-law spectrum is linked to the relief statistics [Eq. (4)] of the types of sediments considered here.

Discrete formulation and implementation of our time-dependent sediment interface backscatter model is described in Appendix B 1.

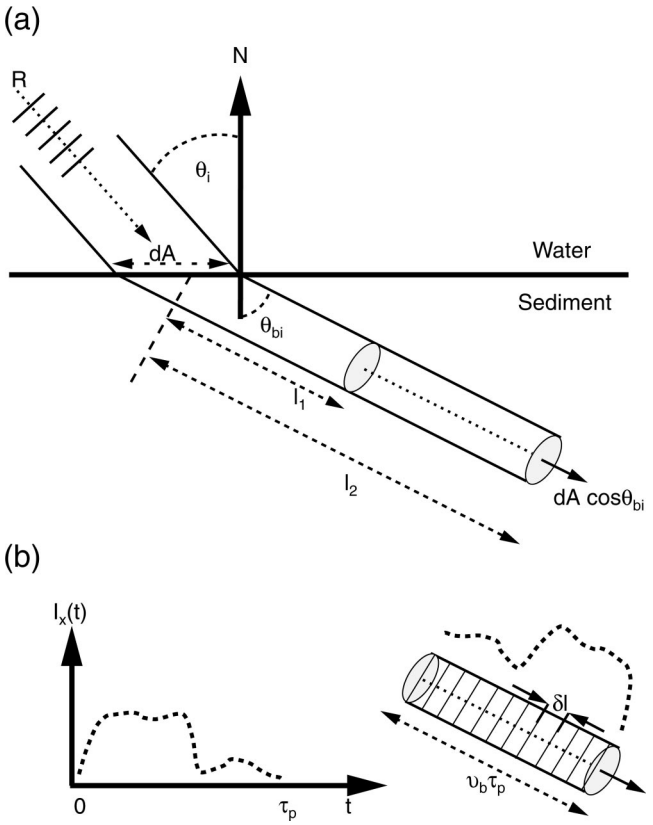


FIG. 4. Integration of backscattered energy from an elemental volume tube: (a) Volume scattering geometry, where scatterers within the tube are insonified from distances l_1 to l_2; (b) illustration of transmitted pulse and juxtaposition with region described in (a).

E. Time-dependent sediment volume backscatter

Single scattering is assumed and the statistics governing subbottom inhomogeneities are assumed to be isotropic and homogeneous, leading to a constant sediment volume scattering coefficient σ_v.

Figure 4(a) shows a closeup of the elemental surface dA at time $t>(R/\nu_w)+\tau_p$, where a portion of the incident energy has refracted into the sediment, and the edges of the pulse have propagation distances (l_1,l_2), referenced from the point of entry into the sediment.

We express the acoustic scattering from the sediment volume as a convolution of the transmit signal $I_x(t)$ with the volume scattering characteristics along the propagation path, taking into account the effects of transducer directivity and of round-trip spherical spreading and absorption in the water column. At time t, the total sediment volume backscattering

field $I_v(t)$ is the sum of all the contributions from volume tubes of cross section dA that converge at the transducer's face. It is expressed in integral form, over the angular intervals $0 \leqslant \theta_i \leqslant \theta_2$ and $0 \leqslant \psi < 2\pi$, as

$$I_v(t) = \int_{\psi=0}^{2\pi} \int_{\theta_i=0}^{\theta_2(t)} \frac{\sigma_v b^4(\theta_i,\psi)}{R^4 10^{\alpha_w R/5}} \mathcal{V}_l(\theta_i) \times \left\{ \int_{l_1(t)}^{l_2(t)} I_x\left(\tau_p - \frac{l - l_1(t)}{\nu_b} \right) e^{-2\beta_e l} dl \right\} dA, \quad (21)$$

with $\theta_2(t)$ defined in Eq. (20), and

$$l_1(t) = \begin{cases} \left[\frac{t}{2} - \tau_p - \frac{R}{\nu_w} \right] \nu_b & t \geqslant 2\left(\frac{R}{\nu_w} + \tau_p \right) \\ 0 & t < 2\left(\frac{R}{\nu_w} + \tau_p \right) \end{cases}$$

and (22)

$$l_2(t) = \begin{cases} \left[\frac{t}{2} - \frac{R}{\nu_w} \right] \nu_b & t \geqslant 2R/\nu_w \\ 0 & t < 2R/\nu_w . \end{cases}$$

Equation (21) is valid from normal incidence to the critical angle.

Spherical spreading within the elemental tubes is considered negligible, and the assumption of statistical volume homogeneity allows placement of σ_v outside the bracketed convolution integral. The integrand includes a spatial representation of $I_x(t)$, tempered by the absorption characteristics of the sediment. The round-trip attenuation between scattering center l and the water–sediment interface is expressed with the exponential intensity attenuation rate β_e, which has units of power nepers per unit distance.

In model computations, the transmitted waveform is digitized and the bracketed integrand of Eq. (21) is evaluated numerically at each time t. If one assumes a perfectly rectangular pulse, I_x may be moved outside of the integral, and the remaining expression evaluates to an attenuation length $L(t)$, which we define as

$$L(t) \equiv \int_{l_1(t)}^{l_2(t)} e^{-2\beta_e l} dl = \frac{1}{2\beta_e} (e^{-2\beta_e l_1(t)} - e^{-2\beta_e l_2(t)}). \quad (23)$$

Note that the time dependency can be removed by assuming an ideal rectangular pulse of length much greater than the energy extinction depth of the substrate, and by evaluating the integral for $l = [0,\infty]$. In this limiting case $L \approx 1/(\mathcal{A}\alpha_b)$, where $\alpha_b = 4.343\beta_e$ is the sediment compressional wave attenuation coefficient and $\mathcal{A} = 2/(10 \log_{10} e)$.

We follow Refs. 3, 4, and 11 and adopt Stockhausen's[31] semiempirical method for calculating the sediment volume backscatter coefficient σ_v (Sec. II C), employing two roughness scales for determining transmission of acoustic energy through the water–sediment interface,[3,4] instead of the physical approach of Chernov[36] and Ivakin and Lysanov,[30] based on estimates of sound speed and density fluctuations in the bottom, and used by Pouliquen and Lurton[13] with a flat water–sediment interface. Our conservative approach sacrifices sound speed and density heterogeniety inversion potential for straightforward estimates of seafloor volume scattering strength. We chose this simpler approach because volume heterogeneity measurements from undisturbed sediments are currently difficult to obtain,[37] and thus existing analytical descriptions of sound speed and density correlation lengths in the sediment are not yet validated. Eventually, the more stringent physical approach to estimating sediment volume characteristics should be a powerful addition to the inversion process.

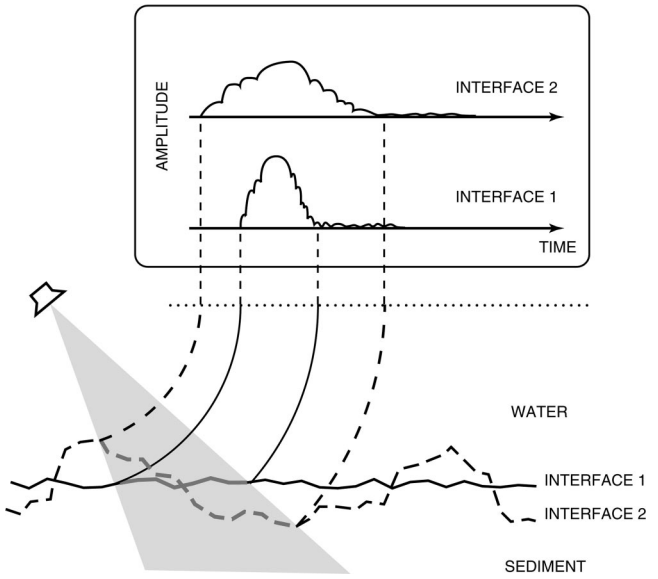


FIG. 5. Lengthening of the echo due to macro-roughness.

The discrete formulation of our time-dependent sediment volume backscatter model is described in Appendix B 2, with its implementation for finite duration, variable amplitude transmit signals.

F. Influence of seafloor macro-roughness

The model described so far predicts the average echo intensity envelope measured with a monostatic transducer aimed at the bottom. The model includes the ratio of scattered to incident energy as the pulse traverses the water–sediment interface and penetrates the substrate. Volume scattering is treated in three dimensions and scattering at the interface is essentially two-dimensional because the ensemble of scattering elements dA are aligned along a mean flat horizontal plane at distance a from the transducer (Fig. 3). Significant deviation of bottom relief from this mean plane at the scale of a beam footprint may result in elongation of the echo, as illustrated in Fig. 5. Temporal stretching of the echo due to large-scale bottom roughness is incorporated by convolving model computations based on the small-scale relief with the height distribution of the interface relief as described by several authors.[9,10,13,38,39] The specifics of our discrete implementation of this macro-roughness effect are described in Appendix B 4.

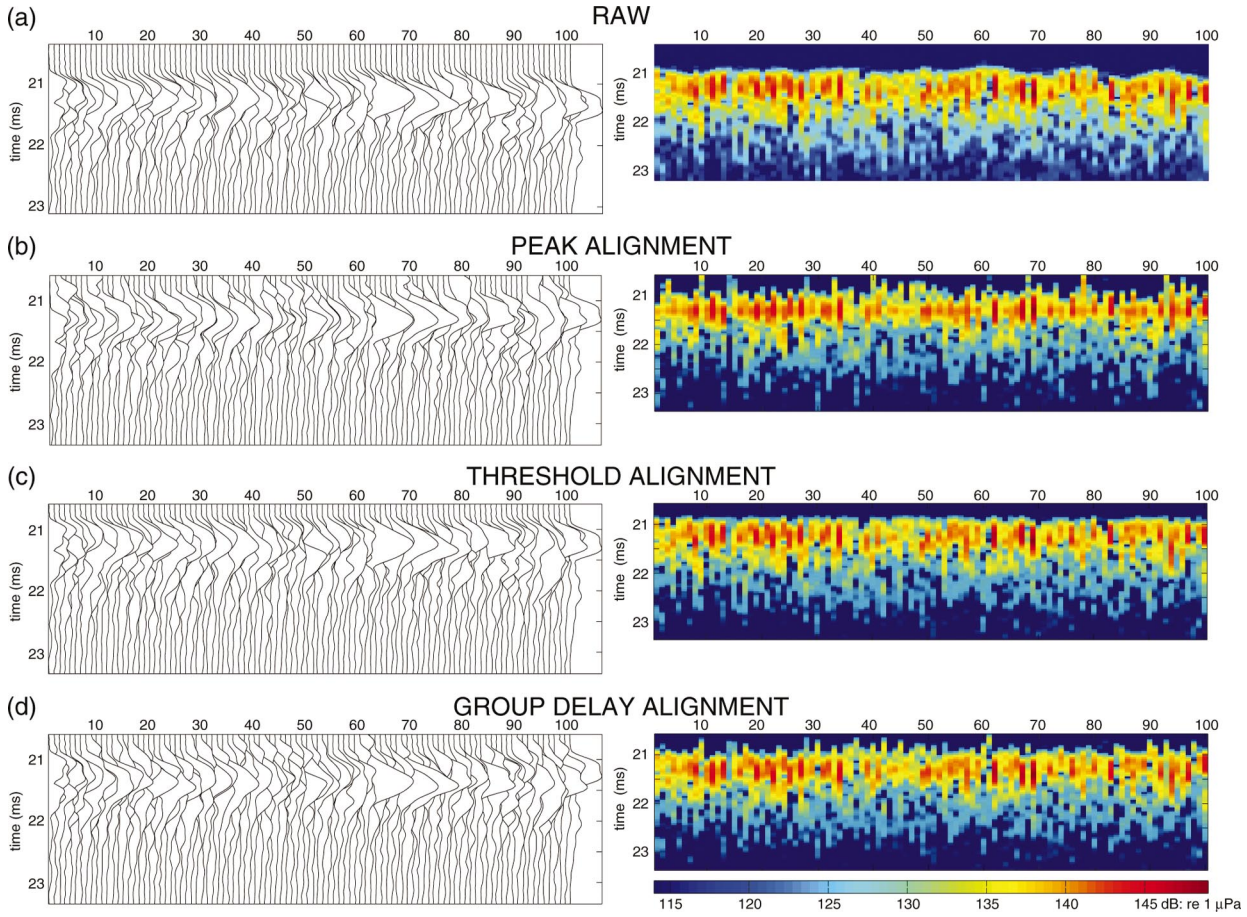


FIG. 6. Waterfall and raster plots illustrating the effect of echo alignment techniques: (a) none; (b) peak with horizontal red line at the signal maxima; (c) leading edge threshold; (d) group delay.

III. DATA

With a calibrated echo-sounder, bottom echoes were collected over substrates in San Diego Bay ranging from clay to sand. The acoustic frequencies (33 and 93 kHz), transducer orientations (maximum response axis at 0°–15° incidence), and −3-dB beamwidths (10°–21°) of the system are consistent with the scattering theory incorporated in the temporal model. The transducer was elevated to a specified angle from nadir in the roll plane. Angles of pitch and roll were digitized for each ping and, combined with knowledge of local bathymetry, used to determine the incidence angle (θ_T) of the transducer's maximum response axis.

The waterfall and raster plots of Fig. 6(a) depict 100 consecutive bottom echoes from the San Diego Bay silt substrate measured at 33 kHz, with $\theta_T = 2°$. With 5 pings per second at a vessel speed of 1 kn, a small bottom patch would typically be sampled 60 times over the −3-dB footprint of the beam; thus, spatial overlap of echoes between consecutive pings is greater than 98%.

Acoustic wavelengths at frequencies greater than 10 kHz are generally small compared to the large-scale relief of the water–sediment interface, resulting in mostly incoherent scattering of the incident acoustic energy [Eq. (1)]. As seen in Fig. 6(a), bottom echoes are incoherent, varying significantly in amplitude and shape as the sonar translates longitudinally above the interface. Because of this variability, echoes must be treated stochastically.

The measured bottom echo consists of a pulsed CW signal modulated by the scattering process. Envelope detection of this signal yields an rms pressure time series, $p(t)$, expressed in units of pascals (Pa), and represented by the discrete sequence $p[n]$ when sampled with a period τ_e. For comparison with the temporal model, an ensemble of M contiguous returns is characterized by the average echo sequence $p_a[n]$ with N samples, and a mean altitude ($\overline{a}$) representing the transducer–bottom distance along a vector normal to the water–sediment interface

$$p_a[n] = \frac{1}{M}\sum_{m=0}^{M-1} p[m,n], \quad n = 0,1,...,N-1. \tag{24}$$

As we only have knowledge of the signal at amplitudes above the noise floor, $p_a[n]$ is truncated at both ends by application of a threshold minimum, P_{th}. The leading edge index corresponding to P_{th} identifies the time of *bottom detect* (t_{bd}), allowing for calculation of the mean altitude

$$\overline{a} = \frac{v_w t_{bd}}{2}, \tag{25}$$

where straight-path association of $\overline{a}$ and t_{bd} is a reasonable assumption for the modest transducer elevation angles char-

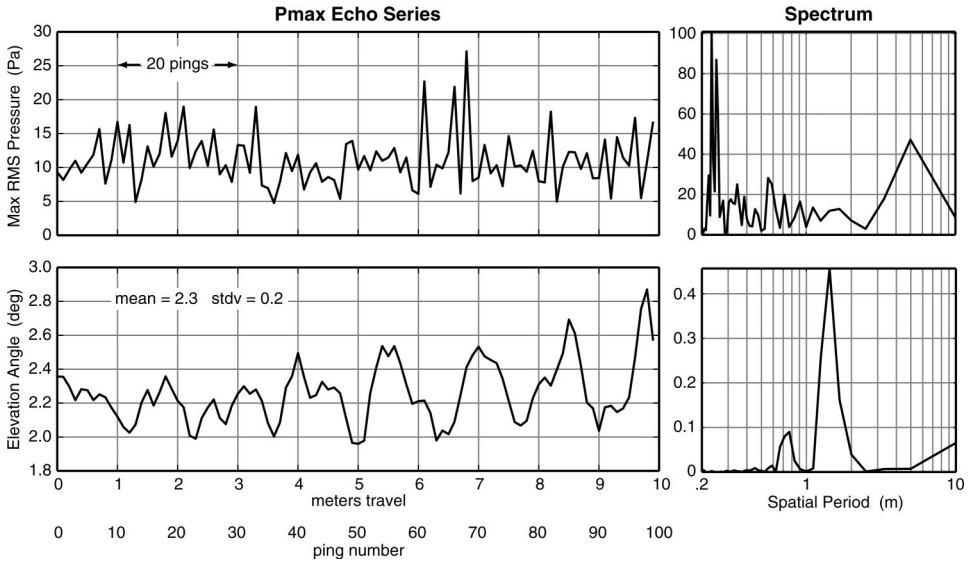


FIG. 7. Amplitude and spectra of peak rms pressure and corresponding transducer elevation angle measured along survey track.

acterizing the measurements. This value of $\overline{a}$ is used for generating the echo envelope model.

In the following sections the ensemble size (M) necessary for computing the average echo envelope is discussed, and alignment techniques compensating for vessel heave and topography variations are described.

A. Ensemble size

The left-hand-side plots of Fig. 7 represent the peak rms pressure (top) and the corresponding elevation angles measured with a clinometer for the 100 sequential echoes of Fig. 6(a). The energy density spectra displayed to the right indicate that most of the energy in the peak rms pressure sequence occurs at short spatial periods, and is minimal around 1.3 m—where the energy in the elevation angle spectrum is maximum. Hence, under stable survey conditions, changes in sensor attitude which are small relative to the beamwidth (standard deviation of 0.2° vs −3-dB beamwidth of 10°) have a negligible effect on the bottom echo statistics.

From analysis of a number of such data sets, we concluded that averaging over an echo ensemble covering the footprint of the −6-dB width of the beam (8.3 m in this scenario) provides adequate statistical representation. For the San Diego Bay measurements, $M=100$ pings meets or exceeds this criterion.

B. Echo envelope alignment

Envelope averaging should be performed on echoes that have been aligned in time, thus removing the effects of transducer heave and of small depth variations over consecutive pings. In this section, we compare the effectiveness of two common alignment techniques based on (1) the threshold minimum (P_{th}) and (2) the echo peak amplitude, and we introduce a third method which exploits phase information in the echo's discrete Fourier transform.

Echo alignment typically relies on tracking and indexing a temporal feature. The echoes within the ensemble are then shifted in time to line up on that feature. For the two-dimensional matrix $p[m,n]$ of Eq. (24), an alignment index (j_m) takes on integer values ($0 \leqslant j_m \leqslant N-1$). A mean alignment index, j, defined by

$$j = \frac{1}{M} \sum_{m=0}^{M-1} j_m , \tag{26}$$

leads to a delay $d_m = j_m - j$ for each of the M echoes. This yields the aligned array $p[m,(n-d_m)]$ that is substituted for $p[m,n]$ in (24) to compute the average echo sequence $p_a[n]$.

The temporal features presented here for comparison are labeled on the canonical echo envelope illustration of Fig. 8. This signal is characterized by a well-defined initial rise and peak amplitude, followed by a slow decay. The threshold minimum and peak amplitude indices associated with the times T_t and T_p are identified by serial search through the sequence, $p[n]$.

For echoes with poorly defined temporal features we employ T_g, which is determined by a process analogous to calculating the signal's group delay. This method provides an alignment index based on energy contributions spanning the entire length of the return, rather than on a single temporal

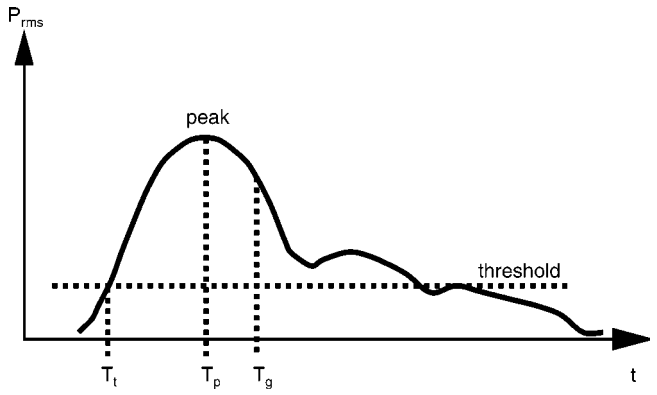


FIG. 8. Signal features used for echo envelope alignment offsets: threshold T_t (rising edge), T_p (peak), T_g (group delay).

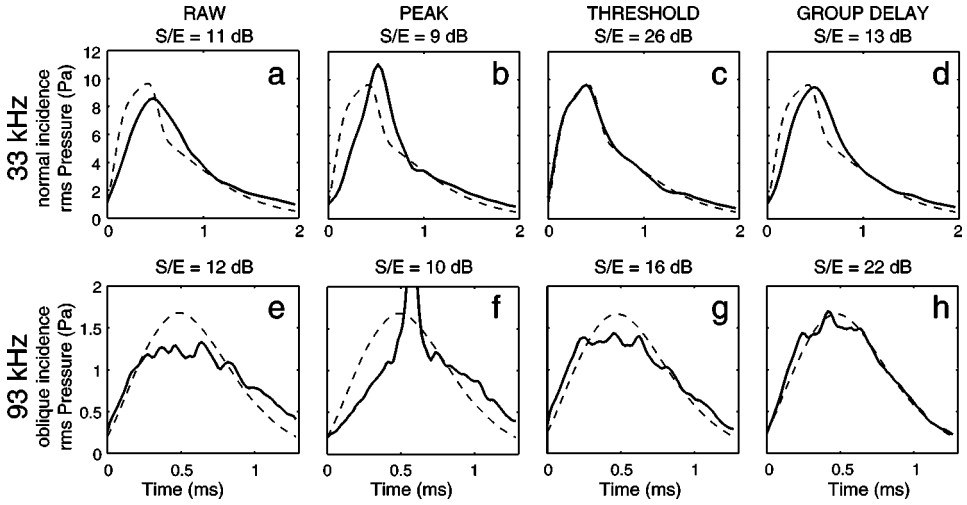


FIG. 9. Comparison of temporal model output (dashed line), given a silt substrate at normal incidence for 33 kHz, and at oblique incidence for 93 kHz, with stacked and averaged echoes (solid lines) for the various alignment techniques. From Sternlicht and de Moustier (Ref. 43).

feature. In this scheme, each sample of $p[n]$ is represented by a phasor $\mathcal{P}[n]$, with amplitude $p[n]$ and phase ϕ_n, such that

$$\mathcal{P}[n]=p[n]e^{i\phi_n}, \tag{27}$$

and

$$\phi_n=\frac{2\pi n}{N}-\pi, \quad -\pi\leqslant\phi_n<\pi. \tag{28}$$

The weighted phase ϕ for the energy in the echo envelope is determined by

$$\phi=\arg\left\{\sum_{n=0}^{N-1}\mathcal{P}[n]\right\}, \tag{29}$$

where ''arg'' refers to the phase of the bracketed quantity. Echoes situated near the beginning of the time window have a negative ϕ, whereas echoes situated toward the end of the window exhibit a positive ϕ. The alignment index (j_m) corresponding to the mth ping in an ensemble is calculated as

$$j_m=\frac{\phi+\pi}{2\pi}(N-1), \tag{30}$$

yielding $T_g=j_m\tau_e$. We refer to this method as group delay alignment, as expanding Eq. (29) reveals a formula similar to the phase slope of the discrete Fourier transform evaluated at the first two Fourier coefficients.

The effects of applying alignment techniques to an ensemble of measured echoes are illustrated in Figs. 6 and 9. The panels of Figs. 6(b)–(d) represent the adjusted envelopes $p[m,(n-d_m)]$ resulting from peak, threshold, and group delay alignments of the data in Fig. 6(a). In the top panels of Fig. 9, average echo envelopes created from these data segments are plotted alongside temporal model computations for which bottom characteristics were determined from grab samples and video images. The signal to error ratio (S/E) quantifies the model-data fits, and is evaluated as follows:

(1) Truncate the leading and trailing edges of the temporal model simulation, $\hat{p}_a[n]$, by applying the threshold minimum P_{th}.
(2) Align $p_a[n]$ and $\hat{p}_a[n]$ along their respective P_{th} leading edge indices.
(3) Truncate the tail end of the longer of the two time series such that $p_a[n]$ and $\hat{p}_a[n]$ are of equal length.
(4) Define n_1 and n_2 as initial and final indices for both waveforms and calculate S/E according to

$$\text{S/E}=\frac{\Sigma_{n=n_1}^{n_2}p_a^2[n]}{\Sigma_{n=n_1}^{n_2}(p_a[n]-\hat{p}_a[n])^2}, \tag{31}$$

for which a high value represents a ''good'' match of model with data.

This figure of merit is independent of signal scale and length, and thus provides a convenient comparison of matching results across data sets.

As seen in Fig. 9(a), averaging ''raw'' data may result in a distorted rising edge—quantified by the low S/E ratio of 11 dB. Alignment by peak tracking [Figs. 6(b), 9(b)] yields a poor representation of the echo. In general, alignments based on signal enhancement techniques (e.g., peak tracking and matched filters) induce vertical disproportions, unsuitable for echo envelope matching.

Bottom echoes from substrates whose relief is small compared to the acoustic wavelength exhibit consistent temporal energy distributions, particularly when measurements are conducted near normal incidence. In these situations, stacking and averaging via minimum threshold [Fig. 6(c)] preserves the integrity of the echo's rising edge, as demonstrated by the 26-dB signal to error match of Fig. 9(c).

In comparison to the other alignment techniques, group delay alignment [Fig. 6(d)] yields a more symmetric distribution of signal energy about the alignment index [Fig. 9(d)], and is less likely to trigger on an early blip or anomalous peak. Threshold alignment may be ineffectual in high-noise environments, or when signal shapes are highly variable—for example, when echoes are measured from substrates which are extremely rough relative to the acoustic wavelength. Under these conditions, group delay alignment may yield average echoes which are more consistent with theoretical predictions. This is illustrated by the 22-dB model–data match of Fig. 9(h), where the data were collected at 93 kHz and oblique incidence.

Based on the above, we have relied on threshold mini-

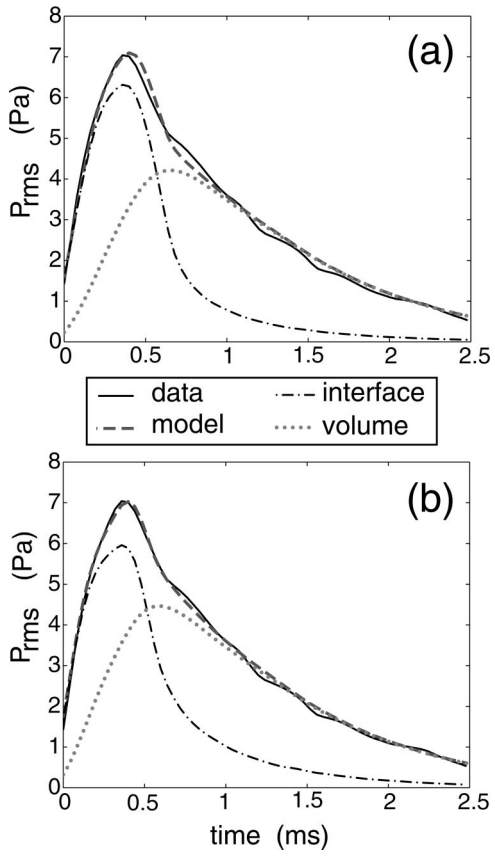


FIG. 10. Example of bottom type aliasing. 33 kHz data (solid line): San Diego Bay silt site. Model (dashed line): $\theta_T=8°$ and "optimum" parameter combinations: (a) "correct" solution: $M_\phi=4.76$, $\gamma=3.26$, $w_2=0.0012$, $\sigma_v=0.091$: S/E=27.6 dB. (b) "aliased" solution: $M_\phi=4.64$, $\gamma=3.00$, $w_2=0.0031$, $\sigma_v=0.111$: S/E=31.4 dB.

mum alignment for data with well-defined energy distributions and on group delay alignment otherwise.

IV. DATA–MODEL COMPARISON

Envelope detection, followed by the alignment and averaging steps described in Sec. III, yields an average rms pressure sequence, $p_a[n]$, whereas model computations yield an echo intensity. We chose to compare model and data as rms pressure sequences because the nonlinear conversion of the measured pressure signal in Pa to a power signal in W/m^2 introduced complications in the echo alignment procedures and in the matching operations. Therefore, a temporal model estimate $\hat{p}_a[n]$ is generated with specified mean altitude and sediment geoacoustic parameters

$$\hat{p}_a[n]=\sqrt{\rho_w \nu_w \hat{I}_a[n]}, \tag{32}$$

where $\hat{I}_a[n]$ represents the intensity of the computed echo envelope [Eq. (B8)] and ρ_w and ν_w correspond to seawater density and sound speed.

Figure 10 shows comparisons of model and averaged data for measurements at 33 kHz over a silt substrate. Henceforth, the sediment–water mass and density ratios (ρ,ν) and the sediment attenuation constant (κ_p) are correlated to the mean grain size parameter (M_ϕ) through the linear regression equations described in Appendix A. The interface contribution dominates the early part of the return, and volume contributions continue as the projection of the pulse exits the transducer beam footprint.

Panels (a) and (b) demonstrate two convincing matches, yet case (b) with the highest S/E represents an aliased solution, whereas the parameters inferred from (a) represent a closer match to those of a silt substrate. Accounting for ambiguous matches is a critical issue in using the model for parameter estimation. In the following sections we consider the influence of each parameter on the temporal model, and propose relief spectrum constraints consistent with ground-truth measurements. These constraints are used in a parameter estimation scheme described in the companion paper.[20]

A. Sensitivity of echo shape to model parameters

The influence of each parameter $(M_\phi,\gamma,w_2,\sigma_v)$ on echo shape is determined by comparing recorded data to iterations of the temporal model. We start in Fig. 11(c) with the closest match between the average echo from another set of 100 sequential echoes measured over a silt bottom, and a model echo with best-fit parameters yielding an S/E of 29.3 dB. Figure 11(d) illustrates the angular dependence curve for a substrate with these parameters: $S=10\log_{10}(s_i+s_{vl})$, where interface and volume scattering coefficients, s_i and s_{vl}, are described by Eqs. (13) and (15), respectively. All other parameters being held constant, each of the four parameters is increased [Figs. 11(a), (b)] or decreased [Figs. 11(e), (f)] relative to its best-fit value to evaluate its effect on the model output.

1. Grain size (Mϕ) influence on echo envelope

An increase in M_ϕ is accompanied by a lower signal amplitude [Fig. 11(a)]. This is due to the smaller impedance contrast $(\rho\nu)$, hence the lower backscatter strength (S), predicted for fine-grain sediments. The higher peak amplitude observed for lower M_ϕ is due primarily to higher impedance contrast [Fig. 11(e)]. The increase of energy in the *tail* of the backscattered signal seen in Fig. 11(a) is explained by the decrease in (κ_p) predicted for fine-grain, water-saturated sediments. These theoretical relationships are supported by the observation that bottom echoes from sand substrates typically exhibit large peak amplitudes (high reflection coefficients), whereas echoes from fine-grain sediments are characterized by long tails indicative of greater subbottom penetration.

2. Relief spectrum (γ,w_2) influence on echo envelope

Changes in either the spectral exponent γ or the spectral strength w_2 produce similar changes in the character of the backscattered echo. Their respective effects on the interface backscatter coefficient (s_i) depend on the range of spectral parameters considered. For the granular sediments studied in this work, the ranges likely to be encountered are $(3.0\leqslant\gamma\leqslant3.6)$ and $(0.0002\leqslant w_2\leqslant0.01)$. Within these boundaries, increasing γ or w_2 has the effect of increasing the cutoff spatial wave number k_c [Eq. (9)]; hence, higher spatial frequencies are included in the theoretical large-scale surface roughness. Likewise, the estimated rms slope (ς) of the bot-

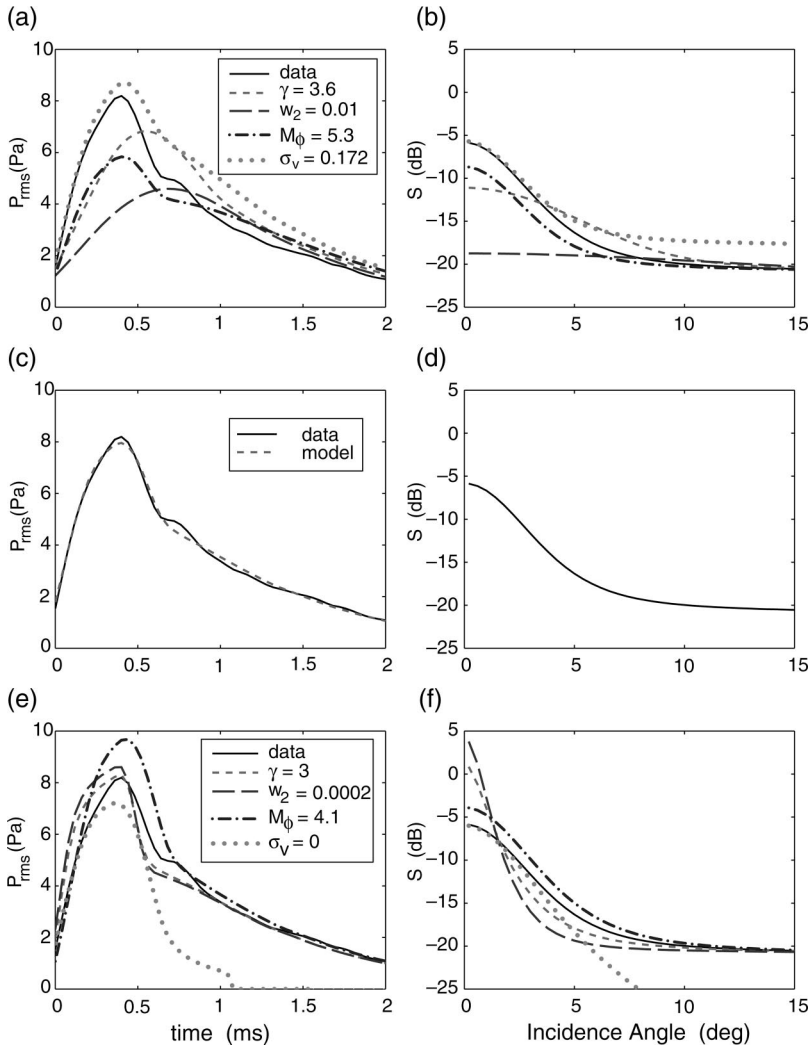


FIG. 11. Effects of geoacoustic parameters on modeled echoes. rms pressure envelopes and corresponding acoustic backscatter strengths (S) (data: solid line, model: dashed line). Best-fit model parameters for (c) and (d): $M_\phi = 4.68$, $\gamma = 3.3$, $w_2 = 0.0009\ \mathrm{cm}^4$, $\sigma_v = 0.086\ \mathrm{m}^{-1}$, $\theta_T = 8°$. (a), (b) and (e), (f) represent, respectively, increases ($M_\phi = 5.3$, $\gamma = 3.6$, $w_2 = 0.01\ \mathrm{cm}^4$, $\sigma_v = 0.172\ \mathrm{m}^{-1}$) and decreases ($M_\phi = 4.1$, $\gamma = 3.0$, $w_2 = 0.0002\ \mathrm{cm}^4$, $\sigma_v = 0\ \mathrm{m}^{-1}$) in the labeled geoacoustic properties, other parameters being held constant at the values in (c) and (d).

tom increases such that more energy is scattered at higher incidence angles, and proportionally less energy is scattered in the vicinity of the normal to the surface.

For higher values of γ or w_2 a gentler slope in the angular response is observed in Fig. 11(b), with a slow initial decay of backscatter strength from normal to higher incidence angles. Here, the backscatter strength at normal incidence is reduced, and there is a smaller dynamic range in levels between 0° and 15°. This is seen in the extreme for $w_2 = 0.01$, where the volume component dominates. Figure 11(a) shows that an increase in γ or w_2 lowers the peak amplitude of the bottom echo, and retards the rise and fall time of the interface scattering component, evidenced by the smearing of the original pulse shape. The shapes of these bottom echoes and angular response curves are characteristic of rough, hard substrates. However, the reduction in backscatter strength and peak echo amplitude observed in the theoretical plots of Figs. 11(a) and (b) are typically compensated by the large reflection coefficients characteristic of these substrates.

Decreased spectral exponent γ and spectral strength w_2 produce a steeper angular response near nadir, characterized by a fast initial decay of backscatter strength from normal to higher incidence angles [Figs. 11(e)–(f)]. The backscatter strength at normal incidence increases, and a large dynamic range is evident between 0° and 15°. The model envelopes of Fig. 11(e) show an increase in amplitude, and sharper rise and decay times.

3. Volume scattering (σ_v) influence on echo envelope

The contribution of subbottom scattering to the received echo is largely determined by the volume scattering coefficient σ_v, which affects energy levels in the tail of the signal. If σ_v is set to zero, the volume component of backscatter is absent. Once the bottom projection of the transmit pulse migrates out of the transducer's main beam, the signal amplitude becomes extremely small, as illustrated in Fig. 11(e). Doubling σ_v from its best-fit value yields a corresponding increase of 3 dB in P_{rms} and S [Figs. 11(a)–(b)].

B. Ambiguities—relief spectrum constraints

The echo envelope model presented here is a function of measurable bottom characteristics, and may be used to predict echo shapes and amplitudes from well-characterized sediments. However, a more useful (and more difficult) application is that of fitting the model to data for sediment classification purposes; that is, extracting parameter combinations which are meaningful, unique, and most importantly,

TABLE I. Published 2D relief spectrum parameters. Comments: ACS=Anisotropic Across-Strike, AAS =Anisotropic Along-Strike. Notation: ~ estimated value, ⇀ graphics offset.

Substrate	Mean grain size (PHI)	Spectral exponent γ	Spectral strength w_2 (cm^4)	Source (Ref. no.)	Comments	Symbol
Coarse sand	0.2~	3.05	0.000 27	26		○
Coarse sand	0.5	2.47	0.006 54	41	Shell hash	
Coarse sand	0.5~	3.46	0.004 73	26	ACS storm gen ripples	○
Coarse sand	0.8	3.12	0.008 49	28	Shell hash	○
Medium sand	1.0	3.0	0.004	25	Shell Fragments	○
Medium sand	1.5~	3.29	0.000 46	26	ACS	○
Medium sand	1.5~	2.33	0.000 38	26	AAS	
Fine sand	2.0	3.0	0.003	25		○
Fine sand	2.5	2.92	0.006 16	42		○
Fine sand	2.5~	3.72	0.000 43	26	Sand dollar smoothing	
Fine sand	2.75~	3.17	0.005 55	26		○
Fine sand	3.0~	3.50	0.000 81	26		○
Very fine sand	3.0	3.3	0.174	25	Dense live shellfish	
Very fine sand	3.0	3.67	0.004 22	26	ACS	
Very fine sand	3.0	3.92	0.005 98	26	AAS	
Silty sand	4.8	3.5	0.004 6	25		*
Mud	5.5	3.18	0.003 18	26	Bimodal clay/sand-gravel	*
Silt	6.4	3.65	0.000 846	27	Prestorm	*
Silt	6.4	3.73	0.000 826	27	AAS poststorm	
Silt	6.4	3.38	0.000 912	27	Prestorm	*⇀6.6
Silt	6.4	3.56	0.001 296	27	ACS poststorm	
Sand–silt–clay	6.5	3.29	0.012 2	28		
Silty clay	9.9	3.42	0.002 31	28	Methane bubbles	*

correct. Unambiguous identification of a best fit is optimized when each of the fitting parameters has a unique effect on the model's output.

To first order, the grain size parameter (M_ϕ) controls the simulated echo's peak amplitude, whereas the volume parameter (σ_v) controls the energy in the signal's tail. The roughness parameters, γ and w_2, control the width and rise time of the signal's peak, but do so in similar fashion as discussed in Sec. IV A 2. The competing effects of these two parameters may lead to several solutions which qualify as ''good'' model–data fits.

Relief spectrum parameters derived from model–data comparisons using four-parameter ($M_\phi, \gamma, w_2, \sigma_v$) unconstrained global model–data optimization techniques (described in Ref. 20) produced substantially different (γ, w_2) combinations. This situation is illustrated by the model–data comparisons of Fig. 10, in which the better fit [Fig. 10(b)] represents a misleading solution according to criteria established below.

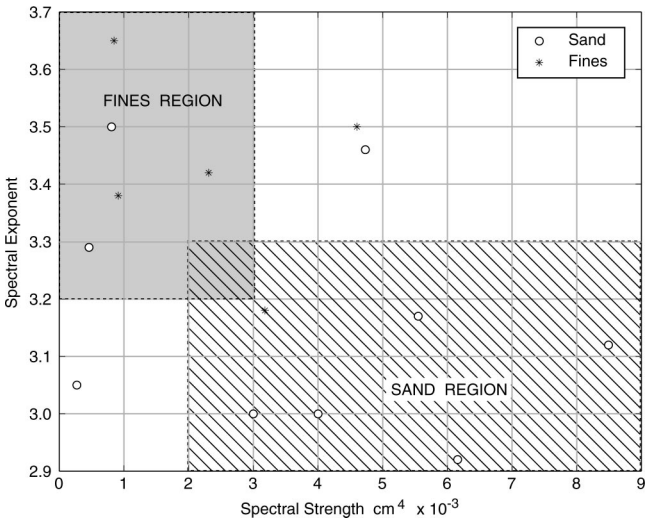


FIG. 12. Published relief spectrum parameters: γ vs w_2 (Table I).

For insight into how relief spectrum parameters relate to bottom type, published relief spectra are summarized in Table I. Measured spectral parameters are listed by bottom type in order of increasing M_ϕ, and a vertical space separates ''sands'' from ''fines.'' The 1D relief spectrum is published in many of the references; the 2D analogs listed are calculated using assumptions of isotropy.[19] Inspection of the table shows the large diversity of spectral parameters measured for sands and fine-grain sediments. To identify trends in the direct roughness measurements, a scatter plot of (γ vs w_2) for the published spectra is presented in Fig. 12. Plots illustrating how γ and w_2 vary with grain size (M_ϕ) are presented in Fig. 13.

Ignoring outliers and redundant measurements, the data in these graphs are a subset of the values listed in Table I—distinguished by the sand symbols (○) and fines symbols (*) listed in the far right column. The shaded γ and w_2 regions (Figs. 12,13) are those which we infer to be characteristic of sand and silts, with boundaries specified in Table II. This grouping of spectral parameters makes physical sense. Large particles are more likely than fines to settle in high-energy environments; hence, sand substrates should exhibit more energy at lower spatial wavelengths (i.e., higher w_2) than their fine-grain counterparts.

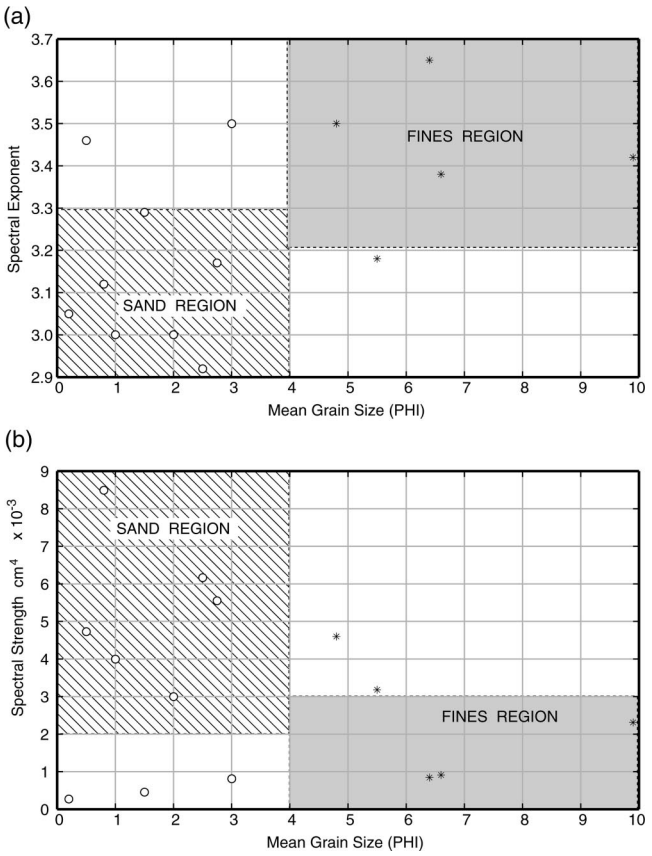


FIG. 13. Published relief spectrum parameters: γ and w_2 vs M_ϕ (Table I).

The exceptions to this rule are disconcerting and warrant closer examination. As evident in Fig. 13, a progressive reversal of the spectral parameters takes place and is most clearly observed for $(M_\phi = 2.5, 2.75, 3.0)$, where large γ values are associated with small w_2 values and, conversely, large w_2 values are associated with small γ values. In Fig. 14, the 2D relief spectra for these three grain sizes are plotted in log–log scale to demonstrate the potential for ambiguities in the linear representation of the power-law relief spectrum, where the spectral ''slope'' is equal to γ, and w_2 represents the energy at the intercept $(k=1)$.

For the relief spectrum parameters (γ, w_2) reported in the literature (Table I): first, the 1D spectra of an ensemble of fine-scale bottom profiles are averaged; the 1D slope and intercept of the average spectrum is determined through linear regression; then, assuming bottom isotropy, the 1D spectral parameters are converted to 2D spectral parameters. The average spectra can be quite noisy, as illustrated by relief spectrum plots presented by Briggs,[26] so a degree of ambiguity is likely in these estimates, which may explain the wide range of spectral parameters reported in the literature. We have observed similar ambiguities in unconstrained optimization of model parameters from our acoustic data, which

TABLE II. Relief spectrum boundaries for granular substrates.

Substrate	M_ϕ	γ	w_2 (cm^4)
Sand	$M_\phi < 4$	$\gamma < 3.3$	$w_2 > 0.002$
Fines	$M_\phi > 4$	$\gamma > 3.2$	$w_2 < 0.003$

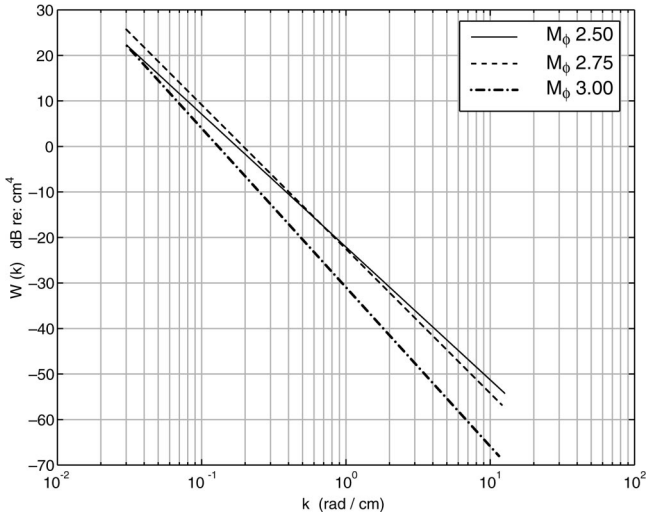


FIG. 14. Relief spectra plots for published parameters derived from topographic data.

suggests that spectral parameter estimates from topographic and acoustic data share similar degrees of uncertainty.[40]

For the San Diego Bay data that we have collected and processed, fewer ambiguous results were obtained when the (γ, w_2) constraints proposed in Table II were enforced during four-parameter model–data matching procedures. This led to the development of a two-stage optimization procedure where for the first stage γ is set to 3.25, the generic value[19] which falls conveniently within the overlap of the regions proposed for sand and fines. For the second stage, γ is set to a value dependent on the initial estimate of mean grain size, $\hat{M}_\phi$. Model–data matching tests using γ constants within the proposed boundaries led us to replace Table II with the simple rule

If $M_\phi < 4.0 \quad \gamma = 3.0$,

If $M_\phi \geqslant 4.0 \quad \gamma = 3.3$.

Subsequently global optimizations were carried out over (M_ϕ, w_2, σ_v), resulting in convergence to unique and sensible solutions.[20]

V. SUMMARY AND CONCLUSIONS

A time-dependent model of the acoustic intensity backscattered by the seafloor was described and compared with data from a calibrated, vertically oriented, echo-sounder operating at 33 and 93 kHz. The model incorporates the characteristics and geometry of the echo-sounder: the transducer's beam pattern, altitude off the bottom, tilt with respect to vertical, and the characteristics of the transmitted sound pulse; and environmental factors: spherical spreading and absorption losses, backscattering of the signal at the water–sediment interface, and by inhomogeneities in the sediment volume.

Scattering from the interface is predicted using Helmholtz–Kirchhoff theory, with physical parameters consisting of the strength (w_2) and exponent (γ) of a power law defining the 2D interface roughness spectrum, the ratio of sediment saturated bulk density over water mass density (ρ),

and the ratio of sediment sound speed over water sound speed (ν). A modified version of the composite roughness model, accounting for the finite duration of the transmitted signal, predicts the energy backscattered from the sediment volume. This contribution to the measured intensity field is determined by the subbottom's volume scattering and attenuation coefficients (σ_v and κ_p, respectively), as well as the water–sediment interface characteristics governing sound transmission into the sediment. Temporal stretching of the echo due to large-scale bottom roughness is incorporated by convolving model computations with the height distribution of the interface relief.

Envelope averaging is performed on echoes aligned in time, where alignments along threshold minima are applied to data with well-defined energy distributions, and offsets calculated from the phase slope of the envelope spectra are applied otherwise. The resulting rms pressure of the measured backscattered signal is compared with estimates of echo intensity converted to pressure [Eq. (31)]. Favorable comparisons of model and data were achieved by correlating the acoustic impedance and attenuation properties of the bottom with measures of its mean grain size (M_ϕ). Estimation of parameters ($M_\phi, \gamma, w_2, \sigma_v$) from fitting the model to data reveal ambiguous ranges for the two spectral parameters. However, analyses of model outputs and of physical measurements reported in the literature yield practical constraints on the roughness spectrum parameters. These constraints are used to estimate geoacoustic parameters through an optimized echo envelope matching techinque described in a companion paper.[20]

A. Limitations

The echo envelope model assumes isotropic and Gaussian-distributed bottom relief, sediment homogeneity in the upper few meters, as well as sediment–water impedance ratios ($\rho\nu$), and sediment acoustic attenuation constants (κ_p) consistent with mean empirical measures. Radical departure from these conditions may invalidate model results. Bottom samples and video images indicated that the sediments surveyed in San Diego Bay generally met these standards. However, model computations may not produce reliable results for complicated biogenic or anisotropic sediments.

Limitations on interface curvature dictated by the Kirchhoff approximation also restrict modeling to sediments having a large rms radius of curvature relative to the acoustic wavelength. This excludes extremely rough (rocky) substrates, or operation at high frequencies (>100 kHz). To evaluate the usefulness of the temporal model for characterizing a broader range of substrates, measurements at a variety of well-characterized sites will be necessary.

ACKNOWLEDGMENTS

We are grateful to the Space and Naval Warfare Systems Center (SPAWAR) for their assistance in collecting acoustic data and bottom samples, Reson Inc. for the generous loan of sonar transducers, and ORINCON Defense for their professional support. We thank Darrell Jackson for his input during the course of this research, and Jo Griffith for her help in preparing the illustrations. This work was funded by the Office of Naval Research under Contract No. N00014-94-1-0121.

APPENDIX A: GENERIC VALUES FOR GEOACOUSTIC MODEL PARAMETERS

The equations used in this work relating geoacoustic model parameters to particle size distribution ($-1 \leqslant M_\phi \leqslant 9$) are adapted from the APL-UW High-Frequency Ocean Environmental Acoustics Models Handbook.[19] The difference between our equations and Ref. 19 is that, in place of the volume parameter (σ_2), we establish a separate particle size mapping to the volume scattering coefficient (σ_v)

$$\begin{aligned}
\rho &\equiv \frac{\rho_b}{\rho_w} \\
&= 0.007\,797\ M_\phi^2 - 0.170\,57\ M_\phi + 2.3139, && -1 \leqslant M_\phi < 1 \\
&= -0.016\,540\,6\ M_\phi^3 + 0.229\,020\,1\ M_\phi^2 - 1.106\,903\,1\ M_\phi \\
&\quad + 3.0455, && 1 \leqslant M_\phi < 5.3 \\
&= -0.001\,297\,3\ M_\phi + 1.1565, && 5.3 \leqslant M_\phi \leqslant 9
\end{aligned} \tag{A1}$$

$$\begin{aligned}
\nu &\equiv \frac{\nu_b}{\nu_w} \\
&= 0.002\,709\ M_\phi^2 - 0.056\,452\ M_\phi + 1.2788, && -1 \leqslant M_\phi < 1 \\
&= -0.001\,488\,1\ M_\phi^3 + 0.021\,393\,7\ M_\phi^2 - 0.138\,279\,8\ M_\phi \\
&\quad + 1.3425, && 1 \leqslant M_\phi < 5.3 \\
&= -0.002\,432\,4\ M_\phi + 1.0019, && 5.3 \leqslant M_\phi \leqslant 9
\end{aligned} \tag{A2}$$

$$\gamma = 3.25, \qquad -1 \leqslant M_\phi \leqslant 9 \tag{A3}$$

$$\begin{aligned}
w_2 &= 0.002\,07\left(\frac{2.038\,46 - 0.269\,23\ M_\phi}{1.0 + 0.076\,923\ M_\phi}\right)^2, && -1 \leqslant M_\phi < 5.0 \\
&= 0.000\,517\,5, && 5.0 \leqslant M_\phi < 9.0
\end{aligned} \tag{A4}$$

$$\begin{aligned}
\kappa_p &= 0.4556, && -1 \leqslant M_\phi < 0 \\
&= 0.0245\ M_\phi + 0.4556, && 0 \leqslant M_\phi < 2.6 \\
&= 0.1245\ M_\phi + 0.1978, && 2.10 \leqslant M_\phi < 4.5 \\
&= 0.200\,98\ M_\phi^2 - 2.5228\ M_\phi + 8.0399, && 4.5 \leqslant M_\phi < 6.0 \\
&= 0.0117\ M_\phi^2 - 0.2041\ M_\phi + 0.9431, && 6.0 \leqslant M_\phi < 9.5 \\
&= 0.0601, && 9.5 \leqslant M_\phi
\end{aligned} \tag{A5}$$

$$\sigma_v = 0.004\alpha_b, \qquad -1 \leqslant M_\phi \leqslant 9. \tag{A6}$$

APPENDIX B: IMPLEMENTATION OF TIME DEPENDENT BACKSCATTER

1. Sediment interface backscatter

Equation (19) is computed numerically using a discrete representation in which the received signal is calculated at intervals of τ_s seconds and indexed by n such that $t=n\tau_s$. Isotropy is assumed in order to reduce the two-dimensional surface integral to a one-dimensional function of the angle of incidence θ_i. Then, the seafloor may be divided into a series of J concentric annuli, with indices j. The area $A[j]$ of each annulus, with range $R[j]=a/\cos(\theta_{ij})$ to its geometric center and radii $r_1[j]$ and $r_2[j]$, is calculated by

$$A[j]=\pi(r_2^2[j]-r_1^2[j]). \tag{B1}$$

Formulation of $r_1[j]$ and $r_2[j]$ depends on the strategy used to partition the insonified area. The indexed angles θ_{ij} may have equal increments in θ_i or they may be calculated via equal increments in the ring radius, r. The former paradigm is used here because it provides finer angular resolution close to normal incidence.

If the change in range from the near to the far edge of $A[j]$ is sufficiently small compared to the length of the pulse, the energy received from $A[j]$ at time index n can be expressed in a form similar to the integrand of Eq. (19) and summed over the elemental annuli, bound by indices $j_1[n] \leqslant j \leqslant j_2[n]$, which overlap the leading and trailing edges of the pulse projection at time index n

$$I_i[n]=\sum_{j=j_1[n]}^{j_2[n]} I_x\left(n\tau_s-\frac{2R[j]}{\nu_w}\right)\frac{s_i[j]A[j]}{att[j]}\mathcal{B}_m[j], \tag{B2}$$

where $s_i[j]$ is the backscatter coefficient, $att[j]$ accounts for the transmission loss in the water column

$$att[j]=R^4[j]10^{\alpha_w(R[j])/5}, \tag{B3}$$

and $\mathcal{B}_m[j]$ is the discrete implementation of the mean directivity function $\mathcal{B}_m(\theta_i)$

$$\mathcal{B}_m(\theta_i)=\frac{1}{2\pi}\int_0^{2\pi} b^4(\theta_i,\psi)d\psi. \tag{B4}$$

When the transmit waveform is represented by a sampled sequence $I_x(n\tau_s)$, the value of $I(n\tau_s-[(2R[j])/\nu_w])$ may be determined by interpolation or rounding.

2. Sediment volume backscatter

The discrete representation of Eq. (21) that we use for numerical computations requires an array for the volume tube [Fig. 4(b)]. An M element array of penetration lengths l_m defines volume cell centers, where l_1 represents the cell at the surface, and l_M the cell at the substrate's *skin depth*, d_s, that we define as the depth at which the signal strength drops by 10 dB

$$d_s=\frac{10+10\log_{10}(1-\mathcal{R}_{\min}^2)}{\alpha_b}, \tag{B5}$$

where for a specific sediment type, $\mathcal{R}_{\min}$ is the minimum angle-dependent reflection coefficient. The width of each cell, δl, is chosen to reflect the resolution of the time-sampled transmit pulse. Here, we chose $\delta l=\nu_b\tau_s/2$, which corresponds to the round-trip path length of the signal, per sample period.

Insonification of the volume continues for some time after the trailing edge of the pulse has passed over the surface patch. Thus, within practical constraints dictated by the skin depth, sediment volume contributions $I_v[n]$ are calculated between nadir ($j=1$) and the annulus overlapping the leading edge ($j=j_2[m]$)

$$I_v[n]=\sum_{j=1}^{j_2[n]}\left(\frac{\sigma_v\mathcal{B}_m[j]}{att[j]}\mathcal{V}_l[j]A[j] \times \sum_{m=m_1[j,n]}^{m_2[j,n]} I_x\left(\tau_p-\frac{l_m-l_{m_1[j,n]}}{\nu_b}\right)L[m]\right), \tag{B6}$$

where $m_2[j,n]$ and $m_1[j,n]$ index the volume cells which overlap the leading and trailing edges of the pulse at $n\tau_s/2$. As before, when the transmit waveform is represented by a sampled time series $I_x(n\tau_s)$, the value of $I_x(\tau_p-[(l_m-l_{m_1[j,n]})/\nu_b])$ may be determined by interpolation or rounding. This function is defined solely within the interval for which the pulse insonifies the volume annulus j bound by the surface and propagation length d_s.

The attenuation length $L[m]$ [the discrete version of Eq. (23)], is an array with elements

$$L[m]=\frac{1}{\mathcal{A}\alpha_b}(e^{-\mathcal{A}\alpha_b(l_m-\delta l/2)}-e^{-\mathcal{A}\alpha_b(l_m+\delta l/2)}), \quad 1\leqslant m\leqslant M. \tag{B7}$$

3. Echo signal intensity

The pressure contributions to the total field that are backscattered by the water–sediment interface and by sub-bottom volume inhomogeneities are assumed to be uncorrelated, and the total intensity received at time $n\tau$ is expressed with a simple addition of these contributions

$$I[n]=I_i[n]+I_v[n], \tag{B8}$$

where for large-scale roughness, $I_i[n]$ is computed with Eq. (B2) and $I_v[n]$ is computed with Eq. (B6).

4. Seafloor macro-roughness

Like Berry and Blackwell,[9] Haines and Langston,[10] Ogilvy,[38,39] and Pouliquen and Lurton,[13] we account for seafloor macro-roughness by convolving the ''smooth surface'' response with a roughness response. Within the length r_f of the footprint of the -3-dB width of the beam pattern along the interface, the macro-roughness corresponds to an altitude change ζ about a plane at mean altitude $\bar{a}$ such that $(\zeta=a-\bar{a})$ is a zero-mean Gaussian random variable with variance σ_a^2. Assuming that a power-law relief spectrum is valid for the macro-roughness at length scales of order r_f, such that mean-square height differences across the footprint increase with increasing footprint size, a measure of σ_a^2 is given by the structure function [Eq. (4)] evaluated at r_f

$$\sigma_a^2=C_\zeta^2 r_f^{2\alpha}. \tag{B9}$$

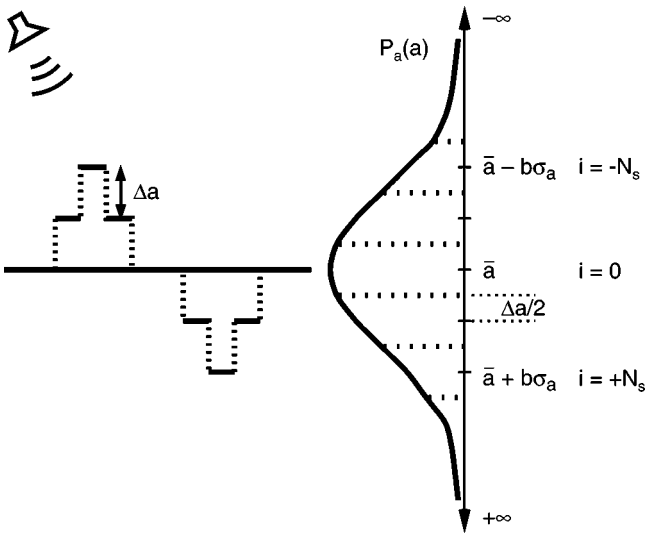


FIG. 15. Probability density function of altitude associated with macro-roughness. The independent axis of the altitude PDF $\mathcal{P}_a(a)$ is oriented along the vertical.

As illustrated in Fig. 15, for numerical computations, we must specify a finite range of altitudes about the mean plane $\overline{a}$ over which the macro-roughness is considered, and then normalize the corresponding area under the altitude probability density function $\mathcal{P}_a(a)$ to 1 in order to retain the total signal energy.

For each seafloor altitude index *i*, illustrated in Fig. 15, a suitably delayed echo is computed and weighted by the probability mass function $\mathcal{P}_a[i]$, which is derived from the Gaussian function illustrated on the right side of the figure. Convolution of the temporal model with this altitude distribution is achieved by summation along the indices of these delayed and weighted model outputs.

For a given angle of incidence θ_o at bottom detect time, discrete altitude increments Δa are computed at the sampling period τ_s

$$\Delta a = \nu_w \tau_s \cos(\theta_o)/2, \tag{B10}$$

and we sum over (N_s) increments above and below the mean plane $\overline{a}$

$$N_s = \mathrm{int}[b\sigma_a/\Delta a], \tag{B11}$$

where b is a fractional constant specifying the percentage of the distribution included in the computation, and ''int'' signifies rounding to the nearest integer value. We use an empirically derived $b=1.28$ (specifying 80% of the distribution) for fine-grain sediments (muds), and $b=0.5$ (40% of the distribution) for coarse-grain sediments (sand). The discrete implementation of the effects of macro-roughness is then expressed as

$$I_m[n] = \begin{cases} \sum_{-N_s}^{N[n]} I[n\tau_s - i\tau_s, a_s[i]]\mathcal{P}_a[i] & n \geqslant (\overline{a} - b\sigma_a)/\Delta a \\ 0 & \text{otherwise,} \end{cases} \tag{B12}$$

where $a_s[i] = \overline{a} + i\Delta a$ represents the height for model index *i*

$$N[n] = \begin{cases} n - \overline{a}/\Delta a & (\overline{a} - b\sigma_a)/\Delta a \leqslant n \leqslant (\overline{a} + b\sigma_a)/\Delta a \\ N_s & n > (\overline{a} + b\sigma_a)/\Delta a, \end{cases} \tag{B13}$$

and the discrete probability mass function $\mathcal{P}_a[i]$ is calculated by integrating the normal distribution between $(\xi_1 = a_s[i] - \Delta a/2)$ and $(\xi_2 = a_s[i] + \Delta a/2)$ such that

$$\mathcal{P}_a[i] = \Phi\left(\frac{\xi_2 - \overline{a}}{\sigma_a}\right) - \Phi\left(\frac{\xi_1 - \overline{a}}{\sigma_a}\right), \tag{B14}$$

where Φ is the distribution function of a normal random variable. For conservation of energy, $\mathcal{P}_a[i]$ must be normalized to 1, such that

$$\mathcal{P}_a[i] = \mathcal{P}_a[i] + \frac{1 - \Sigma_{-N_s}^{N_s} \mathcal{P}_a[i]}{2N_s + 1}. \tag{B15}$$

Macro-roughness lengthens the rise and decay times of the echo envelope and it reduces the maximum signal strength. In the model, these distortive effects are expected to be more pronounced for sand substrates whose macro-roughness standard deviation σ_a is predicted to be about three times as large as that of clay (11.5 cm vs 4.4 cm). A weakness of this approach is that when comparing the model with data, differing values of the macro-roughness parameter (b) are required for sand and fine-grain sediments. For unambiguous estimates of sediment characteristics from bottom echo envelopes, a more sophisticated approach to modeling the effects of macro-roughness will eventually be required.

[1] National Defense Research Council, ''Physics of sound in the sea,'' Technical report, National Research Council, Peninsula Publishing, Los Altos, CA, 1946.

[2] R. J. Urick. *Principles of Underwater Sound* (McGraw-Hill, New York, 1983).

[3] D. R. Jackson, D. P. Winebrenner, and A. Ishimaru, ''Application of the composite roughness model to high-frequency bottom backscattering,'' J. Acoust. Soc. Am. **79**(5), 1410–1422 (1986).

[4] C. de Moustier and D. Alexandrou, ''Angular dependence of 12-kHz seafloor acoustic backscatter,'' J. Acoust. Soc. Am. **90**(1), 522–531 (1991).

[5] S. D. Morgera, ''Signal processing for precise ocean mapping,'' IEEE J. Ocean. Eng. **OE-1**(2), 49–57 (1976).

[6] S. D. Morgera and R. Sankar, ''Digital signal processing for precision wide-swath bathymetry,'' IEEE J. Ocean. Eng. **OE-9**(2), 73–84 (1984).

[7] M. V. Berry, ''On deducing the form of surfaces from their diffracted echoes,'' J. Phys. A **5**, 272–291 (1972).

[8] M. V. Berry, ''The statistical properties of echoes diffracted from rough surfaces,'' Philos. Trans. R. Soc. London, Ser. A **273**, 611–654 (1973).

[9] M. V. Berry and T. M. Blackwell, ''Diffractal echoes,'' J. Phys. A **14**, 3101–3110 (1981).

[10] N. F. Haines and D. B. Langston, ''The reflection of ultrasonic pulses from surfaces,'' J. Acoust. Soc. Am. **67**(5), 1443–1454 (1980).

[11] E. H. Nesbitt, ''Estimation of sea bottom parameters using acoustic backscattering at vertical incidence,'' Master's thesis, University of Washington, 1988.

[12] D. R. Jackson and E. Nesbitt, ''Bottom classification using backscattering at vertical incidence,'' J. Acoust. Soc. Am. Suppl. 1 **83**, S80 (1988).

[13] E. Pouliquen and X. Lurton, ''Identification de la nature du fond de la mer à l'aide de signaux d'écho-sondeurs. I. Modélisation d'échos réverbérés par le fond,'' Acta Acust. European Acoustics Assoc. **2**(2), 113–126 (1994).

[14] X. Lurton and E. Pouliquen, ''Identification de la nature du fond de la mer à l'aide de signaux d'écho-sondeurs. II. Méthode d'identification et résulatats experimentaux,'' Acta Acust. European Acoustics Assoc. **2**(3), 187–194 (1994).

[15] R. Chivers, N. Emerson, and D. R. Burns, ''New acoustic processing for underway surveying,'' Hydrographic J. **56**, 9–17 (1990).

[16] A. S. Tsehmahman, W. T. Collins, and B. T. Prager, ''Acoustic seabed classification and correlation analysis of sediment properties by QTC view,'' in Proceedings of IEEE OCEANS 97, pp. 921–926, 1997.
[17] E. L. Hamilton, ''Geoacoustic modeling of the sea floor,'' J. Acoust. Soc. Am. **68**(5), 1313–1340 (1980).
[18] E. L. Hamilton and R. T. Bachman, ''Sound velocity and related properties of marine sediments,'' J. Acoust. Soc. Am. **72**(6), 1891–1904 (1982).
[19] Applied Physics Laboratory. High-Frequency Ocean Environmental Acoustic Models Handbook. Technical report APL-UW TR9407 AEAS 9501, University of Washington, 1994.
[20] D. D. Sternlicht and C. P. de Moustier, ''Remote sensing of sediment characteristics by optimized echo envelope matching,'' J. Acoust. Soc. Am. **114**, 2727–2743 (2003).
[21] A. Ishimaru, *Wave Propogation and Scattering in Random Media* (Academic, New York, 1978), Vol. 2.
[22] T. H. Bell, Jr., ''Statistical features of sea-floor topography,'' Deep-Sea Res. Oceanogr. Abstr. **22**, 883–892 (1975).
[23] J. M. Berkson and E. Matthews, ''Statistical characterization of seafloor roughness,'' IEEE J. Ocean. Eng. **OE-9**(1), 48–51 (1984).
[24] C. G. Fox and C. E. Hayes, ''Quantitative methods for analyzing the roughness of the seafloor,'' Rev. Geophys. **23**(1), 1–48 (1985).
[25] D. R. Jackson, J. J. Crisp, and P. A. Thompson, ''High-frequency bottom backscatter measurements in shallow water,'' J. Acoust. Soc. Am. **80**(4), 1188–1199 (1986).
[26] K. B. Briggs, ''Microtopgraphical roughness of shallow-water continental shelves,'' IEEE J. Ocean. Eng. **14**(4), 360–367 (1989).
[27] D. R. Jackson and K. B. Briggs, ''High-frequency bottom backscattering: Roughness versus sediment volume scattering,'' J. Acoust. Soc. Am. **92**(2), 962–977 (1992).
[28] D. R. Jackson, K. B. Briggs, K. L. Williams, and M. D. Richardson, ''Test of models for high-frequency seafloor backscatter,'' IEEE J. Ocean. Eng. **21**(4), 458–470 (1996).
[29] E. L. Hamilton, ''Compressional wave attenuation in marine sediments,'' Geophysics **37**, 620–646 (1972).
[30] A. N. Ivakin and Y. P. Lysanov, ''Underwater sound scattering by volume inhomogeneities of a bottom medium bounded by a rough surface,'' Sov. Phys. Acoust. **27**(3), 212–215 (1981).
[31] J. H. Stockhausen, ''Scattering from the volume of an inhomogenous half-space,'' Technical Report 63/9, Naval Research Establishment, Dartmouth, N.S., Canada, August 1963.
[32] S. T. McDaniel and A. D. Gorman, ''An examination of the composite-roughness scattering model,'' J. Acoust. Soc. Am. **73**(5), 1476–1486 (1983).
[33] J. E. Moe and D. R. Jackson, ''Near-field scattering through and from a two-dimensional fluid–fluid rough interface,'' J. Acoust. Soc. Am. **103**(1), 275–287 (1998).
[34] P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces* (Macmillan, New York, 1963).
[35] L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics*, 2nd ed. (Springer, Berlin, 1991).
[36] L. A. Chernov, *Wave Propagation in a Random Medium, Part 2* (McGraw-Hill, New York, 1960).
[37] D. Tang, ''Small scale volumetric inhomogeneities of shallow water sediments: Measurements and discussion,'' in High Frequency Acoustics in Shallow Water, CP-45, pp. 539–546, Lerici, Italy, July 1997. NATO SACLANT Undersea Research Centre.
[38] J. A. Ogilvy, ''Computer simulation of acoustic wave scattering from rough surfaces,'' J. Phys. D **21**, 260–277 (1988).
[39] J. A. Ogilvy, *Theory of Wave Scattering from Random Rough Surfaces* (IOP, London, 1991).
[40] D. D. Sternlicht, ''High Frequency Acoustic Remote Sensing of Seafloor Characteristics,'' Ph.D. thesis, University of California, San Diego, 1999.
[41] S. Stanic, K. B. Briggs, P. Fleischer, W. B. Sawyer, and R. I. Ray, ''High-frequency acoustic backscattering,'' J. Acoust. Soc. Am. **85**(1), 125–136 (1989).
[42] S. Stanic, K. B. Briggs, P. Fleischer, R. I. Ray, and W. B. Sawyer, ''Shallow-water high-frequency bottom scattering off Panama City, Florida,'' J. Acoust. Soc. Am. **83**(6), 2134–2144 (1988).
[43] D. D. Sternlicht and C. P. de Moustier, ''Stacking and averaging techniques for bottom echo characterization,'' in Proceedings: 16th International Congress on Acoustics and 135th Meeting Acoustical Society of America, Vol. IV, pp. 3023–3024, Seattle, Washington, June 1998.

Remote sensing of sediment characteristics by optimized echo-envelope matching[a)]

Daniel D. Sternlicht[b)] and Christian P. de Moustier[c)]
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0205

(Received 1 November 1999; revised 13 June 2003; accepted 26 June 2003)

A sediment geoacoustic parameter estimation technique is described which compares bottom returns, measured by a calibrated monostatic sonar oriented within 15° of vertical and having a 10°–21° beamwidth, with an echo envelope model based on high-frequency (10–100 kHz) incoherent backscatter theory and sediment properties such as: mean grain size, strength, and exponent of the power law characterizing the interface roughness energy density spectrum, and volume scattering coefficient. An average echo envelope matching procedure iterates on the reflection coefficient to match the peak echo amplitude and separate coarse from fine-grain sediments, followed by a global optimization using a combination of simulated annealing and downhill simplex searches over mean grain size, interface roughness spectral strength, and sediment volume scattering coefficient. Error analyses using Monte Carlo simulations validate this optimization procedure. Moderate frequencies (33 kHz) and orientations normal with the interface are best suited for this application. Distinction between sands and fine-grain sediments is demonstrated based on acoustic estimation of mean grain size alone. The creation of feature vectors from estimates of mean grain size and interface roughness spectral strength shows promise for intraclass separation of silt and clay. The correlation between estimated parameters is consistent with what is observed *in situ*. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1608019]

PACS numbers: 43.30.Gv, 43.30.Hw, 43.30.Ft, 43.30.Pc [DLB] Pages: 2727–2743

I. INTRODUCTION

Remote classification of ocean sediments is motivated by mineral resources assessment, cable and pipeline route planning, and mine warfare. In recent years a number of high-frequency (>10 kHz) echo analysis techniques have been developed for characterizing the upper layer of seafloor sediments.

Sediment classification techniques using single-beam sonars are either phenomenological or physical. Phenomenological approaches identify nonparametric measured echo characteristics with core samples or bottom photographs. Such systems typically require calibration of signal characteristics with ground truth at the beginning of each survey, and operation must proceed at a fixed sensor altitude. Pace and Ceen investigated sediment characterization using single-beam echoes,[1] where comparison of the expanded echo (due to temporal spreading) with the transmit pulse was used to infer bottom roughness. Echo durations commensurate with the duration of the transmit pulse were thought to originate from smooth substrates, whereas longer, variably shaped echoes were attributed to coarse materials. Sediment classification techniques that empirically match echo characteristics to ground truth have since been developed. One such system[2] exploits the bottom echo and the first surface multiple (bottom–surface–bottom) by integrating the energy over the tail section of the first return, and integrating over the entire length of the multiple. Representation of these two measures as feature vectors allows segregation of a variety of bottom types. Building on this paradigm, multifeature classification techniques based on higher moment statistics of the recorded waveform are being investigated.[3,4]

Results from Ref. 1 inspired interpretation of the bottom echo's tail as an indicator of bottom roughness, while the energy content of the multiple is considered an indicator of the reflection coefficient, or hardness of the substrate. Theoretical explanations for the success of these systems and modeling of the bistatic geometry are being investigated.[5,6]

In physics-based approaches, sediment characteristics are estimated by comparing measurements to predictions made with physical models—thus minimizing presurvey training requirements and removing limitations on sensor altitude that, typically, are found in phenomenological approaches. One example of physics-based acoustic sediment characterization is described in the works of Schock, LeBlanc, and Mayer,[7,8] wherein broadband (2–10 kHz) echo amplitudes are used to estimate coherent reflection coefficients of sediment layers, and measured distortions of echo spectra yield information on sediment attenuation properties.

The inspiration for our work comes from physics-based echo envelope inversion techniques described by Berry,[9] Nesbitt,[10] Jackson and Nesbitt,[11] and Lurton and Pouliquen.[12] Berry's estimation of irradiated surface characteristics employs half-power lengths of measured and modeled average radar backscatter envelopes. Nesbitt used a least-squares search for matching acoustic backscatter envelopes with models based on reflection loss, sediment absorption coefficient, rms bottom slope, and a sediment volume scattering parameter. His work incorporated up to two sedi-

[a)]Parts of this manuscript were presented in the talks: Sternlicht and de Moustier [J. Acoust. Soc. Am. **105**, 1206 (1999)] and Sternlicht and de Moustier [J. Acoust. Soc. Am. **108**, 2536 (2000)].

[b)]Current address: Dynamics Technology Inc., 21311 Hawthorne Blvd., Suite 300, Torrance, CA 90503. Electronic mail: dsternlicht@dynatec.com
[c)]Current address: Center for Coastal and Ocean Mapping, University of New Hampshire, 24 Colovos Road, Durham, NH 03824. Electronic mail: cpm@ccom.unh.edu

coustic backscatter models for extracting bottom characteristics from single-beam echo-sounder data. Lurton and Pouliquen described a method for sea-bottom identification comparing normalized cumulative functions of the echo envelope derived from measurements and physical backscatter models. Waveform normalization allows for the use of uncalibrated echo-sounders; however, ignoring echo strength limits exploitation of important information such as impedance contrast at the water–sediment interface. Furthermore, integration of echo envelope time series into cumulative form disproportionately represents signal components occurring earlier in time (closer to normal incidence).

In this work, we match a physics-based echo intensity envelope model[13] to seafloor acoustic backscatter measurements, made over substrates ranging from clay to sand, collected with calibrated 33- and 93-kHz echo-sounders whose 3-dB beamwidths (10°–21°) and elevation angles (maximum response axis at 0°–15° incidence) are consistent with the model's underlying Kirchhoff scattering theory. This model incorporates the system's deployment geometry, beam pattern, and signal characteristics, the ocean volume spreading and absorption losses, and solutions of the monochromatic wave equation using boundary conditions described by the sediment geoacoustic characteristics. The time-dependent intensity measured at the transducer face $I(t)$ is modeled as the sum of a sediment interface component $I_i(t)$ and a sediment volume component $I_v(t)$

$$I(t)=I_i(t)+I_v(t), \tag{1}$$

where, following the theoretical work of Jackson *et al.*,[14] the interface backscatter component is obtained from a solution of the Helmholtz diffraction integral using the Kirchhoff approximation, and a composite roughness approach is used to predict scattering from the sediment volume.

Model parameters include the mean grain size (M_ϕ), defined as $M_\phi=-\log_2 D_g$, where D_g is the sediment's mean grain diameter,[15,16] and its correlates, the sediment:water density and sound-speed ratios (ρ,ν), and the sediment's compressional wave attenuation constant (κ_p in dB/m/kHz).[17] Fluctuations of these properties are incorporated into a sediment scattering coefficient, $\sigma_v\,(m^{-1})$, signifying the scattering cross section per unit volume, per unit solid angle. The interface is modeled by a power-law relief energy density spectrum $W(\mathbf{k})=w_2k^{-\gamma}$, where $\mathbf{k}$ is the bottom relief's two-dimensional wave number vector with magnitude k, w_2 is the spectral strength (expressed in units cm^4), and γ is the spectral exponent. The roughness spectrum is bandlimited to wave numbers spanning approximately an order of magnitude above and below the acoustic wave number.

The expected *in situ* ranges of the model components are: $-1\leqslant M_\phi\leqslant 9$, $2.4\leqslant\gamma\leqslant 3.9$, $0.0\leqslant w_2\leqslant 1.0$, $0.8\leqslant\nu\leqslant 3.0$, $1.0\leqslant\rho\leqslant 3.0$, $0.01\leqslant\kappa_p\leqslant 1$, $0.0\leqslant\sigma_v\leqslant 1.0$ α_b. α_b is the sediment compressional wave attenuation coefficient in dB/m, calculated as $\alpha_b=\kappa_p\times f_a$,[18] and f_a is the acoustic frequency in kHz. If the statistics describing the sediment characteristics are consistent over measurement scales commensurate with the geographic range of collected bottom echoes, the geoacoustic parameters described above may be estimated from optimized comparisons of the echo envelope model with aligned and averaged data.

Normalized angular dependence curves of seafloor acoustic backscatter, measured with the 16-beam SeaBeam echo-sounder, fitted with computed curves parametrized by the relief spectrum components (γ,w_2) was presented by Michalopoulou *et al.*[19] Using a least-squares maximum likelihood estimator and chi-square acoustic backscatter intensity statistics, the potential of matching acoustic backscatter models with statistically independent measurements was demonstrated. A drawback of this implementation is its reliance on exhaustive search procedures and its limitation to high-impedance contrast, impenetrable substrates with no demonstrable volume component. Another approach is described in Matsumoto *et al.*,[20] where global optimization by simulated annealing and downhill simplex is used to estimate relief spectrum parameters from the same kind of SeaBeam acoustic data.

The samples of the time series measured with a single echo-sounder are partially correlated, making the statistical approach of Ref. 19 inappropriate. Instead, the model's pressure time series are matched to measured echo envelopes calculated from stacked and averaged data with a two-stage, average echo envelope matching procedure, which builds on the work of Matsumoto *et al.*[20] by expanding the optimization to include relief spectrum parameters and physical quantities related to grain size and sediment volume scattering.

By incorporation of the measurement system's transmit and receive sensitivities, directional characteristics, and a filtering operation for converting voltage waveforms measured at the transducer terminals to pressure waveforms incident at the transducer,[21] the shape and amplitude of the bottom's angular response is exploited in a model–data matching scheme appropriate for simple, inexpensive, single-beam echo sounders. This is distinguished from other physics-based approaches which compare normalized measurements of uncalibrated returns to normalized model realizations,[12] and from phenomenological seafloor characterization techniques[2,4] using correlation analysis of measured echo features (e.g., amplitude and energy in bottom echoes and respective surface multiples) with known ground truth.

Our physics-based model–data optimization procedure generates feature vectors with elements consisting of quantifiable geoacoustic parameters (M_ϕ,w_2,σ_v). This information can be directly associated with bottom type; thus, the procedure is, in theory, independent of specific site characteristics or insonification geometry (such as water depth or transducer orientation). In addition, the sensitivity of this optimization procedure to echo variability can be estimated from the covariance matrix of geoacoustic features, derived from synthetic data sets generated with the data covariance matrix for an ensemble of returns. Furthermore, correlation between the geoacoustic parameters (whether due to natural phenomena or artifacts of the optimization procedure) can be characterized.

Computation of the average echo envelope from data, and of a signal to error ratio in the model–data fit are described in Sec. II, with an example of the data covariance matrix and its implications to the model–data matching pro cedure. The two-stage model–data optimization procedure

for estimating bottom characteristics is presented in Sec. III, with a method of evaluating the error propagation inherent to the matching procedure, using parameter covariance matrices produced from Monte Carlo simulated data sets. The system and data used to validate the echo envelope model and parameter estimation technique are described in Sec. IV, with results presented in Sec. V, and analyses of the effects of echo variability on the optimization procedure given in Sec. VI. Section VII draws conclusions about the usefulness and shortcomings of the approach, and its potential for seafloor classification.

II. AVERAGE ECHO ENVELOPE

The measured bottom echo consists of a pulsed CW signal, modulated by the bottom backscattering process, whose envelope detection and sampling at period τ_e yield an rms pressure sequence, $p[n]$, expressed in units of pascals (Pa). Acoustic wavelengths at frequencies greater than 10 kHz are generally small compared to the relief of the water–sediment interface, and bottom echoes are incoherent, varying significantly in amplitude and shape as the sonar translates longitudinally above the interface. Because of this variability, echoes must be treated stochastically.

For comparison with the temporal model, an ensemble of M contiguous returns is characterized by the average echo sequence $(p_a[n],\ n=0,1,...,N-1)$. To this end, a two-dimensional amplitude array $p[m,n]$ is defined for $(0\leqslant n \leqslant N-1)$ samples per ping and $(0\leqslant m\leqslant M-1)$ pings, incorporating segments of the data presented in Sec. IV

$$p_a[n]=\frac{1}{M}\sum_{m=0}^{M-1} p[m,n], \quad n=0,1,...,N-1. \tag{2}$$

Samples in the echo envelopes from the incoherent returns are Rayleigh distributed, but their ensemble average over many pings is approximately Gaussian. Hence, samples of the average echo envelope are Gaussian distributed. The $N\times N$ covariance matrix C of the average echo is estimated by normalizing the data sample covariance by the number of returns (M). Elements of C are thus

$$C_{ij}=\frac{1}{M}\left\{\frac{1}{M-1}\sum_{m=0}^{M-1}(p[m,i]-p_a[i])(p[m,j]-p_a[j])\right\}, \tag{3}$$

where $(0\leqslant i,j\leqslant N-1)$. Henceforth, C is referred to as the data covariance matrix.

To focus this description, we use the average echo for 100 consecutive returns measured from a vessel underway over a silt substrate in San Diego Bay, plotted in Fig. 1. Prior to averaging, the echoes were aligned along their respective threshold indices as described in Ref. 13. The average echo envelope is bracketed by $p_a[n]\pm\sigma_a[n]$, where variances $\sigma_a^2[n]$ correspond to the diagonal elements of C. Plots of C and its corresponding correlation coefficient matrix Υ, with elements: $\Upsilon_{ij}=C_{ij}/\sigma_a[i]\sigma_a[j]$ (Fig. 2) show that the variance is proportional to signal strength and that neighboring samples are highly correlated. In later sections, synthetic data sets generated with C will help assess the effects of signal variability on the model–data matching procedure.

The average echo is summarily matched by a temporal model estimate ($\hat{p}_a[n]$) generated with specified mean altitude and sediment geoacoustic parameters

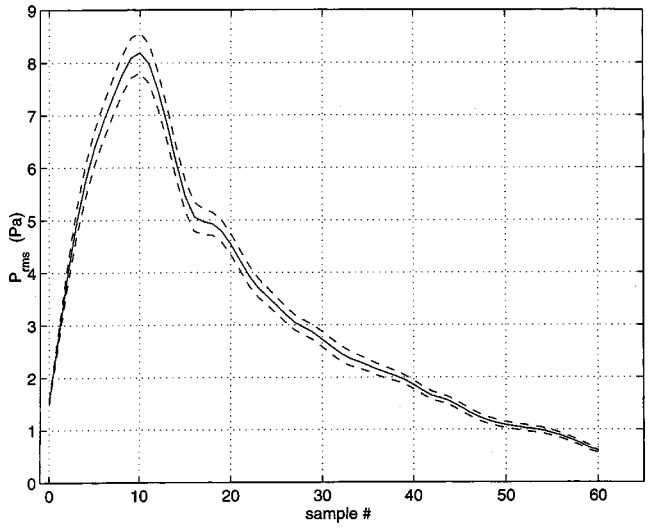


FIG. 1. Average echo envelope for silt substrate: $f_a=33$ kHz, maximum response axis at 8° incidence. Solid line is $p_a[n]$ [Eq. (2)], dashed lines are $p_a[n]\pm\sigma_a[n]$.

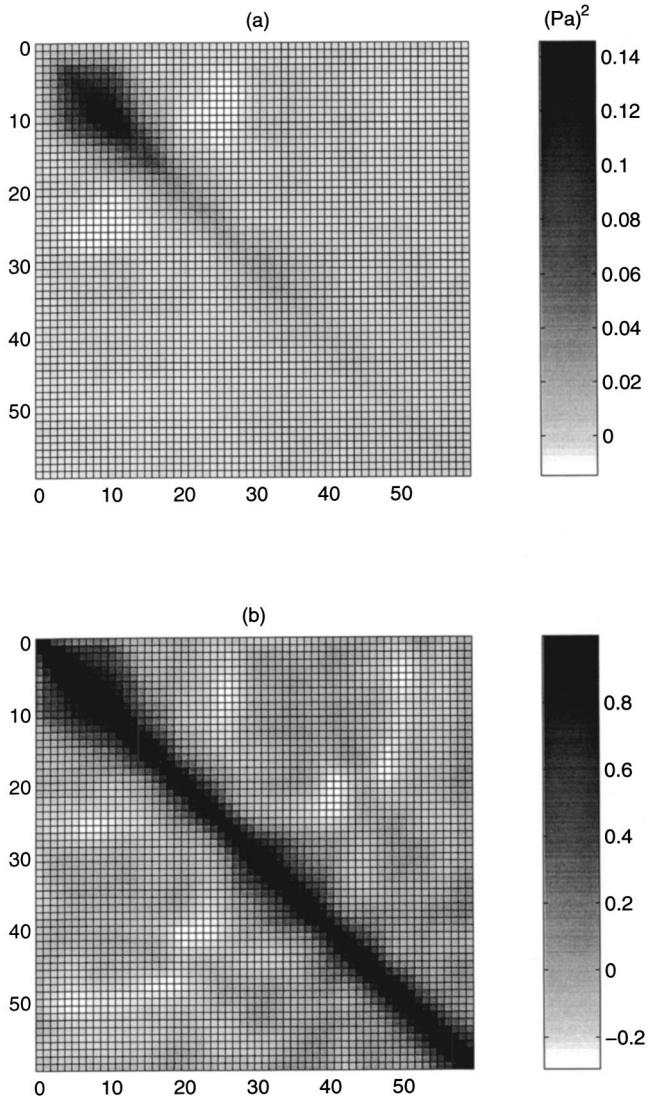


FIG. 2. (a) Data covariance matrix [Eq. (3)], and (b) Correlation coefficient matrix, for average echo envelope of Fig. 1.

$$\hat{p}_a[n] = \sqrt{\rho_w \nu_w \hat{I}_a[n]}, \tag{4}$$

where $\hat{I}_a[n]$ represents the discrete implementation of the echo envelope model, and ρ_w and ν_w correspond to seawater density and sound speed, respectively.

To measure the fit between the model and data, a merit function compares the total energy in the average echo, $p_a[n]$, to a measure of energy representing the discrepancy between model and data. This signal to error ratio (S/E) is expressed as

$$\mathrm{S/E} = \frac{\Sigma_{n=n_1}^{n_2} p_a^2[n]}{\Sigma_{n=n_1}^{n_2} (p_a[n] - \hat{p}_a[n])^2}, \tag{5}$$

where n_1 and n_2 are the initial and final indices for both waveforms. In this scheme a high value of S/E signifies a ''good'' match of model with data. This method provides numerical evaluations which are independent of scale and signal length, and is convenient for comparing results between data sets.

III. GEOACOUSTIC PARAMETER ESTIMATION

The estimation of bottom characteristics from the temporal model depends on a model–data matching paradigm (Fig. 3) that converges to a unique and correct set of bottom parameters. The bottom characteristics which describe the data are determined by comparing the model to the average bottom echo, with the goal of minimizing the error to signal ratio (E/S), i.e., the inverse of Eq. (5). However, estimation of geoacoustic parameters is complicated by the large number of good fits existing in the multidimensional search space, where it is possible to find convincing model–data fits which do not necessarily represent correct solutions.[13] Arriving at sensible solutions requires parsing the problem into manageable parts, establishing the degree of parameter correlations, and constraining the search space.

A. Two-stage parametric optimization

With the goal of deriving unambiguous matches between the temporal model and data, we initially experimented with a one-dimensional (1D) search technique, extracting the best-fit generic characteristics by iterating on the mean grain size parameter (M_ϕ). Here, the six geoacoustic parameters ($w_2, \gamma, \sigma_v, \rho, \nu, \kappa_p$) are related to M_ϕ through linear regression formulas adapted from Refs. 17, 18 and summarized in Appendix A of Ref. 13. As was demonstrated in Ref. 21, the generic parameters produce rough model–data fits for the San Diego Bay substrates investigated. It follows that the solution produced with the 1D search defines a seed vector (M_ϕ, w_2, σ_v) appropriate for a second-stage multiparameter search in which γ is held to a constant. For the second stage, multiparameter local optimization techniques yielded disappointing results marked by convergence to solutions which were unstable and overly sensitive to the choice of seed vector. This led to the development of a model–data matching procedure incorporating the 1D search to establish the general sediment type (sand or fines) and the spectral exponent (γ), followed by a three-dimensional (3D) global optimization using a combination of simulated annealing and downhill simplex searches (SA/DS) over the roughness spectral strength (w_2), the sediment volume scattering coefficient (σ_v), and the mean grain size (M_ϕ) associated with the correlated parameters: ρ, ν, κ_p.

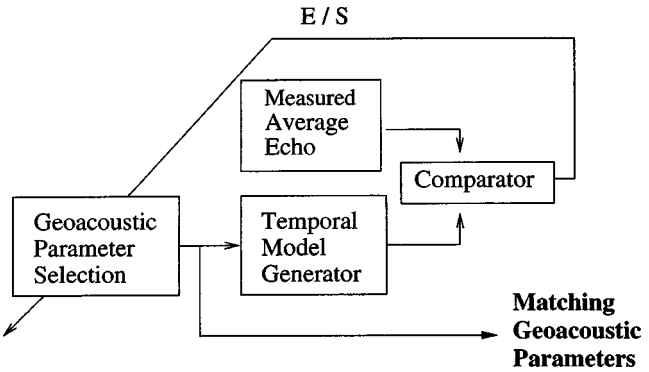


FIG. 3. Geoacoustic parameter optimization procedure: The comparator feeds back (E/S) to the parameter selection module to guide the selection of more promising parameter settings. The system outputs model parameters corresponding to the optimal fit. Careful implementation of the parameter selection module determines the success and tractability of this matching procedure.

1. Stage 1: 1D golden section search and parabolic interpolation

For transducer orientations close to normal incidence, the bottom reflection coefficient is the dominant factor determining the signal amplitude. It follows that the model vs data search space generally has one extremum when described by the single parameter M_ϕ. This situation is illustrated by the E/S vs M_ϕ plot of Fig. 4(a), where the ''best'' solution is found by iteratively bracketing the minimum. For this purpose, we employ a combination of the *golden section* search algorithm coupled with inverse parabolic interpolation, a procedure formulated in Ref. 22. The geoacoustic parameter outputs of stage 1 provide a starting point for the multiparameter global search technique of stage 2.

2. Stage 2: Global simulated annealing—downhill simplex optimization (SA/DS)

After testing a number of local multiparameter search techniques, we found nongreedy, nonexhaustive search procedures to be most appropriate for finding the best-fit geoacoustic parameters. These techniques investigate regions of the parameter space not typically visited by local search techniques, thus increasing the prospects that a true global

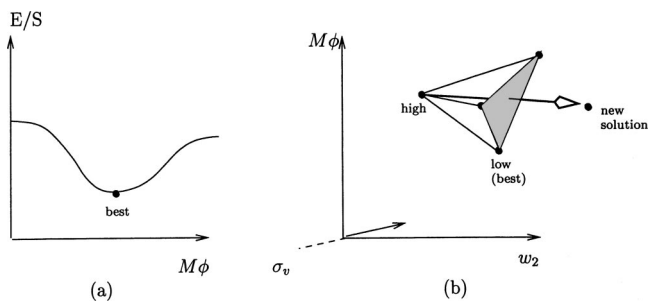


FIG. 4. Parameter space representations for optimization algorithms: (a) 1D search space, E/S vs M_ϕ; (b) *Reflection* across the face of a three parameter simplex.

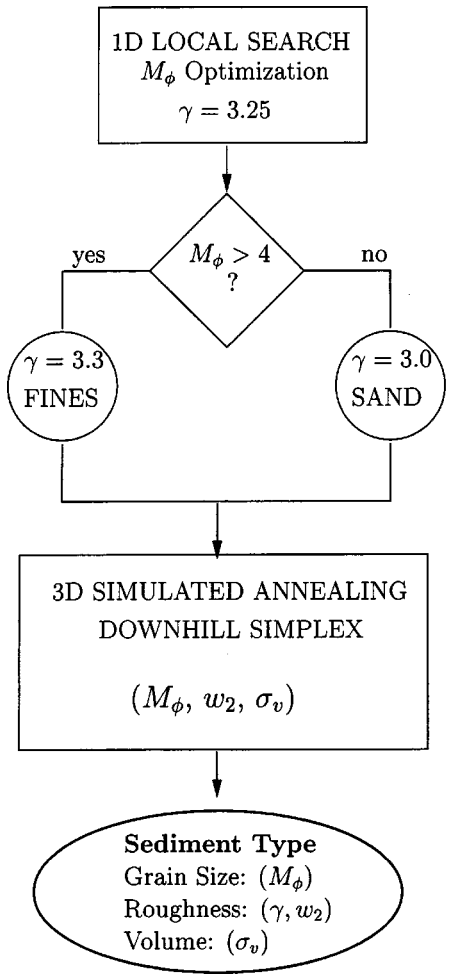


FIG. 5. Flow chart for parameter estimation.

minimum will be found. In *simulated annealing* the system is initialized to some high-energy state and then slowly-brought to the zero state, where a final local search is performed.

A variety of annealing techniques exists, with common reliance on randomly generated numbers for selection of new parameter vectors. We initially tested the best-known version, described in Ref. 23. This method employs the Metropolis algorithm,[24] for which randomly generated parameter vectors, yielding a lower cost than the current vector, are automatically accepted, while those yielding a higher cost are accepted by condition of the Boltzmann probability distribution

$$P(\Delta E) = \exp(-\Delta E/T), \tag{6}$$

where ΔE signifies a positive increase in energy at temperature T. If the search space is vast and/or if calculation of the objective function is computationally intensive, convergence for this method may be unacceptably slow.

Although the temporal model lacks analytic derivatives, it is continuous in the sense that a small change in parameter value is accompanied by a proportional change in the cost function. With this information, faster convergence to a global minimum may be achieved by employing the Nelder–Mead downhill simplex search, modified by random temperature-dependent uphill energy transitions as described in Ref. 25. For a solution space comprised of three parameters (M_ϕ, w_2, σ_v), a simplex of four solution vectors is created as illustrated in Fig. 4(b). The cost function E/S is minimized by reflections, contractions, and expansions of the simplex where, at high temperatures, nonoptimal solutions are occasionally accepted into the simplex at the expense of better solutions. At the final temperature stage $(T=0)$ the simplex is assumed to be in the vicinity of the global minimum, and the Nelder–Mead algorithm is applied in its original form, only accepting better solutions (local search). To maximize the algorithm's effectiveness, the best solution found since initiation of the search is preserved throughout the annealing process.

B. Parameter estimation paradigm

The geoacoustic parameters contained in the temporal model define a complicated search space with numerous local minima. It is thus essential to constrain the solution space using *a priori* knowledge, and to employ practical heuristics in order to reject implausible solutions. For extracting unique and meaningful sediment parameters from the shape and amplitude of measured bottom echoes, we propose the parameter estimation paradigm illustrated by the flow chart of Fig. 5. This technique represents an automated version of the model–data matching guidelines proposed in Ref. 13, where the result of the initial 1D local search (top module of the flow chart) provides the *a priori* information needed to constrain the second stage. The M_ϕ result is fed to a decision junction which determines the general bottom type (sands or fines) and sets the roughness spectral exponent (γ) in preparation for the multiparameter optimization. The 3D global SA/DS procedure iterates over a limited range of M_ϕ, w_2, and σ_v, fine-tuning the impedance contrast, roughness spectral strength, and volume estimates for the substrate. The final result of this procedure provides the general substrate type (sand vs fines), bottom characteristics $(M_\phi, \gamma, w_2, \sigma_v)$, and, indirectly, the sediment geoacoustic parameters correlated to mean grain size.

It should be noted that the search space for the second stage optimization is constrained by restricting the mean grain size to $(\breve{M}_\phi - 1) \leqslant M_\phi \leqslant (\breve{M}_\phi + 1)$, where $\breve{M}_\phi$ represents the seed value from stage one. When contortion of the SA/DS simplex violates these bounds, a suitable penalty is added to the E/S cost function to reject out-of-bound parameter vectors. Broad bounds are similarly applied to the w_2 search space to avoid values unsuitable for the numerical integrations carried out by the temporal model algorithm.

The most important condition imposed on the volume scattering coefficient is $(\sigma_v \geqslant 0)$. However, unreasonably large volume components occasionally produce simulated echoes exhibiting low E/S scores. If the maximum volume component is within 2 dB of the maximum interface component, an empirical penalty, proportional to the severity of this violation, is added to the E/S cost function

$$\text{IF} \quad I_v/I_i > 0.63$$
$$\text{THEN} \quad \text{E/S} = \text{E/S} * 4 * \left\{ 1 + 5 * \left(\frac{I_v}{I_i} - 0.63 \right) \right\}, \tag{7}$$

where I_v and I_i represent the maximum volume and interface intensities, respectively. This is a reasonable restriction except for oblique incidence measurements over fine-grain sub strates, where it is possible for the volume component to

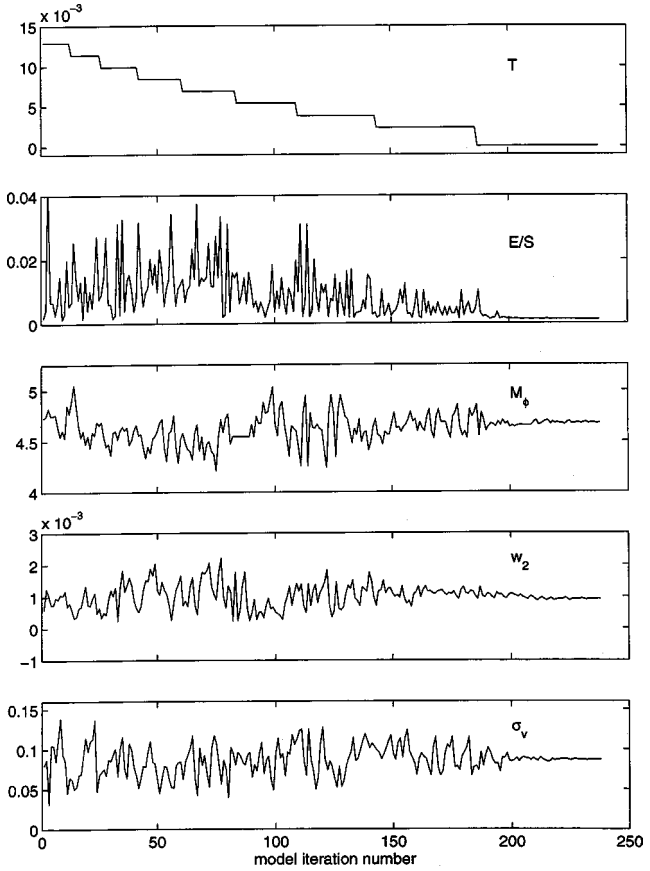


FIG. 6. Annealing process for the data shown in Fig. 1. w_2 in cm^4, σ_v in m^{-1}. The parameter values for the first iteration are the output of the 1D optimization; the final values are the annealing outputs. Note the large variety of nonoptimal solutions investigated before low annealing temperatures constrain the search space. In this particular example, the initial parameters are reasonably close to the final solution.

dominate. In theory, the simulated annealing algorithm converges asymptotically to an optimal solution if the temperature is initially high and allowed to decrease inverse logarithmically with the number of iterations.[26] However, the computational requirements of the cost function in this model–data matching application require a more conservative number of model iterations. After experimenting with the annealing control parameters, adequate solution accuracy and convergence speed were achieved by employing a hybrid linear-exponential cooling schedule with nine discrete temperature levels. In this scheme, the initial temperature T_0 is set to the average E/S for the four initial simplex vertices, where one of these vectors (the seed) is derived from the first local-search stage, and the other three are slightly perturbed replicas. Ten model iterations are initially investigated at T_0 and, for each temperature stage thereafter, the number of iterations increases by 25%, resulting in a total of approximately 230 model iterations (e.g., Fig. 6).

C. Evaluation of error propagation by Monte Carlo simulation

For a given bottom substrate, the average echo can vary from data ensemble to data ensemble. To characterize how this variation affects the results of the model–data matching procedure, K synthetic average echo envelopes are generated with random combinations of signal and noise. Lacking knowledge of the ''true'' signal, the model output resulting from the optimization procedure is distorted by noise characterized in the data's covariance matrix (C). The optimization procedure is applied to each waveform and, using the resulting K solution vectors (M_ϕ, w_2, σ_v), an approximation to the 3×3 parameter covariance matrix is computed and evaluated.

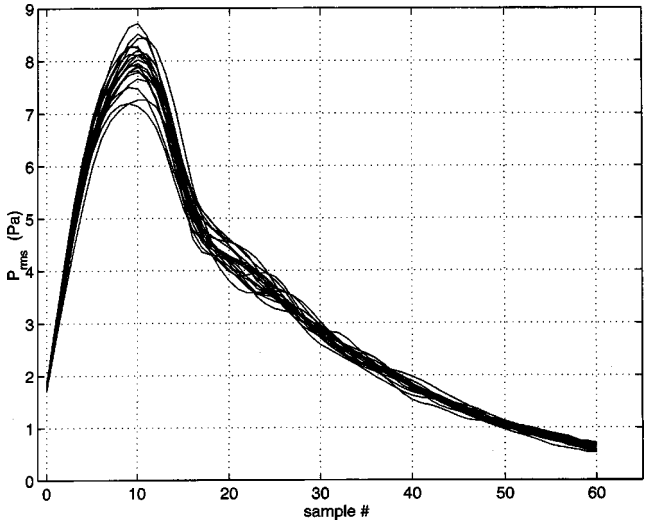


FIG. 7. Monte Carlo simulations for silt substrate in Fig. 1: Model parameters: $f_a=33$ kHz, $M_\phi=4.68$, $\gamma=3.3$, $w_2=0.0009$ cm^4, $\sigma_v=0.086$ m^{-1}, $\theta_T=8°$, $b=1.28$.

With bold lower case letters used to indicate $1\times N$ vectors, a simulated average echo envelope ($\mathbf{p_s}$) is calculated from: the model output $\hat{\mathbf{p}}_\mathbf{a}$, a vector of standard normal random deviates ($\mathbf{x}$), and the upper triangular matrix (A) from Cholesky factorization of the covariance matrix, $C=A^TA$

$$\mathbf{p_s}=\mathbf{x}A+\hat{\mathbf{p}}_\mathbf{a}. \tag{8}$$

Figure 7 shows $K=20$ synthetic ''average'' echo envelopes calculated using the best-fit model for the silt substrate data shown in Fig. 1. These simulations were created using the average echo of Fig. 1 and the covariance matrix displayed in Fig. 2. The amplitude deviations and degrees of correlation between neighboring samples are realistic, as comparison with Fig. 1 confirms. The 20 (M_ϕ, w_2, σ_v) solutions yield the following statistics:

Parameter	Original	Mean	Stdv	*Parameter Pair*	Correl Coeff
M_ϕ	4.68	4.67	0.10	(M_ϕ, w_2)	−0.47
w_2 (cm^4)	0.000 91	0.000 92	0.000 22	(M_ϕ, σ_v)	−0.23
σ_v (m^{-1})	0.086	0.078	0.003	(w_2, σ_v)	−0.14

where ''Original'' refers to the original solution vector. In this example the mean values of the Monte Carlo solutions are similar to the original parameters, the standard deviations are a small percentage of the mean values (with possible exception of w_2), and absolute values of the correlation coefficients are less than 0.5.

In the following sections, plots of M_ϕ and w_2 are used for distinguishing bottom types. Assuming that the solutions are jointly Gaussian distributed, the 90% error ellipse of (w_2 vs M_ϕ) is calculated and plotted in Fig. 8. For this example, the observed echo variability may account for solution intervals: $4.43\leqslant M_\phi \leqslant 4.89$ and $0.0004\leqslant w_2 \leqslant 0.0014$.

IV. SHALLOW-WATER SURVEYS

The sonar system described in Refs. 21, 27 was developed to evaluate the accuracy of the temporal model and its potential for bottom classification over a range of acoustic frequencies and transducer orientations. Circular piston geometries were chosen for their symmetrical directivity patterns, with beamwidths of 21° at 33 kHz and 10° at 93 kHz, so that for each transducer orientation, an adequate range of bottom incident angles could be insonified by a single short pulse of 0.45 msec at 33 kHz and 0.16 msec at 93 kHz.

For meaningful comparison of model and data, the temporal model utilizes a digitized representation of the transmitted signal, and measured voltage waveforms are converted to their respective pressure waveforms using the transducer's mechanical–electrical transfer function.[21,27]

A. Survey site

In January and May of 1997, the dual-frequency echosounder was installed in the instrument well of the 40-ft research vessel ECOS, operated by the Space and Naval Warfare Systems Center (SPAWAR). To validate the temporal model and determine the optimum survey configuration for substrate identification, bottom echoes were recorded from a range of sediment types with the 33- and 93-kHz transducers inclined 0° to 16° from nadir in the roll plane. Data were measured over three sites in San Diego Bay consisting, respectively, of sand, silt, and clay substrates.

Bottom characterization was based on: (1) video coverage recorded during the survey; (2) consulting a sediment data base for the surrounding area; and (3) analysis of particle size distribution for sediment grabs taken during the survey. Sediment samples were separated into size components using sieve separation and pipette settling procedures outlined in Ref. 28. The particle size analyses of these sites are catalogued in the Appendix, Table IV. At these sites, sand particles constituted the largest grain size percentage; however, labels of sand, silt, and clay were determined using the calculated M_ϕ values and observed physical characteristics of the samples.

The sand site consisted of a 50-m N–S trackline running along the jetty at the mouth of San Diego Bay, in water depths of 13–15 m, with mean grain size distributions $M_\phi = 2$, or *medium-fine* sand according to the labeling scheme set forth in Ref. 17. The video images revealed an isotropic bottom characterized by hillocks with crest–trough heights of 40 cm or more over wavelengths of about 8 m, a light sprinkling of shell hash, and an occasional starfish or blade of kelp.

The silt site consisted of a 150-m N–S trackline of the San Diego Bay trough—the deepest part of the bay with water depths of 15–20 m—whose substrate ranged between *clayey sand* and *sandy mud*.[17] The video images revealed long stretches of homogeneous substrate, occasional patches of kelp, and sole blades of sea grass.

The clay site consisted of a 50-m E–W trackline running just north of San Diego Bay's North Island, water depths of 11–13 m, with mean grain size distributions $M_\phi = 7.0$, or *sandy clay*.[17] The grain-size analysis identifies this sediment as borderline silt–clay, but we categorize it as clay because of the relatively smooth seascape observed, and the pasty, waterlogged character of the physical samples. The bottom video revealed a featureless, isotropic bottom, with little flora or fauna except for what appeared to be small burrows less than a centimeter in diameter.

The uncomplicated appearance of these three substrates, the high spatial overlap between consecutive pings, and the generally level bathymetry, were conditions deemed sufficient for testing the accuracy of the temporal model.

B. Data

The acoustic survey for each site was carried out at speeds of 1–2 kn, ping repetition rate of 5 Hz, and horizontal displacements of about 0.1 m per ping. The transducer was elevated to a specified angle from nadir in the roll plane. Angles of pitch and roll were digitized for each ping repetition and used, along with knowledge of local bathymetry, to determine the angle of incidence (θ_T) of the transducer's maximum response axis on the bottom. Sea conditions were generally mild, with pitch and roll standard deviations typically less than 0.5°.

Echoes from the San Diego Bay substrates measured at 33 and 93 kHz are plotted in Figs. 9 and 10. A total of 12 scenarios is analyzed, each characterized by a unique combination of acoustic frequency, sediment type, and transducer orientation (Appendix, Table V). It is tempting to interpret the raster images (Figs. 9, 10) as true geophysical cross sections of the bottom; however, penetration at these high acoustic frequencies is limited and the observed energy is due primarily to scattering from the water–sediment interface.

The raster image of Fig. 9(c) shows a 30-m track segment with a gradual downward slope of the bottom, modulated by the vessel's heave—whose removal is essential for echo alignment and averaging. In contrast, the 40-cmdepth fluctuations apparent in Fig. 10(a) represent actual topography. Therefore, these data sets require a level of scrutiny to identify artifacts that can unfairly bias the shapes and amplitudes of the backscattered echoes. Objects protruding from or suspended over the bottom may cause scattering

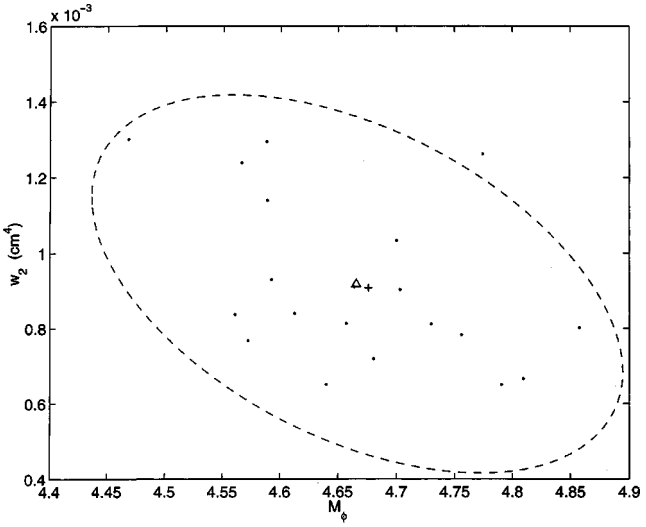


FIG. 8. Scatter plot (w_2 vs M_ϕ) and 90% error ellipse for Monte Carlo simulations. (·) Monte Carlo solutions; (△) mean of Monte Carlo solutions; (+) original solution.

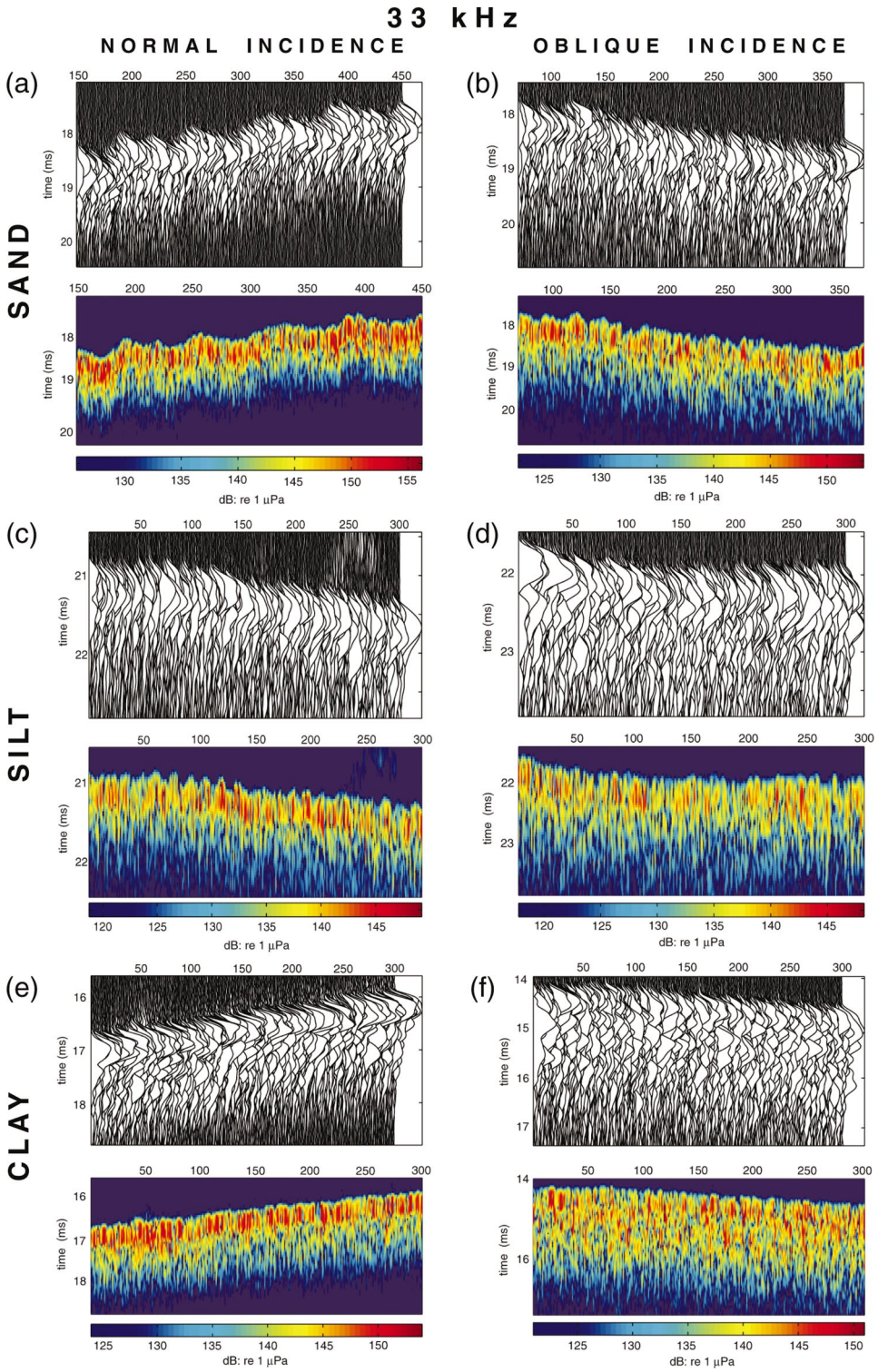


FIG. 9. Waterfall and raster plots for 300 consecutive pings of 33-kHz data. Left: normal incidence, Right: oblique incidence. y=time in ms since transmit, x=ping number.

anomalies and/or reduced signal levels. The early returns evident in pings 230–280 of Fig. 9(c) are most likely caused by a school of fish swimming near the bottom. Similarly, the scattered energy preceding the bottom profile in pings 430–460 of Fig. 10(e) is a strong indication of flora anchored to the sediment. Bubbles on the face of the transducer can cause temporary dropout of signal amplitude, as evident in pings 80–100 of Fig. 10(b). Data segments clearly exhibiting the artifacts described above are rejected.

Segments of these data sets are combined into an average echo envelope (average pressure vs time) for comparison with the temporal model.

V. OPTIMUM FITS OF MODEL WITH DATA

The two-stage parameter estimation technique described in Sec. III was applied to average echo envelopes from the 12 scenarios presented in Sec. IV B. A group delay echo alignment technique was applied to 93-kHz oblique incidence measurements made over sand and silt, and minimum

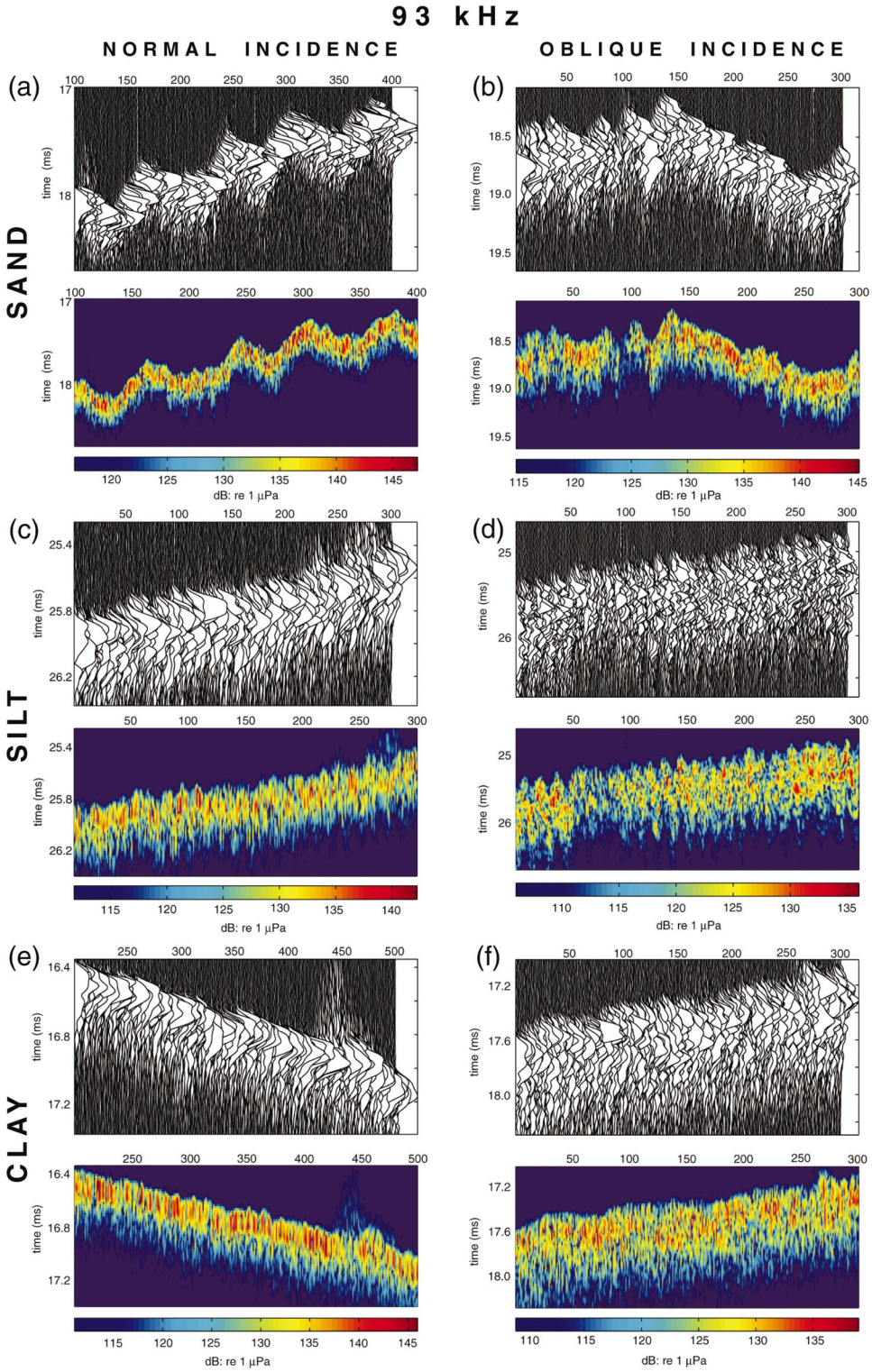


FIG. 10. Waterfall and raster plots for 300 consecutive pings of 93-kHz data. Left: normal incidence, Right: oblique incidence. y=time in ms since transmit, x=ping number.

threshold alignment was used for all other scenarios. Details of the echo alignment techniques are given in Ref. 13.

The volume scatter penalty [Eq. (7)] is applied in all the scenarios, except for 93-kHz oblique incidence measurements on clay and silt. For the latter, large volume contributions are expected to dominate the signal amplitude when transducer elevation angles are large relative to the beamwidth, and in conditions of increased bottom penetration—such as water-saturated sediment and/or low acoustic frequencies.

For each scenario, approximately ten model–data matches were determined with 50% or less overlap between data segments. A summary of model–data matches is presented in Table I. First- and second-order statistics of the results are listed in the Appendix, Tables VI and VII.

To determine the best prospects for sediment classification, we evaluated parameter estimates for the four measurement combinations (two acoustic frequencies, two transducer orientations) and concluded that scatter plots of w_2 vs M_ϕ effectively delineate the bottom substrates (Figs. 11, 12).

TABLE I. Sediment classification summary. Mean values are rounded off to the nearest one-tenth value. Geoacoustic parameters (ρ, ν, κ_p) are calculated from M_ϕ with relationships described in Ref. 13.

Site	Freq (kHz)	Transducer orientation	M_ϕ	γ	w_2 (cm^4)	ν	ρ	κ_p (dB/m/kHz)	σ_v (m^{-1})
Sand	33	Normal	3.9	3.00	0.0050	1.041	1.232	0.680	0.20
		Oblique	3.2	3.00	0.0058	1.072	1.313	0.591	0.04
	93	Normal	2.5	3.00	0.0039	1.109	1.458	0.516	0.95
		Oblique	2.5	3.00	0.0024	1.110	1.464	0.516	0.10
Silt	33	Normal	5.0	3.30	0.0008	1.001	1.170	0.473	0.07
		Oblique	4.7	3.30	0.0012	1.012	1.186	0.653	0.09
	93	Normal	5.2	3.30	0.0007	0.993	1.156	0.363	0.26
		Oblique	5.3	3.30	0.0022	0.989	1.150	0.308	0.31
Clay	33	Normal	5.3	3.30	0.0005	0.990	1.151	0.329	0.05
		Oblique	5.3	3.30	0.0006	0.989	1.149	0.316	0.05
	93	Normal	5.2	3.30	0.0004	0.992	1.154	0.350	0.10
		Oblique	5.2	3.30	0.0005	0.993	1.157	0.365	0.17

A. Mean grain size (M_ϕ) vs relief spectrum strength (w_2)

On the whole, estimated values of mean grain size (M_ϕ) agree with ground-truth measurements presented in Appendix Table IV, and model–data matches for silt exhibit the most consistency across acoustic frequency and transducer orientation. The 33-kHz (M_ϕ) estimates for sand are high, approaching the range characteristic of silts. This may be due to local deviation of impedance contrast from the generic values employed by the model–data matching technique—an inference supported by sediment density measurements described in Refs. 29, 30. As the estimated (M_ϕ) parameter is especially sensitive to changes in measured echo amplitude, errors in field calibration may also contribute to disagreements between model–data matches and ground truth.

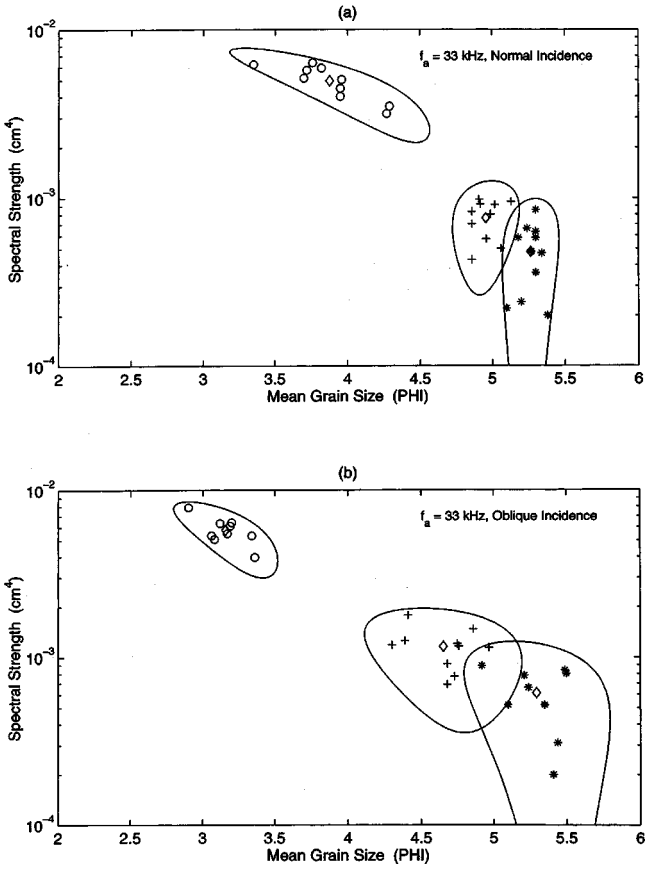


FIG. 11. Scatter plot of model–data matches at 33 kHz: Site locations: (○) Sand; (+) silt; (*) clay; (◇) mean value and center of 90% confidence region (solid line). Transducer orientation: (a) Normal; (b) Oblique.

Mean grain size (M_ϕ) estimates for the clay site (~5.3) are lower than the ground-truth values (>6.5) because the volume signal component is overestimated, due to the sharp decrease of the sediment acoustic attenuation constant (κ_p) correlated with high values of M_ϕ.[17] For these fine-grain sediments, accurate M_ϕ matching is limited, and applying locally determined (ρ, ν, κ_p) trends would probably produce more realistic model–data matches.

Mean grain size (M_ϕ) estimates for sand exhibit greater variability than for fines, with measures of standard deviation ranging from 0.14 to 0.44, and 90%-confidence regions spanning as many as three gradations. In general, M_ϕ estimates for silt and clay exhibit more modest ranges, with standard deviations spanning 0.08 to 0.38.

As seen in Table VI and Figs. 11, 12, estimates of roughness spectral strength (w_2) are greater than 0.001 for the sand site, less than 0.001 for the clay site, and about 0.001 for the silt site. This trend follows the logic that the relief energy density spectra of coarse-grain sediments have more energy than those of fine-grain sediments. Variation in the estimate of w_2 appears greater in fines than in sands. As a percentage of the mean value, w_2 standard deviations for fine-grain sediments (24%–56%) are typically larger than those for sand (19%–38%).

Note from Table VII and Figs. 11, 12 that anticorrelation of M_ϕ and w_2 is also a general bias of the model–data matching procedure. This is especially true of sand measurements, where (M_ϕ, w_2) correlation coefficients range from -0.58 to -0.96, causing the pronounced slope in the sand confidence regions.

In the literature there is agreement that bottom scattering measurements can be matched to general bottom classes (fines, sand, gravel, rock);[31] however, correlation of scattering strength to grain size distribution is thought to be weak within each sediment class. The variability in the individual echo amplitudes that we measured confirms this. If, as indi-

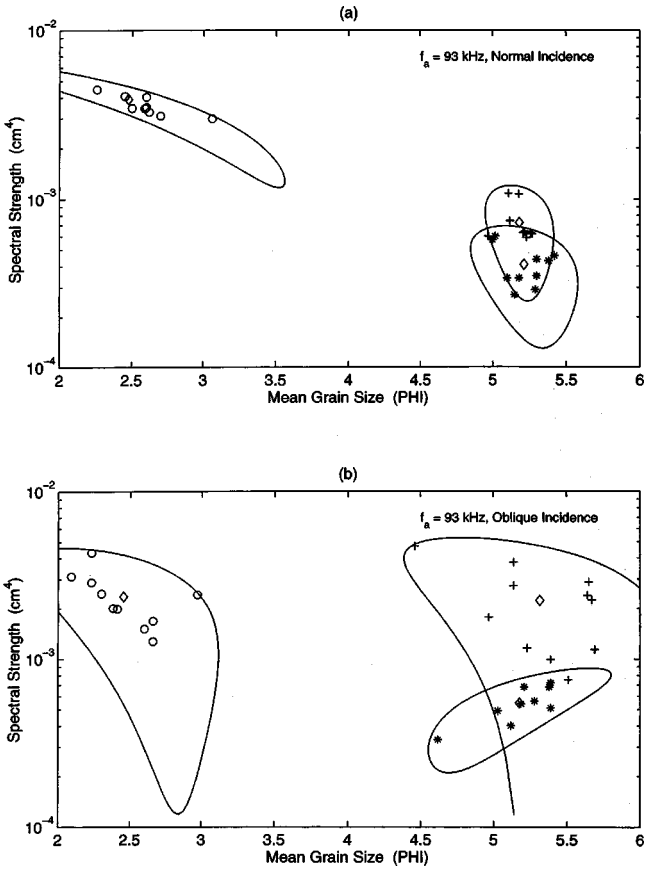


FIG. 12. Scatter plot of model–data matches at 93 kHz: Site locations: (○) Sand; (+) silt; (∗) clay; (◇) mean value and center of 90% confidence region (solid line). Transducer orientation: (a) Normal; (b) Oblique.

cated in this analysis, interface roughness characteristics and grain size distributions were complementary, evaluation of echo shape may allow a degree of *intraclass* separation.

B. Sediment volume scattering coefficient (σ_v)

Estimates of the volume scattering coefficient (σ_v) are perhaps the most difficult to interpret and, as seen in Table VI, standard deviations on the order of 3 dB from the mean value are not uncommon. This variation may be due to real changes in the statistics governing neighboring patches of seafloor. The rough σ_v frequency dependencies inferred from our acoustic backscatter measurements at normal and oblique incidences over sediments in San Diego Bay are $f^{1.2}$ for sand and silt, and $f^{1.0}$ for clay. These values are slightly higher than, but not inconsistent with the $f^{0.7}$ trend inferred from the backscatter measurements analyzed in Refs. 14, 32. Note that an f^4 dependence would indicate Rayleigh scattering from inhomogeneities much smaller than an acoustic wavelength, whereas frequency independence of σ_v would imply geometric scattering from inhomogeneities significantly larger than the acoustic wavelength, as might be the case for our measurements over clay. Our inferred frequency dependencies indicate that the volume scatterers in the San Diego Bay sediments have a range of sizes both smaller and larger than the acoustic wavelength.

At a given frequency and transducer orientation, estimates of σ_v are reasonably consistent for the fine-grain sediments. However, for sand, normal incidence values can exceed oblique incidence values by 10 dB. This may indicate a shortcoming in the model assumption that σ_v is uniform near the sediment–water interface. If σ_v increases with depth, estimated values at normal incidence will appear larger than those for oblique incidence. This is due to acoustic penetration at normal incidence to depths where σ_v is larger—an interpretation consistent with the observations of Refs. 30, 33.

VI. EFFECTS OF ECHO VARIABILITY

Changes in bottom characteristics as well as echo variability due to random constructive/destructive interferences and scattering centers contribute to the observed spread in parameter estimates (Figs. 11, 12). The length scale of the survey and the averaging of 100 pings (corresponding to roughly the along-track extent of the beam's −6-dB footprint) removes some of the ''natural'' variability in the individual ping echoes. To investigate the effects of *residual* echo variability on the outputs of the model–data matching

TABLE II. Monte Carlo statistics.

Site	Freq (kHz)	Transducer orientation	Original M_ϕ	Mean M_ϕ	Stdv M_ϕ	Original w_2 (cm^4)	Mean w_2 (cm^4)	Stdv w_2 (cm^4)	Original σ_v (m^{-1})	Mean σ_v (m^{-1})	Stdv σ_v (m^{-1})	Fig.
Sand	33	Normal	3.72	3.57	0.38	0.005 75	0.006 40	0.001 71	0.201	0.199	0.017	13(a)
		Oblique	3.17	3.21	0.08	0.005 49	0.005 36	0.000 61	0.035	0.037	0.004	13(b)
	93	Normal	2.60	2.51	0.28	0.003 49	0.003 69	0.000 62	0.558	0.572	0.150	14(a)
		Oblique	2.30	2.36	0.12	0.002 46	0.002 36	0.000 19	0.010	0.045	0.061	14(b)
Silt	33	Normal	4.99	4.99	0.05	0.000 80	0.000 80	0.000 22	0.065	0.069	0.006	13(c)
		Oblique	4.68	4.67	0.10	0.000 91	0.000 92	0.000 22	0.086	0.078	0.003	13(d)
	93	Normal	5.12	5.11	0.08	0.000 74	0.000 76	0.000 12	0.288	0.305	0.028	14(c)
		Oblique	5.23	5.16	0.07	0.001 16	0.000 89	0.000 34	0.226	0.273	0.013	14(d)
Clay	33	Normal	5.27	5.27	0.04	0.000 48	0.000 55	0.000 18	0.059	0.067	0.005	13(e)
		Oblique	5.24	5.24	0.05	0.000 66	0.000 78	0.000 36	0.047	0.045	0.003	13(f)
	93	Normal	5.18	5.20	0.04	0.000 34	0.000 33	0.000 05	0.122	0.122	0.014	14(e)
		Oblique	5.19	5.20	0.04	0.000 54	0.000 53	0.000 12	0.181	0.190	0.012	14(f)

TABLE III. Monte Carlo: Parameter correlation.

Site	Freq. (kHz)	Transducer orientation	(M_ϕ, w_2)	(M_ϕ, σ_v)	(w_2, σ_v)
Sand	33	Normal	−0.94	+0.10	−0.19
		Oblique	−0.64	+0.55	−0.45
	93	Normal	−0.87	−0.32	+0.64
		Oblique	−0.85	+0.90	−0.88
Silt	33	Normal	−0.30	−0.31	+0.02
		Oblique	−0.47	−0.44	+0.32
	93	Normal	−0.86	−0.28	+0.12
		Oblique	+0.42	−0.23	−0.14
Clay	33	Normal	+0.05	−0.75	+0.04
		Oblique	+0.44	+0.07	−0.05
	93	Normal	−0.62	−0.66	+0.57
		Oblique	+0.26	−0.24	−0.29

procedure, (M_ϕ, w_2, σ_v) solutions close to the mean value for each (substrate, frequency, orientation) combination were chosen. Then, for each original solution, 20 synthetic average echo envelopes and matched parameter solutions were generated as described in Sec. III C.

For each measurement scenario, an original solution and the mean and standard deviation of the corresponding Monte Carlo solution are summarized in Table II. Correlation between parameter pairs are summarized in Table III. The Monte Carlo solutions for w_2 vs M_ϕ are shown in Figs. 13 and 14 for 33 and 93 kHz, respectively. The distributions of Monte Carlo solutions are adequately represented by the confidence regions—with the exception of Fig. 13(b) (sand, 33 kHz, oblique) which demonstrates one-sided w_2 clustering about 0.0055. The mean values of the Monte Carlo solutions are in general agreement with the original solutions, with the exception of Fig. 14(d) (silt, 93 kHz, oblique).

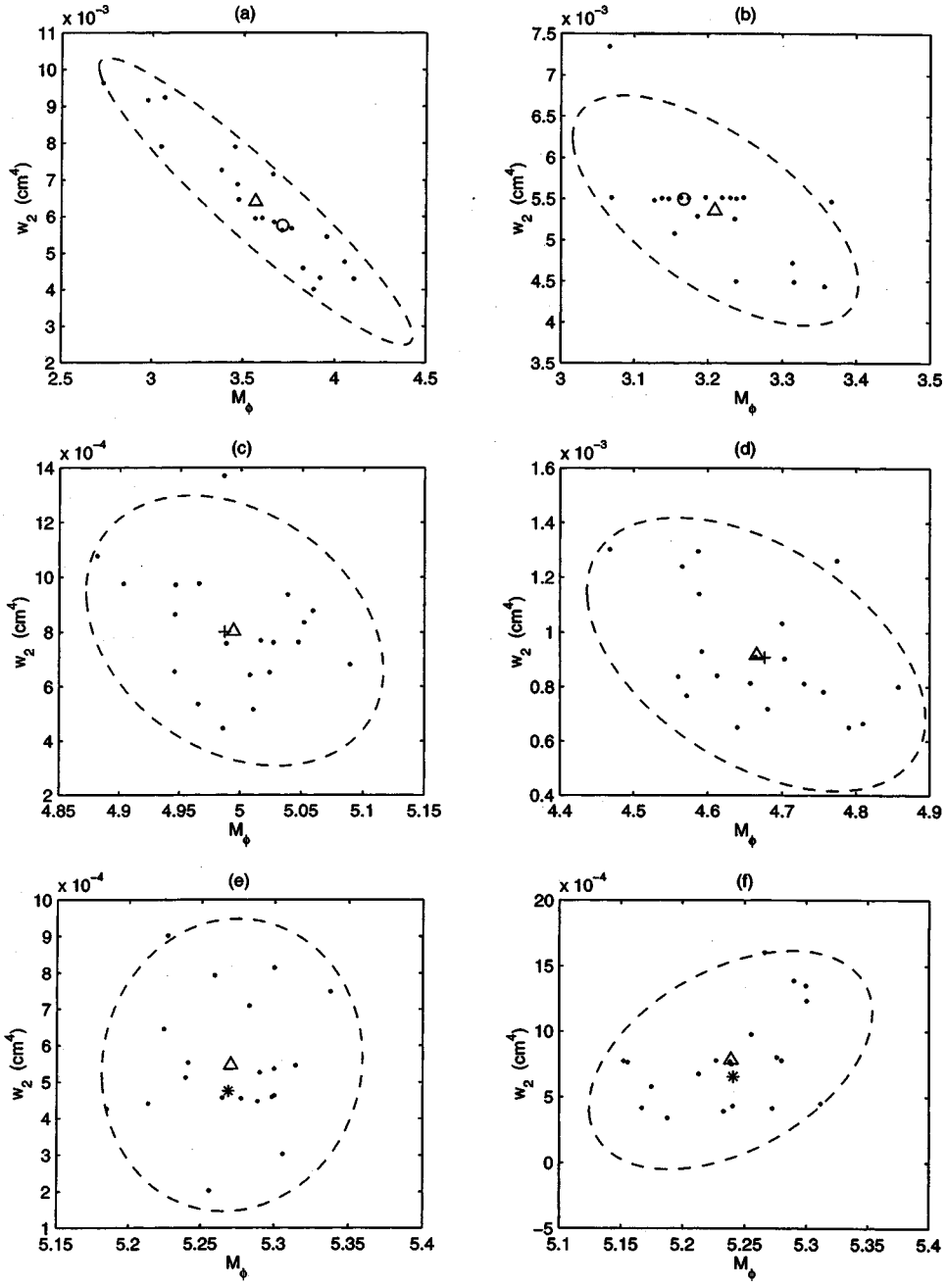


FIG. 13. Scatter plot of Monte Carlo solutions at 33 kHz: Panel descriptions in Table II. (·) Monte Carlo solutions; (△) mean value and center of 90% confidence region; (○) original sand solution; (+) original silt solution; (∗) original clay solution.

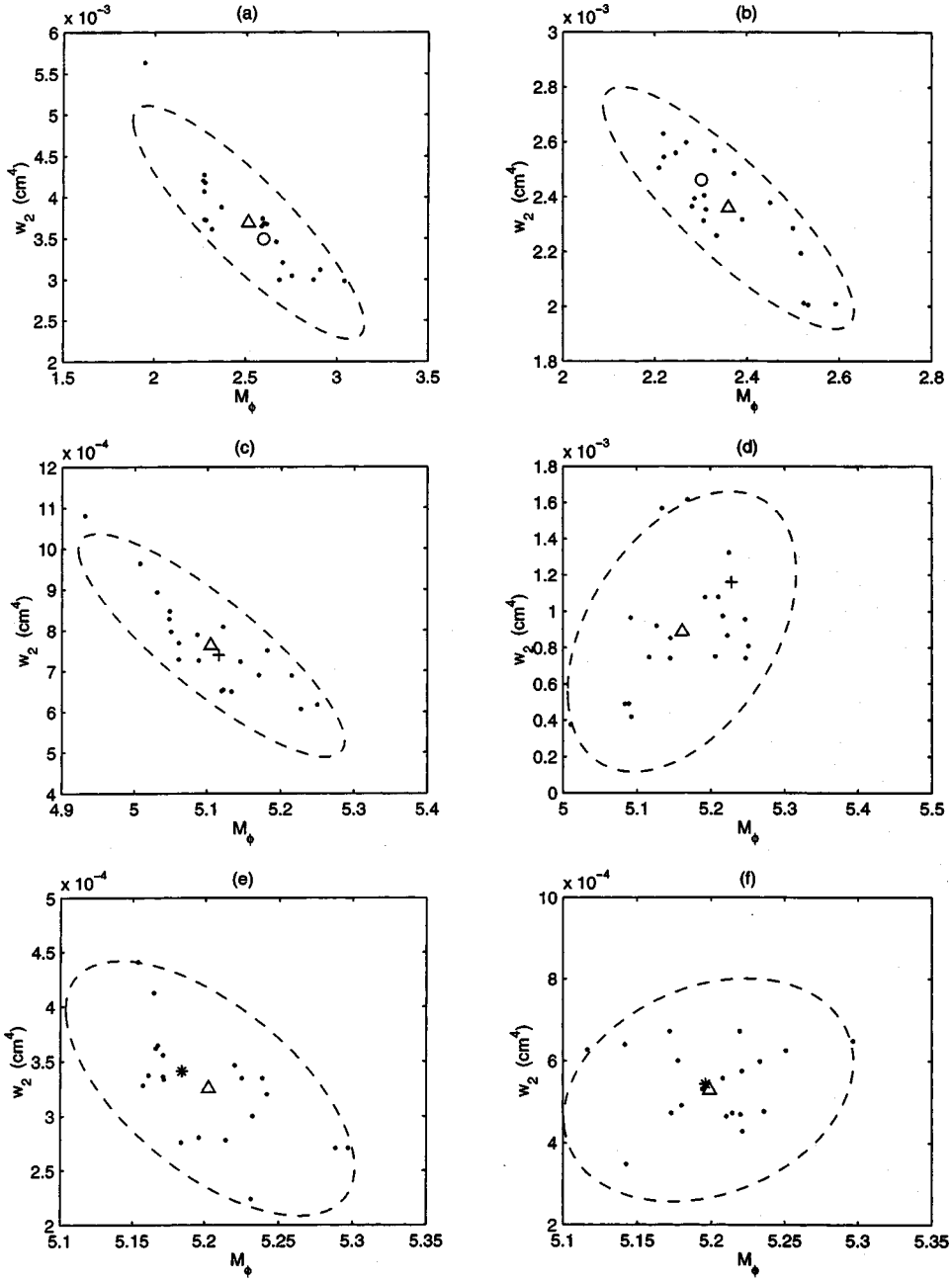


FIG. 14. Scatter plot of Monte Carlo solutions at 93 kHz: Panel descriptions in Table II. (·) Monte Carlo solutions; (△) mean value and center of 90% confidence region; (○) original sand solution; (+) original silt solution; (∗) original clay solution.

In Figs. 15 and 16, 90%-confidence regions for the Monte Carlo solutions are juxtaposed with those for the model–data solutions reported in Sec. V (henceforth called the ''real'' solutions) at 33 and 93 kHz, respectively. In general, the confidence regions of the Monte Carlo solutions fall within those of the real solutions. However, the plots suggest that for normal incidence over sand at 33 kHz, variations in M_ϕ and w_2 are larger than suggested by the limited number of field measurements. The same can be said of w_2 estimates for oblique incidence over clay at 33 kHz. This implies that analysis of larger data sets could yield greater solution variability than what is currently observed.

As observed for the real solutions, the Monte Carlo estimates of M_ϕ for sand exhibit greater variability than for fines—with measures of standard deviation ranging from 0.08 to 0.38 and 0.04 to 0.10, respectively. As a percentage of the mean value, w_2 standard deviations for fine-grain simulations (15%–46%) are typically larger than those for sand (8%–27%). Standard deviations for σ_v are typically less than 10% of the mean value, with exceptions for sand at 33 kHz.

Also, a significant anticorrelation between M_ϕ and w_2 for sand substrates is seen in the real solutions and in the Monte Carlo solutions, with correlation coefficients ranging from -0.64 to -0.94. In the temporal model of acoustic backscatter, increasing either parameter decreases signal peak amplitude, and vice versa. In nature, these quantities are expected to be negatively correlated—i.e., coarser sediments (lower M_ϕ) exhibit more energy in the relief energy density spectrum (larger w_2). When the ''true'' signal is contaminated by ''noise'' the parameters also tend to adjust in opposite directions.

There also appears to be modest anticorrelation between M_ϕ and σ_v in solutions for fine-grain sediments. In these substrates, scattering from the sediment volume typically plays a larger role. An increase in either parameter raises the

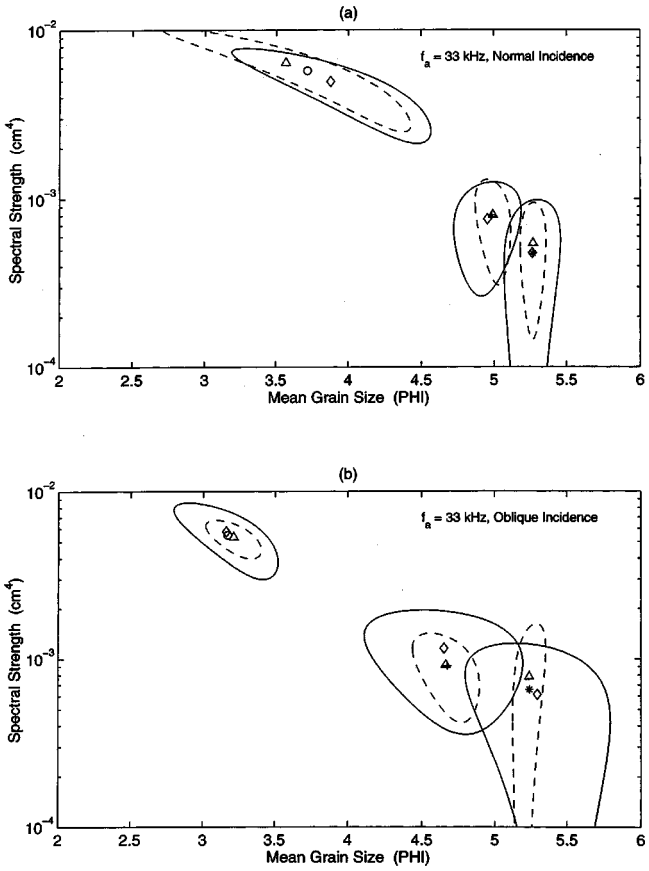


FIG. 15. Confidence regions for Monte Carlo and real solutions (33 kHz). Real solutions: Solid lines ———90%-confidence regions; (◇) mean values. Monte Carlo solutions: Dashed lines — — — 90%-confidence regions; (△) mean values; (○) original sand solution; (+) original silt solution; (*) original clay solution. Transducer orientation: (a) Normal; (b) Oblique.

calculated energy in the signal tail, and apparently the two parameters compete to fit this section of the signal. Too little is known about *in situ* sediment volume scattering characteristics to warrant a physical interpretation.

In theory, values of the data covariance matrix (and thus the solution variance) can be decreased by averaging a larger number of echoes. However, in our data sets, processing ensembles much greater than 100 pings excessively filters the shape characteristics of the average envelope that are essential to the matching procedure. Furthermore, with large ensembles the requirement of bottom homogeneity is more likely to be violated—especially at high survey speeds.

In theory also, the data covariance matrix has potential application in model–data fitting. The nonweighted least-squares merit function of Eq. (5) was chosen over a variance-weighted approach in order to favor peak amplitude model–data matching—emphasizing extraction of mean grain size correlated parameters, such as impedance contrast. Future work with this technique will include testing of the full maximum likelihood estimation (MLE) paradigm; i.e., covariance matrix weighting of the model–data disparity. Variance weighting of each model–data sample disparity should improve model–data fitting at the leading and trailing signal edges—at the expense of precise peak amplitude matching. The effect, however, of the data's covariances should coerce the optimized model to assume the true ''shape'' of the measured envelope. Comparisons between full MLE optimization and peak amplitude (nonweighted) matching will be evaluated in future evolutions of this echo envelope sediment characterization algorithm.

VII. SUMMARY

The method for estimating sediment geoacoustic parameters presented here compares bottom returns measured by a calibrated, moderate beamwidth (10°–21°), vertically oriented (0°–15°) monostatic sonar, with an echo envelope model based on high-frequency (10–100 kHz) incoherent backscatter theory and sediment properties such as mean grain size (M_ϕ), interface roughness (w_2, γ), and sediment volume scattering statistics (σ_v). A two-stage average echo envelope matching procedure was described where: first, the sediment type (sand or fines) is established by iterating on the reflection coefficient to match the peak echo amplitude, and to establish a general fit with generic values of the remaining geoacoustic parameters; then, a three-parameter global optimization is performed using a combination of simulated annealing and downhill simplex searches over the allowable range of interface roughness spectral strength, sediment volume scattering coefficient, and a constrained range of reflection and bottom absorption coefficients correlated to mean grain size. In San Diego Bay, bottom echoes were collected at 33 and 93 kHz over substrates ranging from sand to clay. Application of the sediment characterization method to these data yielded solutions for grain size and geoacoustic properties that are consistent with ground-truth measurements.

The ground-truth measurements, consisting of bottom video, grain size analyses, environmental databases, and associated ranges of geoacoustic parameters, lack direct assessments of the modeled geoacoustic properties. This, and the small number of sites and regions evaluated, limits definitive assessment of the accuracy and robustness of the described inversion technique. Controlled, calibrated surveys over sites characterized for the complete range of geoacoustic parameters must eventually be employed to further evaluate and improve the efficacy of the echo envelope sediment characterization technique.

For the experiments described in this paper, analyses of the estimated geoacoustic parameters for different combinations of sediment type, frequency, and transducer orientation suggest that moderate frequencies (33 kHz) and normal incidence are more suitable for this method of sediment characterization. This may, in part, be due to limitations at high acoustic frequencies (e.g., 93 kHz) of the backscatter model's underlying Kirchhoff theory, and partly due to the simple temporal structure of the returns at lower acoustic frequencies (33 kHz)—simplifying calculation of the average echo envelope. Furthermore, approximate alignment of the transducer's maximum response axis at normal incidence insures that the maximum interface component of the backscattered signal will exceed the maximum volume contribution—a condition necessary for reducing ambiguity in the model–data matching procedure.

The ability to distinguish sands from fine-grain sediments was demonstrated based on acoustic estimation of

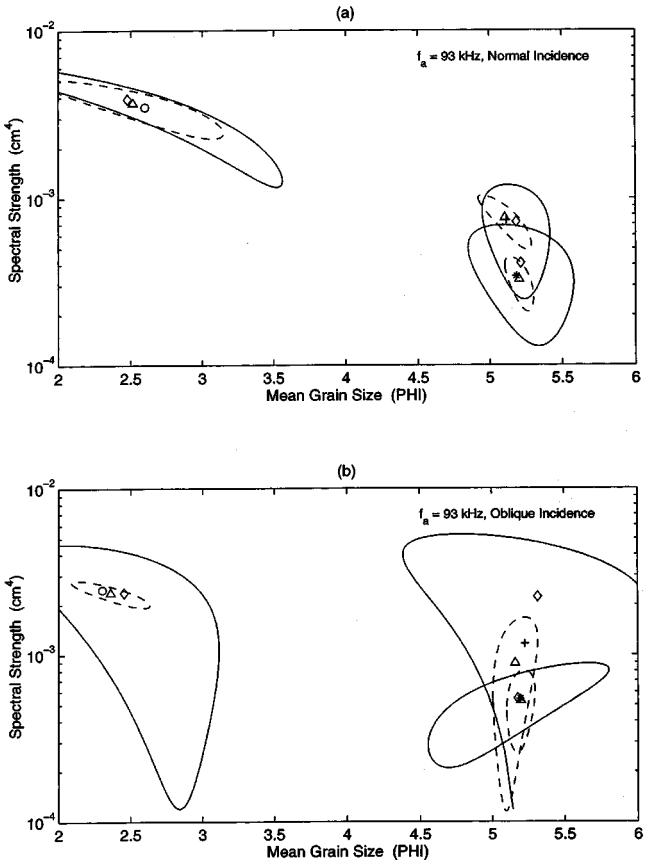


FIG. 16. Confidence regions for Monte Carlo and real solutions (93 kHz). Real solutions: Solid lines ———90%-confidence regions; (◇) mean values. Monte Carlo solutions: Dashed lines — — — 90%-confidence regions; (△) mean values; (○) original sand solution; (+) original silt solution; (∗) original clay solution. Transducer orientation: (a) Normal; (b) Oblique.

mean grain size alone. The creation of feature vectors from estimates of mean grain size (M_ϕ) and interface roughness spectral strength (w_2) showed promise for *intraclass* separation of silt and clay. Limitations on interface curvature dictated by the Kirchhoff approximation restrict application of this scheme to sediments having a large rms radius of curvature relative to the acoustic wavelength. This excludes extremely rough (rocky) substrates, or operation at high frequencies (>100 kHz). Furthermore, local deviations of sediment:water impedance ratio ($\rho\nu$) and sediment acoustic attenuation constant (κ_p) from generic values (mean values correlated with M_ϕ) will result in estimates of M_ϕ, w_2, and σ_v that are distorted from their true values.

Monte Carlo simulations based on a geoacoustic parameter solution set and the data's covariance matrix were described. In the mean, the Monte Carlo solutions agree with the original solution; however, for a given substrate, there is as much variability in the Monte Carlo solutions as there are in an ensemble of real solutions. Therefore, echo variability must be considered during parameter optimization by providing confidence limits on the results.

According to the observed spread of geoacoustic matches from measured signals and synthetic data, roughness spectral strength estimates (w_2) for sand substrates are relatively immune to raw echo variability, whereas mean grain size estimates (M_ϕ) are moderately affected. The opposite is observed for fine-grain substrates: M_ϕ estimates are relatively immune to raw echo variability, whereas w_2 estimates are significantly affected. A more thorough investigation of echo envelope averaging procedures and maximum likelihood–data matching techniques may result in methods to reduce the (M_ϕ,w_2) confidence regions.

Finally, the classification procedure introduces a degree of anticorrelation between M_ϕ and w_2, which is especially large for sand substrates. This trend is consistent with what is expected in nature, where the relief energy density spectra of coarser sediments (lower M_ϕ) exhibit more energy (higher W_2) than those of fine-grain substrates.

ACKNOWLEDGMENTS

We are grateful to the Space and Naval Warfare Systems Center (SPAWAR) for their assistance in collecting acoustic data and bottom samples, Reson Inc. for the generous loan of sonar transducers, and ORINCON Defense for their professional support. We thank Professor William Coles for his input on conducting the error analysis, and Jo Griffith for her help in preparing the illustrations. This work was funded by the Office of Naval Research under Contract No. N00014-94-1-0121.

APPENDIX: SUPPORTING TABLES

TABLE IV. Survey site ground truth. Substrate percentages may not add exactly to 100 due to round-off and a small gravel constituency.

Site	Sample index	Latitude deg min North	Longitude deg min West	Mean grain size (PHI)	Mean grain size (μm)	% Sand	% Silt	% Clay
Sand	1	32 40.760	117 13.653	1.9	268	93	2	4
	2	32 40.650	117 13.585	2.2	218	93	2	4
	3	32 40.647	117 13.626	1.7	308	90	3	4
Silt	1	32 42.265	117 13.927	4.1	58	76	14	10
	2	32 41.887	117 14.153	5.9	17	49	30	21
Clay	1	32 42.997	117 11.728	6.5	11	39	33	27
	2	32 42.995	117 11.767	6.8	9	36	34	29
	3	32 42.997	117 11.814	6.6	10	38	33	29

TABLE V. Survey measurement characteristics. Normal incidence measurements refer to those for which the central axis of the transducer's radiation pattern is aligned with the bottom normal (mean $\theta_T \sim 0$). Oblique incidence measurements exhibit elevation angles roughly corresponding to $\theta_{3dB}/2$ (8°–12° and 6°–9°, respectively, at 33 and 93 kHz). Columns 3–6 represent mean or approximate values.

Site	Freq (kHz)	Transducer orientation	Transducer elevation angle θ_T (deg)	Transducer altitude alt (m)	Along-track 3-dB footprint D_3 (m)	Along-track 6-dB footprint D_6 (m)	Transmission source level dB *re*: 1 μPa @ 1 m	Figure label
Sand	33	Normal	2.0	13.5	5.0	7.0	197.8	9(a)
		Oblique	12.0	13.5	5.0	7.0	197.8	9(b)
	93	Normal	2.0	13.5	2.4	3.1	191.4	10(a)
		Oblique	6.5	14	2.5	3.2	191.4	10(b)
Silt	33	Normal	2.0	16	5.9	8.3	192.4	9(c)
		Oblique	8.0	16.5	6.1	8.5	192.4	9(d)
	93	Normal	2.0	19	3.3	4.3	192.3	10(c)
		Oblique	8.5	19	3.3	4.3	192.3	10(d)
Clay	33	Normal	0	12	4.5	6.2	197.8	9(e)
		Oblique	12.0	11	4.1	5.7	197.8	9(f)
	93	Normal	1.0	12.5	2.2	2.8	191.4	10(e)
		Oblique	7.0	13	2.3	3.0	191.4	10(f)

TABLE VI. Sediment classification statistics.

Site	Freq (kHz)	Transducer orientation	Mean M_ϕ	Stdv M_ϕ	γ	Mean w_2 (cm^4)	Stdv w_2 (cm^4)	Mean σ_v (m^{-1})	Stdv σ_v (m^{-1})
Sand	33	Normal	3.88	0.28	3.00	0.004 96	0.001 13	0.202	0.022
		Oblique	3.16	0.14	3.00	0.005 77	0.001 09	0.041	0.020
	93	Normal	2.48	0.44	3.00	0.003 91	0.001 10	0.945	0.343
		Oblique	2.45	0.27	3.00	0.002 36	0.000 90	0.104	0.112
Silt	33	Normal	4.96	0.09	3.30	0.000 76	0.000 20	0.071	0.010
		Oblique	4.65	0.22	3.30	0.001 15	0.000 32	0.091	0.016
	93	Normal	5.18	0.09	3.30	0.000 72	0.000 19	0.261	0.070
		Oblique	5.32	0.38	3.30	0.002 23	0.001 24	0.314	0.227
Clay	33	Normal	5.26	0.08	3.30	0.000 48	0.000 21	0.048	0.022
		Oblique	5.30	0.20	3.30	0.000 61	0.000 24	0.047	0.015
	93	Normal	5.21	0.15	3.30	0.000 41	0.000 11	0.102	0.040
		Oblique	5.18	0.24	3.30	0.000 55	0.000 13	0.170	0.053

TABLE VII. Sediment classification: Parameter correlation.

Site	Freq. (kHz)	Transducer orientation	(M_ϕ, w_2)	(M_ϕ, σ_v)	(w_2, σ_v)
Sand	33	Normal	−0.86	−0.79	+0.37
		Oblique	−0.72	−0.78	−0.48
	93	Normal	−0.96	−0.79	+0.77
		Oblique	−0.58	+0.66	+0.07
Silt	33	Normal	+0.19	+0.07	−0.6
		Oblique	−0.27	+0.17	−0.74
	93	Normal	−0.23	−0.87	+0.44
		Oblique	−0.58	−0.07	+0.60
Clay	33	Normal	+0.18	−0.50	−0.11
		Oblique	−0.31	−0.64	−0.2
	93	Normal	−0.34	−0.77	+0.17
		Oblique	+0.77	+0.26	+0.07

[1] N. G. Pace and R. V. Ceen, "Seabed classification using the backscattering of normally incident broadband acoustic pulses," Hydrographic J. **26**, 9–16 (1982).

[2] R. Chivers, N. Emerson, and D. R. Burns, "New acoustic processing for underway surveying," Hydrographic J. **56**, 9–17 (1990).

[3] L. A. Mayer. UNB-OMG/UNH-CCOM Multibeam Sonar Traning Course Notes, November 2002.

[4] A. S. Tsehmahman, W. T. Collins, and B. T. Prager, "Acoustic seabed classification and correlation analysis of sediment properties by QTC view," in Proceedings of IEEE OCEANS 97, pp. 921–926, 1997.

[5] G. J. Heald and N. G. Pace, "An analysis of 1st and 2nd backscatter for seabed classification," in Proceedings of III European Conference on Underwater Acoustics, Crete, June 1996, pp. 649–654.

[6] G. J. Heald and N. G. Pace, "Implications of bi-static treatment for the second echo from a normal incidence sonar," in Proceedings: 16th International Congress on Acoustics and 135th Meeting Acoustical Society of America, Seattle, Washington, June 1998, Vol. IV, pp. 3009–3010.

[7] S. G. Schock, L. R. LeBlanc, and L. A. Mayer, "Chirp subbottom profiler for quantitative sediment analysis," Geophysics **54**(4), 445–450 (1989).

[8] L. R. LeBlanc, L. Mayer, M. Rufino, and J. King, ''Marine sediment classification using the chirp sonar,'' J. Acoust. Soc. Am. **91**(1), 107–115 (1991).
[9] M. V. Berry, ''The statistical properties of echoes diffracted from rough surfaces,'' Philos. Trans. R. Soc. London, Ser. A **273**, 611–654 (1973).
[10] E. H. Nesbitt, ''Estimation of sea bottom parameters using acoustic backscattering at vertical incidence,'' Master's thesis, University of Washington, 1988.
[11] D. R. Jackson and E. Nesbitt, ''Bottom classification using backscattering at vertical incidence,'' J. Acoust. Soc. Am. Suppl. 1 **83**, S80 (1988).
[12] X. Lurton and E. Pouliquen, ''Identification de la nature du fond de la mer à l'aide de signaux d'écho-sondeurs. II. Méthode d'identification et résulatats experimentaux,'' Acta Acust. (European Acoustics Association) **2**(3), 187–194 (1994).
[13] D. D. Sternlicht and C. P. de Moustier, ''Time-dependent seafloor acoustic backscatter (10–100 kHz),'' J. Acoust. Soc. Am. **114**, 2709–2725 (2003).
[14] D. R. Jackson, D. P. Winebrenner, and A. Ishimaru, ''Application of the composite roughness model to high-frequency bottom backscattering,'' J. Acoust. Soc. Am. **79**(5), 1410–1422 (1986).
[15] C. K. Wentworth, ''A scale of grade and class terms for clastic sediments,'' J. Geol. **30**(5), 377–392 (1922).
[16] W. C. Krumbein, ''Application of logarithmic moments to size frequency distribution of sediments,'' J. Sediment. Petrol. **6**, 35–47 (1936).
[17] Applied Physics Laboratory. High-Frequency Ocean Environmental Acoustic Models Handbook. Technical report, APL-UW TR9407, AEAS 9501 University of Washington, 1994.
[18] E. L. Hamilton, ''Compressional wave attenuation in marine sediments,'' Geophysics **37**, 620–646 (1972).
[19] Z. H. Michalopoulou, D. Alexandrou, and C. de Moustier, ''Application of a maximum likelihood processor to acoustic backscatter for the estimation of seafloor roughness parameters,'' J. Acoust. Soc. Am. **95**(5), 2467–2477 (1994).
[20] H. Matsumoto, R. P. Dziak, and C. Fox, ''Estimation of seafloor microtopographic roughness through modeling of acoustic backscatter data recorded by multi-beam systems,'' J. Acoust. Soc. Am. **94**(5), 2776–2787 (1993).
[21] D. D. Sternlicht and C. P. de Moustier, ''Temporal modeling of high frequency (30–100 kHz) acoustic seafloor backscatter: Shallow water results,'' in High Frequency Acoustics in Shallow Water, CP-45, pp. 509–516, Lerici, Italy, July 1997. NATO SACLANT Undersea Research Centre.
[22] G. E. Forsythe, M. A. Moler, and B. Cleve, *Computer Methods for Mathematical Computations* (Prentice-Hall, Englewood Cliffs, NJ, 1977).
[23] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, ''Optimization by simulated annealing,'' Science **220**(4598), 671–680 (1983).
[24] M. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. E. Teller, ''Equations of state calculations by fast computing machines,'' J. Chem. Phys. **21**, 1087–1091 (1953).
[25] W. H. Press and S. A. Teukolsky, ''Simulated annealing optimization over continuous spaces,'' Comput. Phys. **5**, 426–429 (1991).
[26] 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).
[27] D. D. Sternlicht, ''High Frequency Acoustic Remote Sensing of Seafloor Characteristics,'' Ph.D. thesis, Univeristy of California, San Diego, 1999.
[28] R. H. Plumb, ''Procedure for handling and chemical analysis of sediment and water samples,'' Technical Report TR-EPA/CE-81-1, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Mississippi, 1981.
[29] P. D. Mourad and D. R. Jackson, ''High frequency sonar equation models for bottom backscattering and forward loss,'' in Proceedings of IEEE OCEANS 89, pp. 1168–1175, 1989.
[30] A. P. Lyons and T. H. Orsi, ''The effect of a layer of varying density on high-frequency reflection, forward loss, and backscatter,'' IEEE J. Ocean. Eng. **23**(4), 411–422 (1998).
[31] S. Stanic, K. B. Briggs, P. Fleischer, W. B. Sawyer, and R. I. Ray, ''High-frequency acoustic backscattering,'' J. Acoust. Soc. Am. **85**(1), 125–136 (1989).
[32] D. R. Jackson, J. J. Crisp, and P. A. Thompson, ''High-frequency bottom backscatter measurements in shallow water,'' J. Acoust. Soc. Am. **80**(4), 1188–1199 (1986).
[33] D. R. Jackson and K. B. Briggs, ''High-frequency bottom backscattering: Roughness versus sediment volume scattering,'' J. Acoust. Soc. Am. **92**(2), 962–977 (1992).

Closed-form expressions for ocean reverberation and signal excess with mode stripping and Lambert's law

C. H. Harrison[a)]
SACLANT Undersea Research Centre, Viale S. Bartolomeo 400, 19138 La Spezia, Italy

(Received 18 November 2002; revised 12 August 2003; accepted 25 August 2003)

Closed-form expressions for two-way propagation and reverberation in variable depth ducts are derived for isovelocity water by using ray invariants and acoustic flux. These expressions include the transition to single mode propagation at long range. Three surface scattering laws are considered: Lambert, Lommel-Seeliger, and angle independent, and these are compared with a point target to give explicit signal-to-reverberation ratios. In particular, there is interesting and sometimes surprising behavior when the propagation obeys mode-stripping (the high angles are preferentially attenuated by bottom losses) whilst the scattering obeys Lambert's law (high angles are preferentially back-scattered). There may be conditions where the signal-to-reverberation ratio is independent of range so that there is no reverberation range limit. Bottom slope dependence of both target echo and reverberation is surprisingly weak. The implications of refraction are discussed. The angle dependence for a point or surface scatterer at a given range can be translated into arrival time, so it is possible to calculate the received pulse shape for one-way or two-way paths. Because the tail is exponential with a range-independent half-life that only depends on bottom reflection properties there is scope for extracting geoacoustic information from the pulse shape alone. This environmental time spread is also of use to sonar designers.
[DOI: 10.1121/1.1618240]

PACS numbers: 43.30.Gv [RAS] Pages: 2744–2756

I. INTRODUCTION

The background against which a sonar detects a target is often reverberation, and for this reason reverberation is of interest in its own right. However for a sonar designer where the signal is a target echo the signal-to-reverberation ratio (SRR) is of more interest still. Typically reverberation has two forms, one a locally uniform type, diffuse reverberation, the other, consisting of resolvable pointlike scatterers, geoclutter.[1] This paper concentrates on diffuse reverberation, and includes derivations of closed-form expressions for this quantity in a range-dependent environment. Although numerical models exist[2,3] the value of an analytical approach lies in the ability to see the relative importance of the many physical mechanisms. These include scattering phenomena: angle-dependence of the scattering law, geographic variation of the scattering strength, and local tilt of bottom facets; and propagation phenomena: range-dependent propagation intensity and range-dependent arrival angle at the scatterer.

Thus using a closed-form expression one can see straight away which aspects need to be modelled carefully when designing a numerical model, what environmental quantities need to be mapped and data-based, how many environmental parameters one could expect to get from an inversion, and so on. In the context of broad band active sonar an even stronger case can be made for simple flux and analytical models. Most broad band sonars do not rely on knowing exactly where every interference peak is, rather they respond to frequency-smoothed propagation (this is more or less the same as proportional range smoothing[4]), and so the smooth analytical expression is a useful point of comparison. In fact the expressions form useful benchmarks for testing other models, and the formulas here have already been used in this manner.[5]

This analytical approach is based on Weston's formulas for propagation in a lossy (i.e., mode-stripping) range-dependent environment.[6–8] The method is extended by investigating the effect of an angle-separable scattering law, in particular, Lambert's law and then by comparing the reverberation result with the echo from a point target. Of course Lambert's law may not be true everywhere (or anywhere) and one aim is that this approach to calculation of reverberation and propagation will provide ideas for experiments that give a more definitive answer. Then one can use the same mathematical technique on different scattering laws. Although Weston's approach includes refraction, and the implications of refraction will be discussed here, the results in this paper are for isovelocity water.

The angle dependence associated with the mode stripping can also be translated into an arrival time dependence. Thus there is potential to make environmental deductions from the pulse shape returned by a localized scatterer. These possibilities are considered later.

II. ONE-WAY PROPAGATION

A. Range-independent propagation

From a simple physics point of view propagation in shallow water may follow several well known distinct range laws[9] including spherical spreading, cylindrical spreading, mode-stripping and single mode propagation. The next section concentrates on mode-stripping because, for bottom re-

a)Electronic mail: harrison@saclantc.nato.int

verberation, it is more or less inevitable at practical ranges and frequencies. Some variations are investigated in subsequent sections before going on to range-dependent environments. These help to define the range bounds of mode-stripping.

The term mode-stripping refers to the fact that higher order modes have higher attenuations so the number of significant modes reduces as range increases. In fact the magnitude of the effect can be derived using rays just as well as modes.[10] For simplicity this paper adopts the ray, reflection, angle description of propagation. The most effective rays are inside the seabed's critical angle and have a reflection loss (dB) proportional to angle $R_{\mathrm{dB}}=10\log(R)=\alpha_{\mathrm{dB}}\theta$. In other words, the reflection coefficient R can be written as

$$R=\exp(-\alpha\theta) \tag{1}$$

with $\alpha=\alpha_{\mathrm{dB}}/(10\log(e))$. [Note that $\alpha_{\mathrm{dB}}=10\log(\exp(-\alpha))$, it is therefore not simply the quantity α in dBs.] This relation is certainly a good approximation for a half-space[9] and is also a reasonable approximation for layered sediments.

The number of reflections is determined by the ray cycle distance r_c which for isovelocity is given in terms of the water depth H by

$$r_c=2H/\tan\theta. \tag{2}$$

This results in an exponential decay in range but a gaussian angle distribution after making the small angle approximation,

$$\mathrm{Atten}=10^{(-R_{\mathrm{dB}}\times r/r_c)/10}=\exp\left(-\frac{\alpha\theta^2}{2H}r\right). \tag{3}$$

Considering an eigenray sum where each eigenray contributes an intensity of r^{-2}, there are four groups of eigenrays, each with angular separation of $2H/r$. So the sum Σ_n can be converted to an integral $\int dn$ which in turn can be written as $(r/2H)\int d\theta$. Including the boundary losses and integrating up to the seabed's critical angle (assuming infinite losses beyond) the result is

$$I=\frac{2}{rH}\int_0^{\theta_c}\exp\left(-\frac{\alpha\theta^2}{2H}r\right)d\theta$$

$$=\sqrt{\frac{2\pi}{H\alpha r^3}}\,\mathrm{erf}(\sqrt{\alpha r/2H}\,\theta_c) \tag{4}$$

which becomes the familiar three-halves law or $15\log(r)$ (Ref. 9) for long range [i.e., the Gaussian is narrow compared with the critical angle so that the argument of the erf is large and $\mathrm{erf}(\sqrt{\alpha r/2H}\,\theta_c)=1$]. One can see in Eq. (4) that if the large angle forms had been retained the only part of the exponent that would change significantly would be the part that is already very small. It is easy to show numerically that deviations of the ratio from unity are guaranteed to be less than 1%, 2%, 3% for critical angles 23°, 33°, 40°, respectively (regardless of the Gaussian's width). For reference 0.1 dB is 2.3%. Comparable small errors are associated with the slightly different integrals that arise later.

As range increases, the width of the gaussian decreases until ultimately its width is comparable to the angular separation of modes [$\sim\lambda/2H$ (Ref. 11)]. Near this point it is as if the Gaussian were sampled at intervals $\lambda/2H$. It is easy to see that for one remaining mode [average $\phi_1^2(z)=1/H$] at angle $\lambda/2H$ (Ref. 9) the intensity is

$$I=\frac{\lambda}{rH^2}\exp(-\alpha\lambda^2 r/(8H^3)). \tag{5}$$

Note that this effect is quite distinct from mode cutoff; it is true that many of the modes may indeed have cutoff, but the first mode is much less attenuated by reflection than all the others.

Rather than rely on two separate formulas for mode stripping and single mode, i.e., Eqs. (4) and (5), as Weston did one can combine the two as follows. The discrete modes can be represented graphically in the angle continuum by contiguous rectangles of height $\exp(-\alpha\theta_m^2 r/2H)$ and width $\lambda/2H$, each rectangle centered on the angle $\theta_m=m\lambda/2H$; $m=1,2,\ldots$. One can estimate the effect of the transition to single mode by adding the mode-1 contribution, i.e., $\lambda/2H\exp(-\alpha\theta_1^2 r/2H)$ to the remaining continuum contribution. Thus the continuum contribution needs to be truncated at the edge of mode-1, i.e., at angle $\theta'=(\theta_1+\theta_2)/2=\theta_1\times 1.5$ rather than zero. (One can easily demonstrate that this is numerically a good approximation by comparing the result with the complete *sum* of Gaussians rather than their integral.) Thus any of the subsequent integrals of the form

$$I=a_0\int_0^{\psi}F(\theta)\exp(-g_0\theta^2)\,d\theta \tag{6}$$

can be converted to

$$I=a_0\left(\int_{\theta'}^{\psi}F(\theta)\exp(-g_0\theta^2)\,d\theta+F(\theta_1)\exp(-g_0\theta_1^2)\times\lambda/2H\right) \tag{7}$$

with $\theta_1=\lambda/2H$; $\theta'=1.5\,\theta_1$ to include this single-remaining-mode effect in a continuous fashion.

There are many variants of these formulas for propagation[7,10] and also for ambient noise.[12] Later they are adapted for reverberation and there are still more variants covering refraction[13] and bistatic sonar.[14]

B. Range-dependent propagation

Weston introduced the concept of a ray invariant[15] as a means of working out the ray angle given the water depth and sound speed profile but without actually tracing rays. The method is used in Refs. 6–10, 12, 14 and explained in Ref. 11. It is closely allied to the adiabatic mode approximation,[11,16] and both are valid as long as the environment changes slowly. Note that a truly adiabatic process is reversible (upslope/downslope) and it is the individual modes that behave adiabatically, not the environment. This is because the slow variation requirement is that the ray cycle distance should be small compared with the horizontal scale of the bathymetric features. Thus high order modes (steep rays) may behave adiabatically while low order modes may not. Once the mode reaches cut-off or the ray reaches the

critical angle it can no longer contribute although ray angles are well predicted by the ray invariant up to this point.

In the attenuation term of Eq. (3) now both R_{dB} and r_c vary with range because the angle changes. However, in its simplest isovelocity form, the ray invariant states that

$$H \sin\theta = H_s \sin\theta_s = H_r \sin\theta_r = \text{const}, \tag{8}$$

so the attenuation can be written as

$$\begin{aligned}\text{Atten} &= \exp\left(-\int_0^r \frac{\alpha\theta'^2}{2H'}\, dr' \right) \\ &\cong \exp\left\{ -\alpha H_s^2 \int_0^r \frac{\theta_s^2\, dr'}{2H^3(r')} \right\}. \end{aligned} \tag{9}$$

The intensity formulas of Refs. 6 and 7 can be derived from the incoherent adiabatic mode sum[11]

$$I = 2\pi \sum_m \phi_m^2(z_s)\phi_m^2(z_r) \Big/ \int K_m\, dr \tag{10}$$

by converting it into an angle integral and noting that the depth averages of the normalized modes $\phi_m(z_s)$, $\phi_m(z_r)$ are, respectively, H_s^{-1} and H_r^{-1}. Combining this with the attenuation the result is

$$I = \frac{2}{rH_r} \int_0^{\theta_{\mathrm{lim}}} \exp\left\{ -(\alpha H_s^2/2) \int_0^r \frac{dr'}{H^3(r')}\, \theta_s^2 \right\} d\theta_s. \tag{11}$$

The range integral is just a function of r, say $J(r)$, so whatever its value is, the angle integrand is still a Gaussian and the angle integral can be solved. The angle limit θ_{lim} is the value of θ_s (at the source) for which the ray hits the critical angle somewhere, which will be at a minimum depth H_c in the profile (possibly H_s or H_r) given by

$$\theta_{\mathrm{lim}} H_s = \theta_c H_c. \tag{12}$$

Weston[6] defined an effective depth

$$H_{\mathrm{eff}} = (H_r^2 H_s^2/r) \int_0^r \frac{dr'}{H^3(r')} \tag{13}$$

such that on substitution into Eq. (11) one obtains

$$I = \sqrt{\frac{2\pi}{\alpha r^3 H_{\mathrm{eff}}}}\, \mathrm{erf}\left\{ \sqrt{\frac{\alpha r H_{\mathrm{eff}}}{2}} \frac{\theta_c H_c}{H_r H_s} \right\}. \tag{14}$$

Immediately one sees that the formula has the same form as the range-independent case. Also one sees that the intensity obeys reciprocity, because it only depends on the depths at the ends, the separation of source and receiver, and the minimum depth H_c. This simplifies matters when considering two-way paths.

It is emphasised that throughout this paper the formulas containing H_{eff} are valid for general bottom profiles provided they are slowly varying. All the depths H_{eff}, H_r, H_c are known functions of range which can be calculated from the given bottom profile $H(r)$.

Noting that the range-dependent formula started in the form of Eq. (11) one can incorporate the single-remaining-mode effect by applying Eqs. (6) and (7). Thus Eq. (14) becomes

$$\begin{aligned} I = \sqrt{\frac{2\pi}{\alpha r^3 H_{\mathrm{eff}}}} \Bigg(& \mathrm{erf}\left\{ \sqrt{\frac{\alpha r H_{\mathrm{eff}}}{2}} \frac{\theta_c H_c}{H_r H_s} \right\} \\ & - \mathrm{erf}\left\{ \sqrt{\frac{\alpha r H_{\mathrm{eff}}}{2}} \frac{\theta'}{H_r} \right\} \Bigg) + \frac{\lambda}{r H_r H_s} \exp\left(-\frac{\alpha H_{\mathrm{eff}} \theta_1^2 r}{2H_r^2} \right), \end{aligned} \tag{15}$$

where $\theta_1 = \lambda/2H_s$ and $\theta' = \mathrm{MIN}(1.5\times\theta_1, H_c\theta_c/H_s)$. Note that the single mode term is still reciprocal, as it should be. It is worth stopping to check the origin of this term in Eq. (11). Following a ray initially at θ_s, the ray steepens in the upslope case, resulting in the exponential which is a Gaussian function of the initial ray angle θ_s. Regardless of how much the modes have strengthened or weakened along the way, the Gaussian at the receiver range has a certain width, and mode one at the receiver propagates at angle $\theta_r = \lambda/2H_r$. It is this latter angle that determines the truncation width Eq. (7). However putting this back into the integral in terms of θ_s results, from Eq. (8), in $\theta_s = \lambda/2H_s$. Thus, up-slope or down-slope, the same size chunk is always taken out of the middle of the θ_s integral to represent the single-remaining-mode effect. For this reason the effect is more likely to be seen up-slope than down because the Gaussian from which the chunk is extracted is narrower.

C. Mode stripping on a uniform slope

On a uniform slope with gradient ϵ the various water depths are given by

$$\begin{aligned} H(r) &= H_s + \epsilon r, \\ H_{\mathrm{eff}} &= (H_s + H_r)/2 = H_s + \epsilon r/2, \\ H_c &= \mathrm{MIN}(H_s, H_r) \end{aligned} \tag{16}$$

but

$$H_s H_r / H_c = \mathrm{MAX}(H_s, H_r) = H_s + \mathrm{MAX}(\epsilon, 0) r \tag{17}$$

so

$$\begin{aligned} I = \sqrt{\frac{2\pi}{(H_s + \epsilon r/2)\alpha r^3}}\, & \mathrm{erf}\{ \sqrt{\alpha r (H_s + \epsilon r/2)/2}\ \theta_c / \\ & (H_s + \mathrm{MAX}(\epsilon, 0) r) \} \end{aligned} \tag{18}$$

or

$$I = \sqrt{\frac{2\pi}{(H_s + \epsilon r/2)\alpha r^3}} \tag{19}$$

for large arguments.

Surprisingly almost all the ray angle and reflection loss effects are counteracted by the simple spreading caused by the change in water depth at the receiver. For instance, halving the depth upslope results in $H_{\mathrm{eff}} = 3/4\, H_s$ and only 0.63 dB rise in one-way path intensity. These effects are demonstrated graphically later along with the reverberation formulas. It is possible to confirm the magnitudes, for instance in the case of a wedge, by numerically calculating the changing ray angles and summing the attenuated contributions.

D. Refraction effects

Analytical formulas exist[13] for environments with uniform sound speed gradient and constant depth. Qualitatively the effects can be understood as follows. Because the rays are curved there is a low angle duct with single boundary interaction as well as the high angle duct with two-boundary interaction considered here. Thus the two-boundary contribution tends to behave as the mode-stripping regime except that there is a small gap in the middle of the Gaussian angle distribution filled by the single-boundary contribution. At long range, therefore, the mode-stripping $r^{-3/2}$ range law reverts to the exponential decay of the single-boundary duct. (The decay is exponential because the cycle distance is proportional to angle.) This behavior is similar mathematically to the isovelocity transition to single mode already discussed. Naturally the solution depends on the depth of the source and receiver, and there is a different result for scattering from the high and the low sound speed sides of the duct. Nevertheless general behavior is similar to the isovelocity case.

III. SCATTERING LAWS

One of the aims of this paper is to combine analytical propagation with scattering, but one needs to know the scattering law to do this, and generally this is poorly known; to date, there is no database of scattering strength. Even the dependence on angle is not known for certain, although Lambert's law is often invoked especially in the context of inversion.[17,18] While it is possible to make local measurements of both scattering and reflection properties[19] it is more difficult to deduce the scattering angle dependence from long range reverberation measurements. For this reason interpretations of long range reverberation measurements such as Refs. 20 and 21 tend to assume Lambert's law and concentrate on inverting for bottom reflection properties.

In the following the scattering coefficient S is defined as the linear analogue of scattering strength (scattering strength$=10\log S$). It is the ratio of emitted intensity at one meter from an elementary area to the incident intensity. It is convenient to be able to separate the scattering coefficient into functions of incoming θ_{in} and outgoing θ_{out} elevation angles. One could also separate out the dependence on bistatic angle β and position, i.e., scatterer number i, thus

$$S_i(\theta_{\text{in}},\theta_{\text{out}},\beta)=S_{\text{in}}(\theta_{\text{in}})S_{\text{out}}(\theta_{\text{out}})S_{\text{bis}}(\beta)F_i\,. \tag{20}$$

A possible objection is that forward scatter, near specular, is definitely not separable (since the specular reflection direction depends on the incident direction), but near-specular scatter is most easily handled as a multiply reflected, not scattered path, in other words the 'direct blast'. Thus (in contrast to, e.g., Ref. 22) it is easier to think of a scattered part that is always separable and an additional part that is neglected in the reverberation calculation because it is already included in the direct blast calculation.

A. Lambert's law

Lambert's law, in the context of underwater acoustics is usually stated as

$$S=\mu\sin(\theta_{\text{in}})\sin(\theta_{\text{out}}) \tag{21}$$

where μ is an empirical constant, typically $\mu=10^{-2.7}$ (i.e., -27 dBs).

B. Lommel-Seeliger

An alternative bottom scattering law is the Lommel-Seeliger law[23]

$$S=\mu_1\sin(\theta_{\text{in}})\sin(\theta_{\text{out}})/\{\sin(\theta_{\text{in}})+\sin(\theta_{\text{out}})\}. \tag{22}$$

This originated in the optics of lunar astronomy and is still widely used in determining the shapes of asteroids and the terrain of planetlike objects.[24] Unfortunately, although it obeys reciprocity, it is not separable like Eq. (20). This means that numerical reverberation calculations cannot take advantage of the potential speed increase, but, as will be seen later, the monostatic, large critical angle case is still analytically tractable.

C. Other bottom scattering laws

There are many other possibilities when one considers multiple layers and volume scattering. For instance, if there is penetration into the sediment at long range because of low sediment sound speeds, then there is the possibility of scattering from volume inhomogeneities in the sediment and scattering from a lower boundary.[25,26] Assuming that there is a single scatter in the round trip path (all other interactions being reflection or refraction) then both volume and lower surface scattering have their steep angle responses mapped by Snell's law into low angles in the water. In the case of a lower Lambert's law scattering layer, almost any low angle in the water maps to a fixed angle in the sediment which selects a fixed value of $\sin\theta$. Therefore the scattering strength will tend towards a fixed value for low angles rather than the rapid fall-off of $10\log(\sin\theta)$. However this amplitude is also modified by the (in-going and out-going) transmission coefficient into the sediment which typically rises linearly with θ but much more rapidly than $\sin(\theta)$. The combination of sediment sound speed, roughness of the upper and lower boundary of a sediment layer, and volume scattering means that it is possible to find quite complicated dependences on angle even with a very unsophisticated theory.

IV. TWO-WAY PROPAGATION AND REVERBERATION

A. Range-independent reverberation

1. Two-way propagation; point targets

In a bistatic arrangement a point target at range r_1 from a source and range r_2 from a receiver would first receive multiple ray arrivals then retransmit them over a spread of angles to the receiver. Thus, defining the linear analogue S_T of target strength (target strength$=10\log S_T$) one obtains

$$I=\frac{1}{r_1 H}\int_{-\theta_c}^{\theta_c}\exp\left(-\frac{\alpha\theta_1^2}{2H}r_1\right)d\theta_1\frac{1}{r_2 H}\times\int_{-\theta_c}^{\theta_c}\exp\left(-\frac{\alpha\theta_2^2}{2H}r_2\right)d\theta_2\,S_T \tag{23}$$

and for monostatic this reduces to

$$I = \frac{2\pi S_T}{\alpha H r^3}(\mathrm{erf}(\sqrt{\alpha r/2H}\,\theta_c))^2. \tag{24}$$

Modal effects will be included in the equivalent range-dependent formulas in Sec. IV B.

2. Reverberation with angle-independent scattering strength

For comparison purposes only we introduce the concept of an angle-independent scattering surface which is often invoked in the context of signal-to-reverberation-ratio calculations (e.g., Ref. 27, Sec. 8.6). To avoid possible problems with conservation of energy for very low grazing angles the scattering strength S_B is defined to be constant for all angles above some limit ϵ and zero below. The effect of ϵ on the subsequent angle integrals is assumed to be negligible. Considering an elementary area determined by the spatial pulse length $p = ct_p/2$ and the spatial horizontal beam width $r\Phi$ the reverberation is

$$I = \frac{Q^2 \pi S_B \Phi p}{2\alpha H r^2}(\mathrm{erf}(\sqrt{\alpha r/2H}\,\theta_c))^2. \tag{25}$$

It falls off more slowly than the target echo because of the widening of the scattering area with range. Eventually at long range reverberation will dominate, and at the transition there is a "reverberation limit." Note that the one-way propagation loss includes four eigenrays for each θ (a full set of up- and down-going paths at source and receiver), but with a notional receiver on the seabed this reduces to two. A factor Q has been introduced throughout to allow for this. To be consistent with a scattering-experimentalist's viewpoint where there is only one eigenray illuminating the seabed this factor is set to unity for the outward and return paths. Otherwise if the full field is taken to be the incident field at the seabed it should be set to 2. The correct value of Q depends on the definitions used during construction of the scattering coefficients.

3. Reverberation with Lambert's law scatterers

Assuming that the scattering strength is separable in the incoming (θ_1) and outgoing (θ_2) ray angles,

$$S = \mu \sin(\theta_1)\sin(\theta_2), \tag{26}$$

the intensity

$$I = \frac{Q^2}{4r^2H^2}\int_{-\theta_c}^{\theta_c}\int_{-\theta_c}^{\theta_c}|\sin\theta_1||\sin\theta_2|\exp\left(-\frac{\alpha\theta_1^2}{2H}r\right) \times \exp\left(-\frac{\alpha\theta_2^2}{2H}r\right)d\theta_1\, d\theta_2\, \mu r\Phi p \tag{27}$$

also separates to give a closed form solution for the reverberation

$$I = \left(\frac{Q}{2rH}\int_{-\theta_c}^{\theta_c}\sin(\theta)\exp\left(-\frac{\alpha\theta^2}{2H}r\right)d\theta\right)^2 \mu r\Phi p = \frac{Q^2\mu}{\alpha^2 r^3}\Phi p(1-\exp(-\alpha r\,\theta_c^2/2H))^2. \tag{28}$$

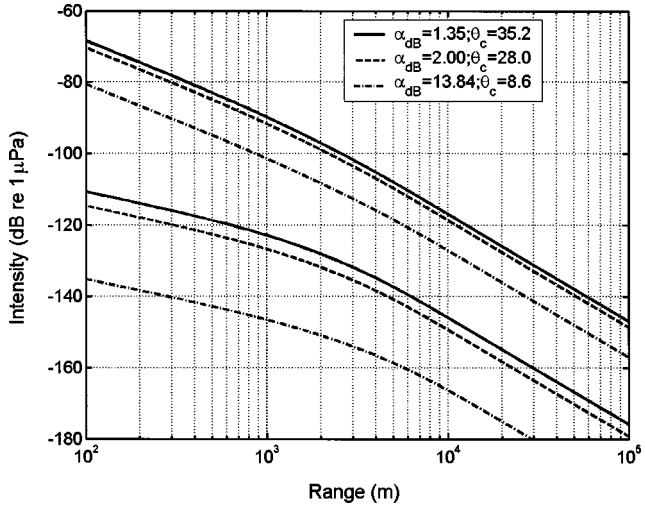


FIG. 1. Target echo (upper three curves) and reverberation (lower three curves) in a mode-stripping duct of depth 100 m. The solid, dashed, and dashed–dotted line pairs are for three realistic environments with reflection loss gradient α_{dB} (dB/rad) and θ_c (degrees) indicated. The quantity α ($\alpha = \alpha_{\mathrm{dB}}/10\log(e)$) has corresponding values 0.31, 0.48, 3.19 rad^{-1}. Note that the bend occurs at a few km in all cases. Target and scattering parameters are given in the text.

In Eq. (24) the mode-stripping three-halves range law resulted in r^{-3} in two-way propagation when the gaussian was much narrower than the critical angle (i.e., in the long range limit). Interestingly the Lambert's law reverberation case in Eq. (28) shows exactly the same range dependence; the usual proportionality of scattering area to range that alters the range dependence in Eq. (25) is counteracted by mode stripping in this new integral.

In this long range regime the range-dependence of the reverberation is always r^{-3}; it neither depends on reflection properties α nor scattering strength μ. Furthermore in this regime the value of reverberation depends on the ratio μ/α^2, so μ and α are *inseparable* in an inversion of reverberation measurements. At shorter ranges where the critical angle truncates the gaussian the erf function of Eqs. (24) and (25) is replaced by a (1-exponential) and there are differences between the range dependences of reverberation and target echo. Another interesting point is that the dependence of reverberation on bottom loss (α) is stronger than that of the target. In fact, imposing low bottom loss increases the reverberation more than the target echo which seems counterintuitive.

These points are illustrated in Fig. 1 for several realistic values of α_{dB} (and α) and θ_c derived (using equations from Ref. 9) from Hamilton's data[28] (a slightly updated version of Ref. 29) and shown in Table I. The upper three curves are target echo and the lower three are reverberation. In all figures there are several quantities that are simple multipliers. For reverberation the multiplier is $\mu\Phi p$, and for target it is S_T. These are assumed to be

$p = ct_p/2 = 1500\times 0.01/2 = 7.5$ m (10 ms pulse length),

$\Phi = 1/60 = 0.0167$ radians (1° beam),

$\mu = 10^{-27/10}$,

TABLE I. Typical sediment parameters (sound speed, density, and absorption) (Ref. 27). These are converted to reflection loss gradient α_{dB}, critical angle θ_c, and the transition range factor r_o/H using equations from Ref. 9.

Type	Speed c_2 (m/s)	Specific gravity ρ_2	Absorption a_2 (dB/λ)	α_{dB} (dB/rad.)	θ_c (deg)	r_o/H $=2/(\alpha\theta_c^2)$
Coarse sand	1836	2.03	0.6	1.35	35.22	17.02
v. fine sand	1709	1.88	0.5	2.09	28.63	16.59
Silt	1615	1.74	0.3	2.82	21.75	21.4
Silty clay	1517	1.48	0.1	13.84	8.59	27.9

$S_T=10$ at 1 m ($10\log S_T=10$ dB at 1 m).

Similarly for clarity a source level of SL=0 dB *re* 1 μPa at 1 m is assumed. Thinking in terms of dBs one can easily see the effect of altering these parameters on any of the plots in this paper.

Interestingly, the transitional bend at several km in all curves in Fig. 1 does not move very much for most realistic geoacoustic parameters (and fixed water depth). The reason for this is that the transition occurs when $\alpha\theta_c^2 r/2H=1$, and if one substitutes for α with formulas from Ref. 9 one finds the transition range r_o to be

$$r_o=\frac{2H}{\alpha\theta_c^2}=H\times\frac{2\pi\theta_c}{a_2}\frac{\rho_1 c_2^2}{\rho_2 c_1^2}10\log(e). \tag{29}$$

Taking typical mean values for the quantities c_2, ρ_2, and a_2 as shown in Table I (acknowledging that there is a spread about the mean, particularly for absorption) one can see that the correlations between them conspire to keep the factor $2/(\alpha\theta_c^2)$ close to about 20. Thus with a 'half-life' of 2000 m in 100 m of water the formulas responsible for Fig. 1 fall to their asymptotic values after a few km.

4. Reverberation with Lommel-Seeliger scatterers

Following the procedure in the preceding section it is possible to find formulas for any separable scattering law provided the integral is tractable. Here, using the Lommel-Seeliger law as an example, it is shown that it may also be possible to solve with non-separable scattering laws. The equivalent of Eq. (27) with Lommel-Seeliger scatterers is

$$I=\frac{Q^2}{r^2H^2}\int_0^{\theta_c}\int_0^{\theta_c}\frac{\sin\theta_1\sin\theta_2}{(\sin\theta_1+\sin\theta_2)}\exp\left(-\frac{\alpha\theta_1^2}{2H}r\right)\times\exp\left(-\frac{\alpha\theta_2^2}{2H}r\right)d\theta_1\,d\theta_2\,\mu r\Phi p \tag{30}$$

which, as noted earlier, is not separable. Ignoring the difference between $\sin\theta$ and θ one can solve the integral. First note that for Lambert's law one could have solved Eq. (27) for the case where θ_c is large enough to avoid truncation by thinking of θ_1, θ_2 as cartesian coordinates x,y and then converting to polar coordinates ρ,ϕ with

$$\theta_1=\rho\cos\phi,\quad \theta_2=\rho\sin\phi. \tag{31}$$

Equation (27) becomes

$$I=\frac{Q^2}{r^2H^2}\int_0^{\infty}\exp\left(-\frac{\alpha\rho^2}{2H}r\right)\rho^3\,d\rho\times\int_0^{\pi/2}\sin\phi\cos\phi\,d\phi\times\mu r\Phi p. \tag{32}$$

The ρ integral is $2(H/\alpha r)^2$ and the angle integral is simply $\frac{1}{2}$, and the solution,

$$I=\frac{Q^2\mu}{\alpha^2 r^3}\Phi p \tag{33}$$

agrees with Eq. (28) for θ_c large, as assumed here. The Lommel-Seeliger equivalent is

$$I=\frac{Q^2}{r^2H^2}\int_0^{\infty}\exp\left(-\frac{\alpha\rho^2}{2H}r\right)\rho^2\,d\rho\int_0^{\pi/2}\frac{\sin\phi\cos\phi}{(\sin\phi+\cos\phi)}\times d\phi\,\mu r\Phi p. \tag{34}$$

The ρ integral is now $(2\pi)^{1/2}(H/\alpha r)^{3/2}$, and the angle integral is $1-2^{1/2}\ln(2^{1/2}+1)=0.3767\equiv\epsilon$. The result is

$$I=\frac{Q^2(2\pi)^{1/2}\epsilon\mu}{H^{1/2}\alpha^{3/2}r^{5/2}}\Phi p \tag{35}$$

which is, not surprisingly, intermediate between Eqs. (28) and (25) in the large θ_c limit.

B. Range-dependent reverberation

1. RD two-way propagation; point targets

Equation (14) or (15) is demonstrably reciprocal and therefore valid in both directions. Therefore for propagation from source to target or target back to monostatic receiver the result is

$$I=\sqrt{\frac{2\pi}{\alpha r^3 H_{\mathrm{eff}}}}\,\mathrm{erf}\left\{\sqrt{\frac{\alpha r H_{\mathrm{eff}}}{2}}\frac{\theta_c H_c}{H_t H_s}\right\} \tag{36}$$

with H_{eff} and H_c still defined by Eqs. (16) and (17) but with the water depth at the target H_t replacing that at the receiver H_r. So the intensity reduces to

$$I=\frac{2\pi}{\alpha r^3 H_{\mathrm{eff}}}\left(\mathrm{erf}\left\{\sqrt{\frac{\alpha r H_{\mathrm{eff}}}{2}}\frac{\theta_c H_c}{H_t H_s}\right\}\right)^2 S_T. \tag{37}$$

A complete solution including the single-mode effect [Eqs. (6) and (7)] is given by

$$I=(I_o)^2 S_T, \tag{38}$$

where

$$I_o=\sqrt{\frac{2\pi}{\alpha r^3 H_{\mathrm{eff}}}}\left(\mathrm{erf}\left\{\sqrt{\frac{\alpha r H_{\mathrm{eff}}}{2}}\frac{\theta_c H_c}{H_t H_s}\right\}-\mathrm{erf}\left\{\sqrt{\frac{\alpha r H_{\mathrm{eff}}}{2}}\frac{\theta'}{H_t}\right\}\right)+\frac{\lambda}{r H_t H_s}\exp\left(-\frac{\alpha H_{\mathrm{eff}}\theta_1^2 r}{2H_t^2}\right) \tag{39}$$

and $\theta_1=\lambda/2H_s$ and $\theta'=\mathrm{MIN}(1.5\times\theta_1, H_c\theta_c/H_s)$.

2. RD reverberation with constant scattering strength

By analogy with Eq. (25) the reverberation from an elementary area is

$$I=\frac{Q^2\pi}{2\alpha r^2 H_{\text{eff}}}\left(\text{erf}\left\{\sqrt{\frac{\alpha r H_{\text{eff}}}{2}}\frac{\theta_c H_c}{H_t H_s}\right\}\right)^2 S_B \Phi p \tag{40}$$

denoting depth at scatterer by H_t. Again a complete solution including the single-mode effect is

$$I=(Q/2)^2(I_o)^2 r S_B \Phi p. \tag{41}$$

3. RD reverberation with lambert's law scatterers

Equation (11) operating two-ways with a Lambert's law scatterer becomes

$$I=\left(\frac{Q}{rH_t}\int_0^{\theta_{\text{lim}}}\sin(\theta_t)\times\exp\{-\alpha H_s^2\theta_s^2 J(r)/2\}d\theta_s\right)^2\mu r\Phi p. \tag{42}$$

Substituting Eqs. (8) and (13) and replacing $\sin\theta$ by θ the soluble integral becomes

$$I=\left(\frac{2QH_s}{r^2\alpha H_{\text{eff}}}\int_0^{\theta_{\text{lim}}}\theta_s\exp\{-\theta_s^2\alpha r H_{\text{eff}}/(2H_t^2)\}d\theta_s\right)^2\mu r\Phi p$$

$$=\frac{1}{r^3}\left(\frac{QH_s}{\alpha H_{\text{eff}}}\{1-\exp(-\theta_c^2 H_c^2\alpha r H_{\text{eff}}/(2H_t^2 H_s^2))\}\right)^2\mu\Phi p. \tag{43}$$

The complete solution including the single-mode effect is

$$I=(Q/2)^2(I_1)^2 r\mu\Phi p, \tag{44}$$

where

$$I_1=\frac{2H_s}{\alpha r^2 H_{\text{eff}}}(\exp(-\theta'^2\alpha r H_{\text{eff}}/(2H_t^2))-\exp(-\theta_c^2 H_c^2\alpha r H_{\text{eff}}/(2H_t^2 H_s^2)))+\frac{\lambda}{rH_t^2}\theta_1\exp\left(-\frac{\alpha H_{\text{eff}}\theta_1^2 r}{2H_t^2}\right) \tag{45}$$

and again $\theta_1=\lambda/2H_s$ and $\theta'=\text{MIN}(1.5\times\theta_1, H_c\theta_c/H_s)$. One can convert this to an explicit function of range for a uniform slope by making the substitutions of Eq. (16) for H_t and H_{eff}. Notice that the comments on the range dependence of reverberation in Sec. IV A 3 apply equally well in a range-dependent environment.

Examples of upslope, flat bottom and downslope environments are illustrated in Fig. 2. In each case the solid lines are target echo and reverberation for the high frequency case using Eqs. (37) and (43), and dashed lines are for 100 Hz using Eqs. (38) and (44). The slope is 1 in 200 so in the upslope case the curve truncates at 20 km where the water depth reaches zero. The environmental parameters are the middle ones of Fig. 1, namely $\alpha_{\text{dB}}=2$ dB/rad and $\theta_c=28°$. A slightly different view of these cases to blow up the differences is shown in Fig. 3 where up, flat, and down are seen on one graph. Finally a gray scale geographic plot of reverberation is shown in Fig. 4 and target echo in Fig. 5 for a 1 in 500 slope. Zero depth is at the bottom of each picture.

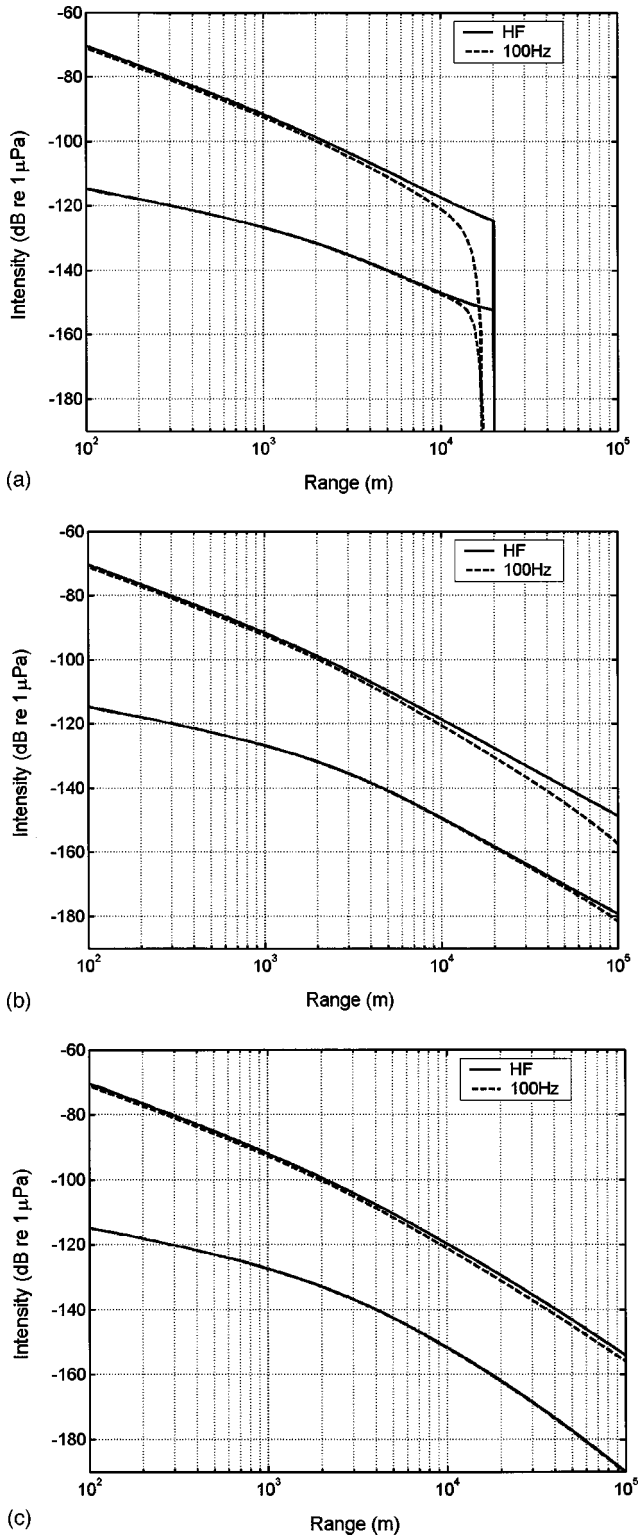


FIG. 2. Target echo (upper pair) and reverberation (lower pair). Solid line is the high frequency formula; dashed line includes single mode regime at 100 Hz. Source in 100 m of water (a) up-slope gradient 1:200, (b) flat, (c) down-slope 1:200.

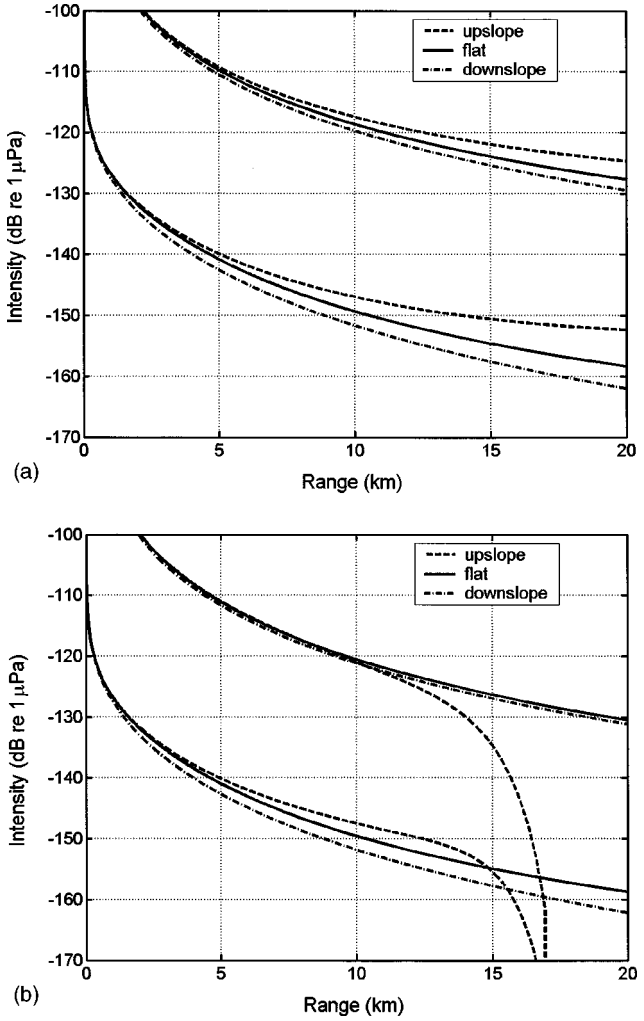


FIG. 3. Blow-up of Fig. 2 with up-slope, flat, down-slope superimposed: (a) HF limit, (b) including single mode regime at 100 Hz.

V. SIGNAL-TO-REVERBERATION RATIO

Taking formulas from the earlier sections one can now write expressions for the signal-to-reverberation ratio (SRR). These general formulas take account of range dependence, large or small critical angle, and high or low frequencies. Various intermediate approximations can be made, and three useful ones are made below: high frequency (ignoring the single mode regime); range independent; large critical angle. All cases consider target echo, reverberation with constant scattering strength and reverberation obeying Lambert's law. For clarity the formulas are grouped in tables starting with the simplest approximations. The later figures are for uniformly sloping bottoms, whereas the tables show formulas for arbitrary profiles using the formula Eq. (13) for H_{eff}. In all tables Q is set to $Q=1$ (down-going incident rays only at the seabed). The bottom left corner of Table II shows the simplest state of affairs where reverberation always dominates at long range because the scattering area is proportional to range. Also the SRR has no environmental dependence other than the scattering strength S_B; the other parameters are functions of the sonar and the target. In contrast the bottom right entry, for Lambert's law combined with mode stripping, tells a different story. First there is no range dependence. This is because the widening of the scattering area with range is exactly compensated by the narrowing of the Gaussian angle distribution with range. Therefore the signal-to-reverberation ratio is just a number, independent of range, and consequently *there is no such thing as reverberation limiting in this regime* since an intersection of target and reverberation levels is impossible. Second the signal-to-reverberation ratio is proportional to bottom reflection loss through α. Not only does the SRR depend on the propagation environment now, but *the SRR is improved by poor propagation*, which seems counterintuitive. Referring to the other entries in the table one can see that this is because the target echo is weakened by poor propagation (in proportion to α), as expected, but the reverberation is even more weakened (in proportion to α^2). For similar reasons the SRR is expected to be better in shallow water than in deep.

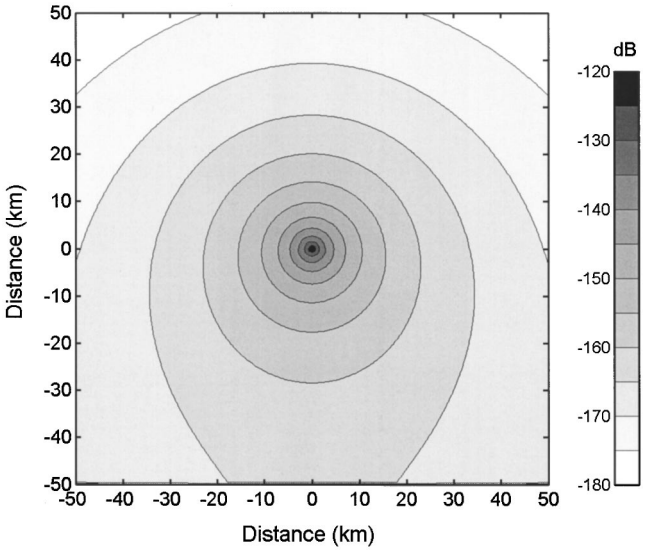


FIG. 4. Geographic projection of monostatic sonar reverberation (in dB *re* 1 μPa) on a slope. Source in 100 m of water at center with zero depth contour at the bottom of the figure.

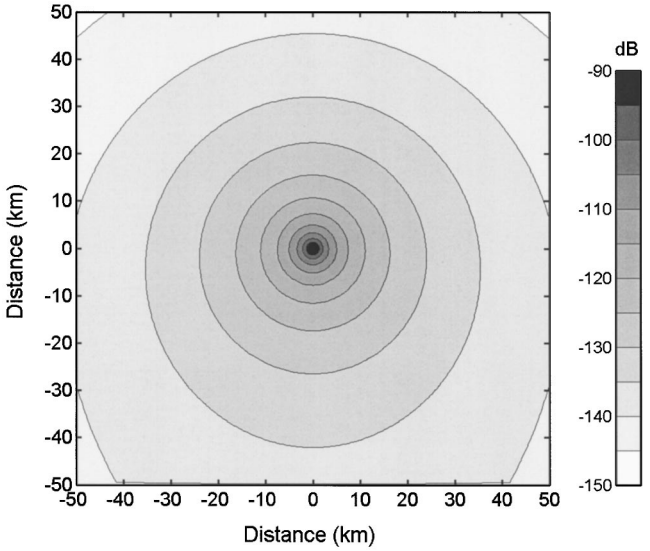


FIG. 5. Geographic projection of monostatic sonar target echo (in dB *re* 1 μPa) on a slope. Source in 100 m of water at center with zero depth contour at the bottom of the figure.

TABLE II. Formulas for target echo, reverberation, and their ratio SRR in the approximation of range-independence, large critical angle, and high frequency.

	$S=$ constant[a]	$S=$ Lambert[a]
Target	$S_T\frac{2\pi}{H\alpha r^3}$	
Reverberation	$\frac{\pi}{2H\alpha r^2}S_B\Phi p$	$\frac{1}{\alpha^2 r^3}\mu\Phi p$
SRR	$\frac{4S_T}{rS_B\Phi p}$	$\frac{2S_T\pi\alpha}{H\mu\Phi p}$

[a]$p=ct_p/2$.

Although the inclusion of refraction in the water column introduces more regimes[13] (e.g., differences between scattering from the high speed and the low speed side of the duct) the general behavior is similar to the isovelocity case and the above effects may still be found. In practice, with or without refraction, other mechanisms may become important in the long range condition. If there is no reverberation range limit there may be an ambient noise limit. As the Lambert's law reverberation from the assumed single boundary decays the reverberation from the opposite boundary may dominate, depending on its scattering law behavior. Similarly, from simple geometric considerations, volume reverberation from the water column (e.g., fish) could become important.

The finite critical angle introduces the exponential and the erf terms in Table III. Clearly for large W_{RI} there is no difference between these formulas and those in Table II since the Gaussian angle distribution is well inside the critical angle (the erf tends to unity and the exponential tends to zero). Small W_{RI} results in $\mathrm{erf}(W_{RI})=(2/\sqrt{\pi})W_{RI}$, and $(1-\exp(W_{RI}^2))=W_{RI}^2$, so the SRR, i.e., the ratio squared is $4/(\pi W_{RI}^2)$. This leads to $\mathrm{SRR}\propto 1/r$. Notice that the constant scattering strength term remains unchanged.

The forms of the formulas are unchanged by introducing range dependence (Table IV), one simply finds extra ratios of water depths. The critical angle transition in range is modified by W_{RD}. In the long-range limit of an up-slope environment the target echo and the constant scattering strength reverberation are both increased slightly by the $1/H_{\mathrm{eff}}$ term which is marginally greater than $1/H_s$. Thus their ratio is again unchanged. However the up-slope angle changes have a stronger effect on the Lambert's law reverberation, which is seen as the extra $(H_s/H_{\mathrm{eff}})^2$ term. The final SRR result comparing with the flat bottom case (depth H_s) differs only by a factor of H_s/H_{eff} though.

TABLE III. Formulas for target echo, reverberation, and their ratio SRR in the approximation of range independence, and high frequency.

	$S=$ constant[a]	$S=$ Lambert[a]
Target	$\frac{2\pi S_T}{\alpha H r^3}(\mathrm{erf}(W_{RI}))^2$	
Reverberation	$\frac{\pi S_B\Phi p}{2\alpha H r^2}(\mathrm{erf}(W_{RI}))^2$	$\frac{\mu\Phi p}{\alpha^2 r^3}(1-\exp(-W_{RI}^2))^2$
SRR	$\frac{4S_T}{rS_B\Phi p}$	$\frac{2S_T\pi\alpha}{H\mu\Phi p}\left(\frac{\mathrm{erf}(W_{RI})}{1-\exp(-W_{RI}^2)}\right)^2$

[a]$W_{RI}=\sqrt{\alpha r/2H}\,\theta_c$.

TABLE IV. Formulas for target echo, reverberation, and their ratio SRR including range dependence, and finite critical angle, but only high frequencies.

	$S=$ constant[a]	$S=$ Lambert[a]
Target	$\frac{2\pi S_T}{\alpha r^3 H_{\mathrm{eff}}}(\mathrm{erf}(W_{\mathrm{RD}}))^2$	
Reverberation	$\frac{\pi S_B\Phi p}{2\alpha r^2 H_{\mathrm{eff}}}(\mathrm{erf}(W_{\mathrm{RD}}))^2$	$\frac{\mu H_s^2\Phi p}{\alpha^2 r^3 H_{\mathrm{eff}}^2}(1-\exp(-W_{\mathrm{RD}}^2))^2$
SRR	$\frac{4S_T}{rS_B\Phi p}$	$\frac{2S_T H_{\mathrm{eff}}\pi\alpha}{H_s^2\mu\Phi p}\left(\frac{\mathrm{erf}(W_{\mathrm{RD}})}{1-\exp(-W_{\mathrm{RD}}^2)}\right)^2$

[a]$W_{\mathrm{RD}}=\sqrt{\alpha r H_{\mathrm{eff}}/2}\,\theta_c H_c/H_t H_s$.

Despite the complexity of the formulas in Table V one can see that the constant scattering strength SRR is again unchanged. For high frequencies with Lambert's law the I_o and I_1 terms revert to those of Table IV. At low frequencies only the last terms of I_o and I_1 remain (the single mode terms) leading to $I_o/I_1=H_t/(\theta_1 H_s)=2H_t/\lambda$. Thus the SRR for Lambert's law reduces to $(2H_t/\lambda)^2 4S_T/(r\mu\Phi p)$ which resembles that of the constant scattering strength scatterer rather than the Lambert's law scatterer in its range dependence and the absence of dependence on bottom loss.

Figures 6(a)–6(c) show a number of interesting features that can be deduced from the SRR formulas. Each plot shows the high frequency formula (solid line), the low frequency formula (100 Hz, dashed line), and the large critical angle approximation (dashed–dotted). Each picture corresponds to a target and reverberation plot shown in the last section. For short ranges the SRR always increases as $\mathrm{SRR}\propto 1/r$, as already noted. Following the solid or dashed line from the left, again one sees the transition at a few km caused mainly by the bend in the reverberation curves. As mentioned earlier, the ratio of this range to the water depth appears not to vary very much for most realistic bottom types. In the flat bottom case the SRR levels out at greater ranges. In the upslope case there is a downward trend in the SRR which is truncated by zero depth water. In the downslope case the weakening reverberation makes the SRR rise with range even at low frequencies.

At first sight the absolute magnitude of the SRR looks rather high [e.g., Fig. 6(b)], but the order of magnitude can be checked against the range-independent Lambert's law formula in Table II. In contrast the Table entry for an angle-independent scatterer with scattering strength $S_B=\mu$ and any propagation law (not just mode-stripping) has SRR $=10\log(4S_T/(r\mu\Phi p))=12.0$ dB at 10 km and only 5.0 dB at 50 km.

A geographic presentation of the SRR on a slope corresponding to Figs. 4 and 5 is shown in Fig. 7. Close to the source/receiver one can see that SRR is inversely proportional to range. Elsewhere the SRR contours clearly follow the depth contours because SRR is proportional to H_{eff} (see

TABLE V. General formulas for target echo, reverberation, and their ratio SRR (range-dependent, finite critical angle, and any frequency).

	$S=$constant[a]	$S=$Lambert[a]
Target	$I=(I_o)^2 S_T$	
Reverberation	$I=(I_o)^2 r S_B \Phi p/4$	$I=(I_1)^2 r \mu \Phi p/4$
SRR	$\frac{4S_T}{rS_B \Phi p}$	$\frac{4(I_o)^2 S_T}{(I_1)^2 r \mu \Phi p}$

[a] $I_o=\sqrt{2\pi/\alpha r^3 H_{\text{eff}}}(\text{erf}(W_c)-\text{erf}(W_2))+(\lambda/rH_tH_s)\exp(-W_1^2)$,
$I_1=(2H_s/\alpha r^2 H_{\text{eff}})\times(\exp(-W_2^2)-\exp(-W_c^2))+(\lambda/rH_t^2)\theta_1\exp(-W_1^2)$,
$W_1=\sqrt{\alpha r H_{\text{eff}}/2}(\theta_1/H_t)$,
$W_2=\sqrt{\alpha r H_{\text{eff}}/2}(\theta_2/H_t)$,
$W_c=\sqrt{\alpha r H_{\text{eff}}/2}(H_c\theta_c/H_sH_t)$,
$\theta_1=\lambda/2H_s$,
$\theta_2=\text{MIN}(1.5\times\theta_1, H_c\theta_c/H_s)$.

Table IV) which for a tilted plane is only a function of depths at source and scatterer.

VI. TIME SMEARING

A. One-way-propagation

In all the earlier range-independent formulas it has been angle dependence that determines the behavior, and angle was treated as a continuum. Because delay time is related to angle the angle dependence can be translated into time dependence simply by a change of variable. In isovelocity the relation for total travel time t is

$$ct=r\sec\theta\cong r+r\theta^2/2. \tag{46}$$

If instead time τ is measured from the first arrival (when $t=t_o\equiv r/c$), then

$$\tau=t-r/c=r\theta^2/2c \tag{47}$$

and differentiating

$$d\tau=(r/c)\theta\, d\theta. \tag{48}$$

So a spread in angle of the form $\exp(-a\theta^2)$ becomes a spread in time of the form

$$p\, d\tau=\exp(-a\theta^2)d\theta=\sqrt{c/2r}\frac{\exp\{-a(2c/r)\tau\}}{\sqrt{\tau}}d\tau. \tag{49}$$

This equation says that for a one-way path, a point source and a point receiver, one should expect a (power) impulse response of the form $\exp(-b\tau)/\sqrt{\tau}$. Note that the assumed continuum of angles obscures any discrete eigenray arrivals.

The reason that this pulse shape is of interest is that the coefficient b is a function of the environment. In the range-independent case [Eq. (3)]

$$b=a(2c/r)=(\alpha r/2H)(2c/r)=\alpha c/H=\alpha/t_H, \tag{50}$$

where α is related to the reflection coefficient R through $R=\exp(-\alpha\theta)$ and t_H is the time taken to traverse the water depth once. In other words the pulse has a characteristic time width t_H/α, *which not only depends on the environment but is constant for all ranges*. Now Eq. (49) becomes

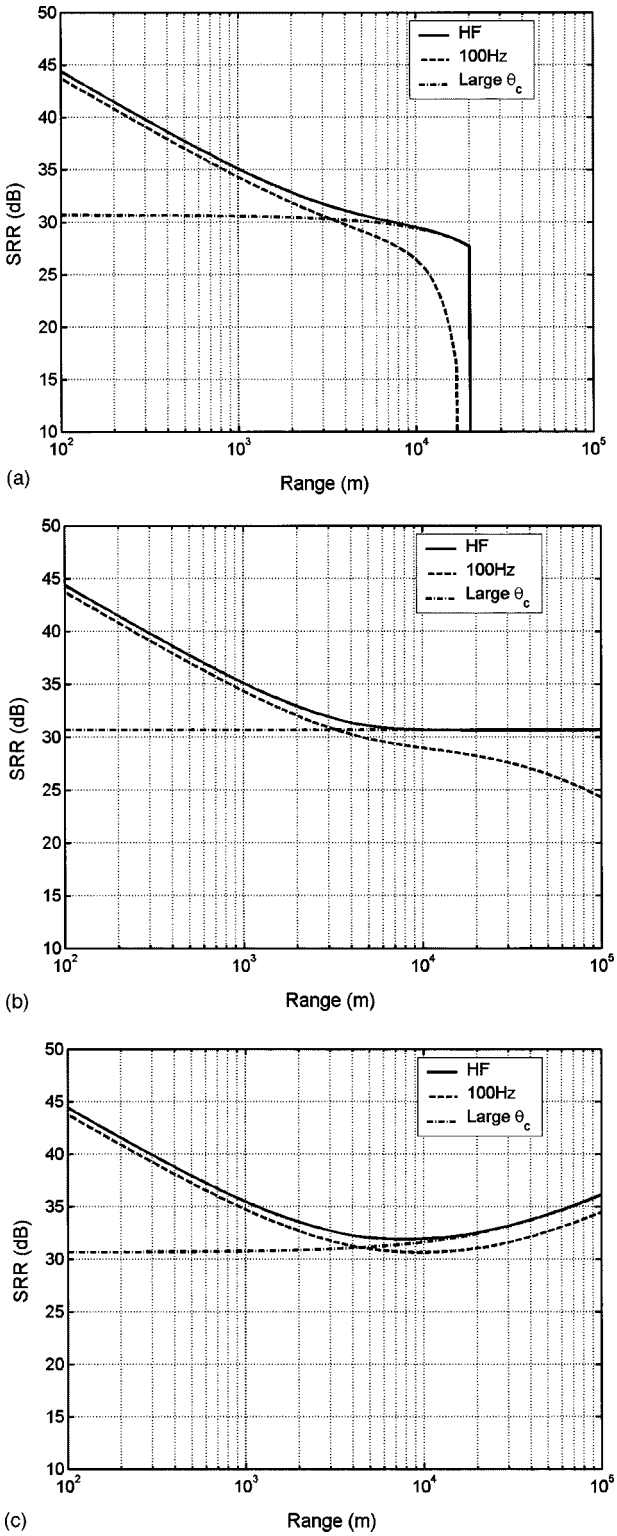


FIG. 6. Signal-to-reverberation-ratio curves corresponding to Fig. 2 for (a) up-slope, (b) flat, (c) down-slope. Solid line is the high frequency formula; dashed line includes single mode regime at 100 Hz. The dotted–dashed line is the high critical angle limit. Note the rising SRR with range in the down-slope case.

$$p\, d\tau=(2t_o)^{-1/2}\frac{\exp\{-\alpha\tau/t_H\}}{\sqrt{\tau}}d\tau, \tag{51}$$

where $t_o\equiv r/c$, the delay time to the first arrival.

The first implication of this is that generally the pulse

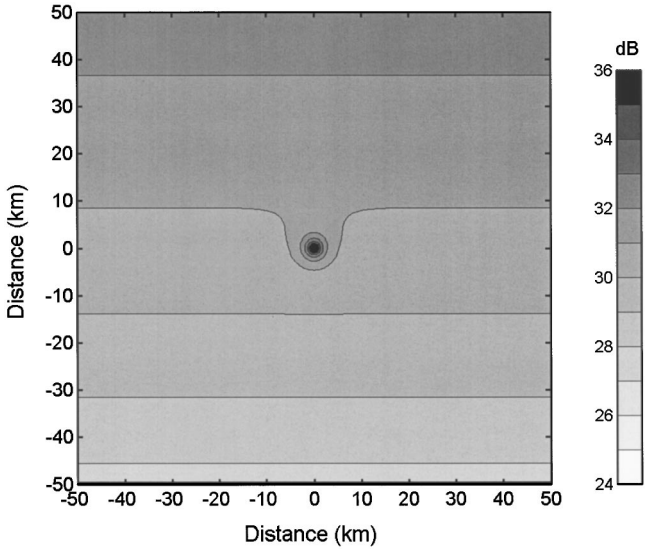


FIG. 7. Geographic projection of monostatic sonar signal-to-reverberation ratio on a slope, corresponding to Figs. 4 and 5.

shape is a function of the environment, and so *there is potential for determining the environment from the pulse shape*. The second implication is that this could form the basis for the choice of integration time for active sonar systems in multipath environments.

B. Two-way propagation (point target)

For a two-way path via a point scatterer the spread in time is

$$p\,d\tau_1\,d\tau_2 = \frac{c}{2r}\,\frac{\exp\{-a(2c/r)\tau_1\}}{\sqrt{\tau_1}}\,\frac{\exp\{-a(2c/r)\tau_2\}}{\sqrt{\tau_2}}\,d\tau_1\,d\tau_2 \tag{52}$$

but in order to collect the arrivals at time $T=\tau_1+\tau_2$, one changes variables to T, τ_1,

$$p\,dT = c/2r\int_{\tau_1=0}^{\tau_1=T}\frac{\exp\{-a(2c/r)T\}}{\sqrt{\tau_1(T-\tau_1)}}\,d\tau_1\,dT. \tag{53}$$

This can be recognized as the convolution of the one-way response with itself. By making the substitutions $\tau'=\tau_1-T/2$ followed by $\sin x=2\tau'/T$ one finds

$$p\,dT = \pi c/2r\exp\{-a(2c/r)T\}dT = \pi/2t_o\exp\{-\alpha T/t_H\}dT. \tag{54}$$

Thus the returning pulse shape for, say, a false target, is exponential in time. Or since $\exp(-\alpha)=10^{-\alpha_{\mathrm{dB}}/10}$ the pulse arrival rolls off at α_{dB} dBs per t_H seconds (where $t_H=H/c$). This means that the tails of targets or false targets or any spikelike response that results in multipaths could provide bottom reflection information. Remember that α, the initial slope of the reflection loss curve, lumps almost all the relevant geoacoustic information into a single number. One could argue that the critical angle has similar status, particularly at short range where it limits the number of eigenrays. In this case the critical angle will truncate the pulse and the delay where this occurs can be measured just as easily given adequate signal-to-noise.

In principle, a single shot in a locality with many false targets could provide bottom reflection properties over a wide area independently of scattering properties. Note that although this analysis is only for isovelocity environments the main effect of refraction, for instance in a surface or bottom duct, will be seen near the pulse's leading edge. In shallow water as soon as rays are steep enough to hit both boundaries the exponential tail is as inevitable as the Gaussian angle distribution. Thus refraction effects are expected to be over by a time $t_o(c_{\max}-c_{\min})/c_{\mathrm{av}}$ after the first arrival, where $c_{\max}$, $c_{\min}$, and c_{av} are maximum, minimum, and average quantities for the water column sound speeds. Furthermore the derivation of the pulse length $(\alpha c/H)^{-1}$ in Eq. (50) can be extended to range-dependent environments. The gaussian in Eq. (49) takes the form of Eq. (9), and the relation between travel time and angle can be found by integrating the phase along the ray $\omega\,dt=k\,dr\cos\theta+k\,dz\sin\theta$. The result with initial angle θ_s is[30]

$$c\tau=(H_s^2\theta_s^2/2)\int_0^r dr/H^2 \tag{55}$$

which replaces Eq. (47). Substituting for θ_s the time dependence (for, e.g., a two-way path) is still $\exp(-b\tau)$ (with pulse length b^{-1}) but

$$b=\alpha c\int_0^r dr/H^3\Big/\int_0^r dr/H^2=\alpha c/H_a. \tag{56}$$

For linear bathymetry $H_a=2H_rH_s/(H_r+H_s)$ which is an intermediate depth close to the arithmetic and geometric mean.

C. Lambert's law reverberation

Imagining the rather unlikely eventuality of a localized Lambert's law scatterer, the extra $\sin\theta_1\sin\theta_2$ term [from Eq. (27)] turns Eq. (52) into

$$p\,d\tau_1\,d\tau_2=(c/r)^2\exp\{-a(2c/r)\tau_1\}\times\exp\{-a(2c/r)\tau_2\}d\tau_1\,d\tau_2. \tag{57}$$

Again collecting arrivals at time $T=\tau_1+\tau_2$ and by integrating over τ_1 one finds a Rayleigh-type shape to the pulse

$$p\,dT=(c/r)^2T\exp\{-a(2c/r)T\}dT = (1/t_o)^2T\exp\{-\alpha T/t_H\}dT. \tag{58}$$

In practice it is unlikely ever to see this behavior unless there is a sudden change in bottom properties.

VII. CONCLUSIONS

This paper has demonstrated that one can still get a long way with closed-form expressions. In particular one can deal with propagation and reverberation at high and low frequencies and in isovelocity range-dependent environments with various scattering laws. These formulas are not intended to be a substitute for more computer-intensive numerical techniques, but it should be said that, in the context of broad band active sonar, the interference effects familiar from

single frequency wave solutions are, in any case, smoothed out as these formulas would suggest. Trends are therefore well explained. Indeed one could argue that operational research studies and trials planning hardly need more sophistication.

In Sec. II the propagation formulas for mode stripping and single remaining mode were reviewed and developed. After briefly investigating the likely variations in scattering law in Sec. III the propagation formulas were extended in Sec. IV to calculate target echo and reverberation under all the earlier propagation conditions including Lambert, Lommel-Seeliger, and angle-independent scattering laws.

Signal-to-reverberation ratio was calculated and tabulated in Sec. V. Graphs of SRR show three dominant regimes. At ranges less than a few km SRR tends to decrease with range largely because of the usual increase of scattering area and hence reverberation with range. At intermediate ranges there is a regime where *SRR is independent of range*, in other words, in which there can be no reverberation limiting. Furthermore *SRR is improved by high bottom loss*. Beyond this, depending on frequency, there is a regime where only one mode remains and so target echo and reverberation decay at least exponentially, however target always decays a bit more than reverberation so again SRR decreases with range. There are a few more surprises with range-dependent SRR. Going up-slope, propagation is more lossy because of the steeper ray angles, but this is counteracted by the concentration of flux in the shallower water. The result is that one-way propagation rises only weakly in proportion to $\sqrt{(H_s/H_{\text{eff}})}$ where H_{eff} is the ''effective depth,'' intermediate between source and scatterer depth (in fact, it is the average in the case of a uniform slope). Consequently two-way target echo is proportional to H_s/H_{eff}, but reverberation goes as $(H_s/H_{\text{eff}})^2$. This is only a modest increase compared with the flat bottom case. For instance, on a uniform slope halving the depth ($H_{\text{eff}}=\frac{3}{4}H_s$) results in only a 2.5 dB increase in reverberation, and doubling it ($H_{\text{eff}}=\frac{3}{2}H_s$) results in a 3.5 dB decrease. Going downslope the SRR actually rises as range increases so that there is a benefit in looking towards deep water. All the range-dependent formulas in this paper are for general bathymetric profiles although the illustrations are for uniform slope.

In all these cases bottom losses control the angle distribution, and in Sec. VI this distribution was converted into a distribution in time or a pulse shape (envelope) for a point scatterer. Interestingly this spread pulse has a constant length irrespective of range, and the shape is exponential in time for a two-way angle-independent path. This finding was shown to be true in a range-dependent, refracting environment. Calculations are also possible for Lambert's law but the pulse shape can, of course, never be seen in the reverberation because of the scatterer's finite extent. The constant pulse length has two implications. One is that it depends on the reflection (and not scattering) properties of the environment in a simple way, and therefore could, in principle, be used as a way of measuring bottom reflection properties. The other is that a known and constant time spreading is a bonus for active sonar processing designers.

ACKNOWLEDGMENTS

Thanks are due to Karna Bryan who stimulated the author's thoughts on the subject by puzzling over parallel target and reverberation curves from a well-known model. Further thoughts arose as a result of conversations with John Preston, Dale Ellis, and Kevin LePage during the BOUNDARY2001 experiment. Another driving force was to produce some (smooth) benchmark solutions against which Mark Prior and Kevin LePage could test their respective models, SUPREMO and R-SNAP.

[1] N. C. Makris, L. Z. Avelino, and R. Menis, ''Deterministic reverberation from ocean ridges,'' J. Acoust. Soc. Am. **97**, 3547–3574 (1995).
[2] H. Weinberg, ''Generic sonar model,'' Naval Underwater Systems Center, Newport, RI, NUSC TD 5971D, 1985.
[3] A. Baldacci and C. H. Harrison, ''SUPREMO Version 1.0 User's Guide,'' SACLANT Undersea Research Centre, La Spezia, Italy, Report M-141, 2002.
[4] C. H. Harrison, ''A simple relationship between frequency and range averages for broadband sonar,'' J. Acoust. Soc. Am. **97**, 1314–1317 (1995).
[5] K. LePage, ''Monostatic reverberation in range dependent waveguides: the R-SNAP model,'' SACLANT Undersea Research Centre, La Spezia, Italy, Report SR-363, 2002.
[6] D. E. Weston, ''Propagation in water with uniform sound velocity but variable-depth lossy bottom,'' J. Sound Vib. **47**, 473–483 (1976).
[7] D. E. Weston, ''Acoustic flux formulas for range-dependent ocean ducts,'' J. Acoust. Soc. Am. **68**, 269–281 (1980).
[8] D. E. Weston, ''Acoustic flux methods for oceanic guided waves,'' J. Acoust. Soc. Am. **68**, 287–296 (1980).
[9] D. E. Weston, ''Intensity-range relations in oceanographic acoustics,'' J. Sound Vib. **18**, 271–287 (1971).
[10] C. H. Harrison, ''Signal and reverberation with mode-stripping and Lambert's law,'' SACLANT Undersea Research Centre, La Spezia, Italy, Report SR-356, 2002.
[11] L. Brekhovskikh and Yu. Lysanov, *Fundamentals of Ocean Acoustics* (Springer-Verlag, New York, 1982), pp. 132, 141, and 142.
[12] C. H. Harrison, ''Noise directionality for surface sources in range-dependent environments,'' J. Acoust. Soc. Am. **102**, 2655–2662 (1997).
[13] C. H. Harrison, ''Signal and reverberation formulae including refraction,'' SACLANT Undersea Research Centre, La Spezia, Italy, Report SR-370, 2002.
[14] C. H. Harrison, ''Fast bistatic signal-to-reverberation-ratio calculation,'' J. Acoust. Soc. Am. (submitted).
[15] D. E. Weston, ''Guided propagation in a slowly varying medium,'' Proc. Phys. Soc. London **73**, 365–384 (1959).
[16] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, New York, 1994), pp. 320–323.
[17] D. D. Ellis and P. Gerstoft, ''Using inversion techniques to extract bottom scattering strengths and sound speeds from shallow-water reverberation data,'' Proceedings of 3rd European Conference on Underwater Acoustics, Heraklion, Greece, June, 1996.
[18] J. L. Bishop, M. T. Sundvik, and D. W. Grande, ''Inverting seabed acoustic parameters from reverberation data,'' *Full Field Inversion Methods in Ocean and Seismo-Acoustics* (Kluwer, Dordrecht, The Netherlands, 1995).
[19] C. W. Holland, ''Shallow water coupled scattering and reflection measurements,'' SACLANT Undersea Research Centre, La Spezia, Italy, Report SR-344, 2001.
[20] D. D. Ellis, J. R. Preston, R. Hollett, and J. Sellschopp, ''Analysis of towed array reverberation data from 160 to 4000 Hz during rapid response 97,'' SACLANT Undersea Research Centre, La Spezia, Italy, Report SR-280, 2000.
[21] J. R. Preston and D. D. Ellis, ''Summary of bottom reverberation findings and model/data comparisons during three rapid environmental assessment trials,'' J. Acoust. Soc. Am. **105**, 1040 (1999).
[22] D. D. Ellis and D. V. Crowe, ''Bistatic reverberation calculations using a three-dimensional scattering function,'' J. Acoust. Soc. Am. **89**, 2207–2214 (1991).
[23] H. C. Van de Hulst, *Multiple Light Scattering* (Academic, New York, 1980).
[24] M. Kaasalainen, K. Muinonen, and T. Laakso, ''Shapes and scattering

properties of large irregular bodies from photometric data,'' Opt. Express **8**, 296–301 (2001).
[25] C. W. Holland, R. Hollett, and L. Troiano, ''Measurement technique for bottom scattering in shallow water,'' J. Acoust. Soc. Am. **108**, 997–1011 (2000).
[26] R. J. Greaves and R. A. Stephen, ''Low-grazing-angle monostatic acoustic reverberation from rough and heterogeneous seafloors,'' J. Acoust. Soc. Am. **108**, 1013–1025 (2000).
[27] R. J. Urick, *Principles of Underwater Sound* (McGraw-Hill, New York, 1975).
[28] E. L. Hamilton, ''Acoustic properties of sediments,'' in *Acoustics and Ocean Bottom*, edited by A. Lara-Saenz, C. Ranz-Guerra, and C. Carbo-Fite, Proceedings of F.A.S.E. Conference, Madrid, June, 1987.
[29] E. L. Hamilton, ''Geoacoustic modeling of the sea floor,'' J. Acoust. Soc. Am. **68**, 1313–1340 (1980).
[30] C. H. Harrison and M. Siderius, ''Effective parameters for matched field geoacoustic inversion in range-dependent environments,'' IEEE J. Oceanic Engineering, Vol. 28, No. 3, July (2003).

Acoustic backscattering by Hawaiian lutjanid snappers. I. Target strength and swimbladder characteristics

Kelly J. Benoit-Bird and Whitlow W. L. Au[a)]
Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

Christopher D. Kelley
Hawaii Undersea Research Laboratory, University of Hawaii, Honolulu, Hawaii 96822

(Received 13 September 2002; accepted for publication 11 August 2003)

The target strengths and swimbladder morphology of six snapper species were investigated using broadband sonar, x rays, and swimbladder casts. Backscatter data were obtained using a frequency-modulated sweep (60–200 kHz) and a broadband, dolphinlike click (peak frequency 120 kHz) from live fish, mounted and rotated around each of their three axes. X rays revealed species-specific differences in the shape, size, and orientation of the swimbladders. The angle between the fish's dorsal aspect and the major axis of its swimbladder ranged from 3° to 12° and was consistent between individuals within a species. This angle had a one-to-one relationship with the angle at which the maximum dorsal aspect target strength was measured ($r^2=0.93$), regardless of species. Maximum dorsal aspect target strength was correlated with length within species. However, the swimbladder modeled as an air-filled prolate spheroid with axes measured from the x rays of the swimbladder predicted maximum target strength significantly better than models based on fish length or swimbladder volume. For both the dorsal and lateral aspects, the prolate spheroid model's predictions were not significantly different from the measured target strengths (observed power >0.75) and were within 3 dB of the measured values. This model predicts the target strengths of all species equally well, unlike those based on length. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1614256]

PACS numbers: 43.30.Sf [WMC] Pages: 2757–2766

I. INTRODUCTION

The Hawaiian bottomfish fishery is comprised of 12 species, the most important of which are seven snappers in the family Lutjanidae that dwell near the bottom in deepwaters (100–400 m). Although all of these species are federally managed, two species, the onaga or long-tailed red snapper (*Etelis coruscans*) and the ehu or red snapper (*Etelis carbunculus*), have become depleted in the main Hawaiian Islands. A third species, the opakapaka or pink snapper (*Pristipomoides filamentosus*), is considered critical to the fishery (Western Pacific Regional Fisheries Management Council, 1999). These three species, the most commercially important bottomfish, are the primary species of concern in this study. Preliminary management efforts for these species have been made; however, the potential effectiveness is undetermined because very little is known about these deepwater fish and their habitat.

The use of various types of acoustic instrumentation such as side-scan sonar, split beam sonar, multi-beam sonar, and doppler current profile sounders could potentially be used to address the problems of monitoring these deepwater fish. However, all of these acoustic technologies share some common and unique disadvantages, one of which is the need for detailed acoustic backscattering data for targeted species (MacLennan, 1990). Field application of acoustic methods to estimate animal abundance requires information on the acoustic size, target strength or backscattering cross section of individual organisms (MacLennan, 1990; Thiebaux *et al.*, 1991). No dorsal aspect measurements of acoustic scattering strength are available for deepwater snappers.

Swimbladders have been identified as the primary cause of acoustic backscattering in several species (Clay and Horne, 1994; Foote, 1980), accounting for as much as 90%–95% of echo energy. Some studies have found that the scattering field for the entire fish can be reconstructed mainly from the properties of the swimbladder (Clay and Horne, 1994; Foote, 1980; Foote and Ona, 1985). However, no information on the size, shape, or other characteristics of the swimbladders of Lutjanid snappers is available.

The objectives of Part I of this work were to obtain target strength and physiological data that could contribute to the acoustic assessment of Hawaiian deepwater snappers in the field. Measurements of backscatter strength taken *in situ* from a manned submersible show that the shape and size of the swimbladder is roughly conserved in these fish (Benoit-Bird *et al.*, 2003). However, these backscatter measurements are only from the lateral aspect of the fish. The first objective was to obtain controlled measurements of acoustic backscatter strength from all aspects of these fish, particularly the two depleted species and the pink snapper, which is the most commercially valuable species. The second was to measure the shape, size, and orientation of the swimbladders of the same fish. Objective three was to assess how acoustic backscattering strength was related to the swimbladder measures and other physiological characters such as length, biovolume, and wet weight.

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

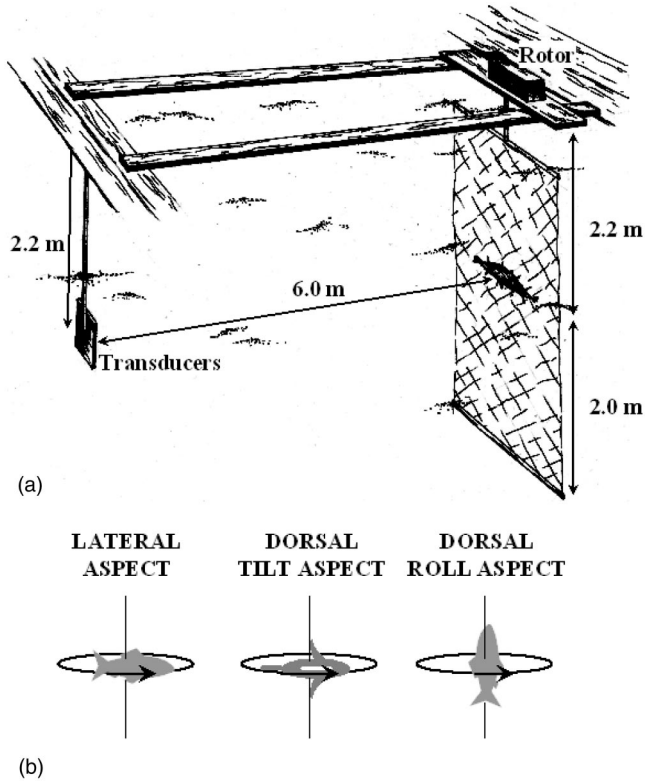


FIG. 1. (a) Experimental setup showing the position of the fish tied into a net bag and mounted to a larger, weighted net turned by a rotor. Both the transmit and receive transducer were located 6 m from the fish, 2.2 m deep, the same depth as the fish. (b) Orientations of target fish as they were rotated about their axes.

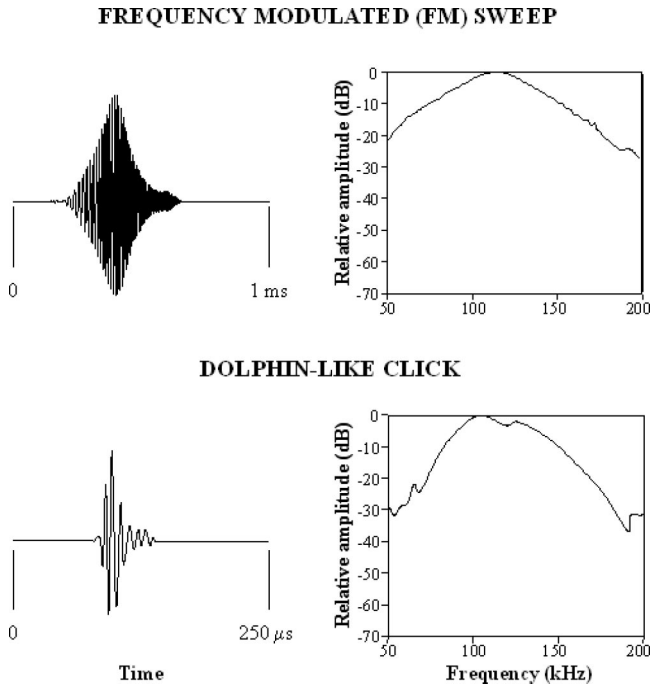


FIG. 2. Waveforms and spectrograms of the incident sonar signals.

II. METHODS

A. Backscatter

Specimens of five of the seven Hawaiian deepwater snapper species and one introduced shallow water snapper were captured off the coasts of the Hawaiian Islands using standard, bottomfishing techniques. The fish were kept alive by immediately deflating their over-expanded swimbladders with a hypodermic needle, releasing the pressure caused by the rapid change in depth. Fish were then transported to the Hawaii Institute of Marine Biology's bottomfish hatchery on Oahu. There, they were maintained in tanks or net pens for a minimum of 8 days to allow them to acclimate to ambient conditions and heal their swimbladders before their backscattering properties were measured. A live, individual fish that had been starved for one day was then transferred into a bath containing 1 mL of 2-phenoxy-ethanol per 10 L of seawater. Once anesthetized, the fish was enclosed in a fitted monofilament net bag to restrain its movements. The net bag was mounted to a large, weighted, monofilament net that could be rotated 360° by a rotor [Fig. 1(a)]. The fish was sequentially mounted in three orientations, the order of which was randomized, for rotation about each of its three axes [dorsal tilt, dorsal roll, and lateral, Fig. 1(b)]. Ten to 11 specimens each of the three primary species of bottomfish were acoustically examined while their stomachs were empty.

A bi-static system was used to measure the echoes from the fish. Planar circular transducers were used as the projector and receiving hydrophone. The transducers were located side by side with a center-to-center separation distance of 10.5 cm. The transmit and receive transducers were both set up 2.2 m deep, the same depth as the mounted fish, approximately 6 m from the fish. Two signals were used, a linear, frequency-modulated sweep with a frequency range of 60 to 200 kHz and a broadband, dolphinlike click with a peak frequency of 120 kHz and a 60-kHz bandwidth, as shown in Fig. 2. The overall 3-dB beamwidth of the transducer assembly at the peak frequency of the signals was 12°. The outgoing signals were produced using a function generator computer plug-in board. The function generator also produced a trigger signal for each transmission. After a delay related to the two-way travel time from the signal to the target, a delayed trigger prompted a Rapid System R1200 analog-to-digital (A/D) converter to digitize and store a block of 1024 sample. Sampling rates of 1 MHz were used for the function generator and the A/D converter. The delayed trigger also caused the rotor and net to advance by an incremental angle. Echoes were collected in 1.5°–2.5° increments about each of the fish's three axes for both source signals.

The incident signals were first measured and digitized with the receiving hydrophone located at the position of a target fish, directly facing the projecting transducer. Target strength based on the signal amplitudes as a function of frequency was calculated by comparing the reflected signal to the incident signal using the equation

$$\mathrm{TS}(f) = 20\log\left[\frac{|V_e(f)|}{|V_i(f)|}\right] + 20\log(R), \tag{1}$$

where $|V_e(f)|$ is the absolute value of the Fourier transform of the echo voltage after correcting for gain, $|V_i(f)|$ is the absolute value of the Fourier transform of the incident volt-

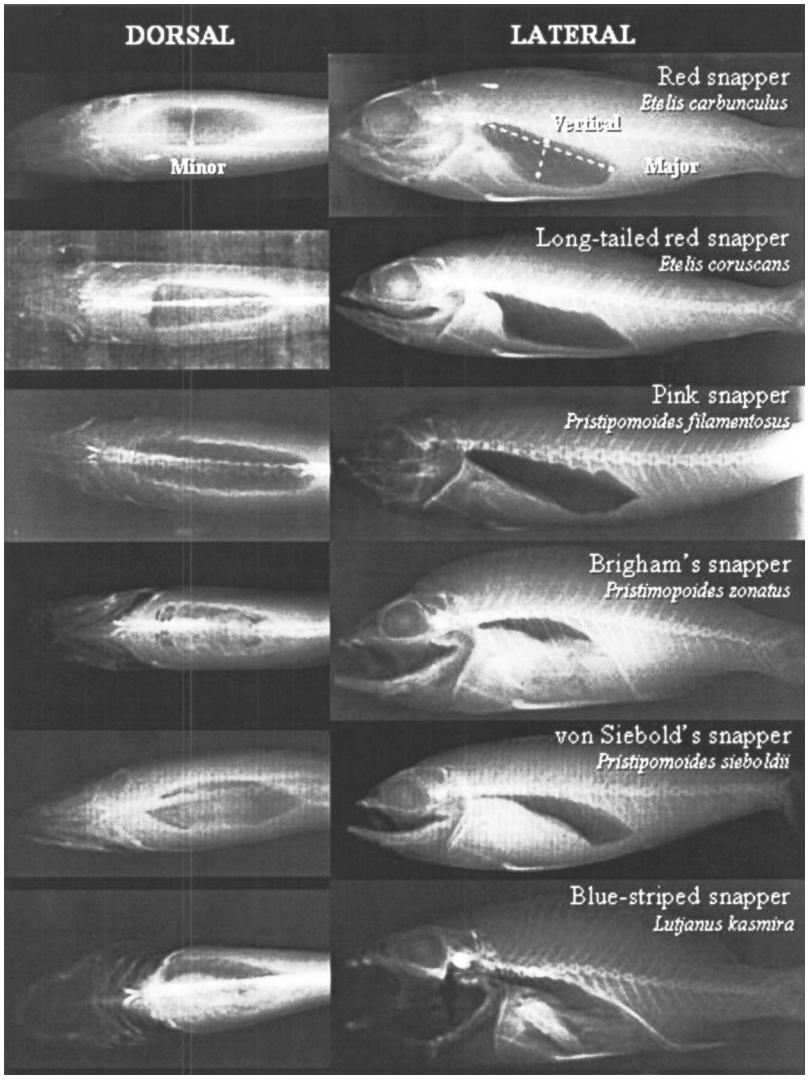


FIG. 3. Dorsal and lateral x rays of five species of Hawaiian, lutjanid snappers. The swimbladders of each fish are evident as the dark areas behind the eye and below the spine of each fish. Fish are all scaled to the same length to permit comparison of interspecies differences in swimbladder size. The axes of the swimbladder of each fish were measured from the x rays, as shown in the top panel.

age, and R is the distance between the transducers and the target. Target strength based on the energy in the incident and echo signals were also calculated using the equation

$$\mathrm{TS}_E(f) = 10 \log[E_e/E_i] + 20 \log R, \tag{2}$$

where E is the value of the integral with respect to time of the respective instantaneous voltage squared. Target strength based on the absolute value of the voltage in the frequency domain was calculated at frequencies of 60, 100, 150, and 200 kHz using Eq. (1) and on energy using Eq. (2). Echoes from the empty net apparatus were also measured at different orientations. The net apparatus echoes interfered with the fish echoes only within $\pm 15°$ of the head and tail of the fish. Echoes from these orientations were removed from the analyses.

After acoustic measurement, fish were sacrificed by over-anesthetization using 2 mL of 2-phenoxy-ethanol per 10 L of seawater. The standard length, total length, displacement volume, and wet weight of each fish were measured after which specimens were immediately frozen.

B. Swimbladders

Frozen fish were taken to Queen's Medical Center where they were x-rayed from both their dorsal and lateral aspects. The lengths of each of the axes of the swimbladder and the angle between the dorsal aspect of the fish and the major axis of the swimbladder were measured directly from the x rays (Fig. 3). X rays also revealed swimbladders that were damaged during decompression. In these fish, air was evident in other cavities within the body and the bottom edge of the swimbladder was indistinct. These fish were not included in the acoustic or swimbladder analyses.

After x-raying, fish were partially thawed and a small slit was made on their ventral side to expose the ventral-most portion of the swimbladder. A syringe with an 18-gauge needle was used to inject Plaster of Paris (2 parts plaster to 1 part water) into the swimbladders following the general technique of Do and Surti (1990). A small amount of food coloring was mixed into the plaster to allow it to be seen through the translucent swimbladder wall, making it easier to determine when the swimbladders were full. When injected, the fish were partly frozen which maintained the structures around the swimbladders, preserving their shape as much as possible. The fish were refrigerated for 24 h postinjection after which the hardened swimbladder casts were extracted. Each cast was then sealed with a spray varnish and its displacement volume measured.

C. Data analysis

The correlation between the target strengths measured using the two signal types was tested using a linear correlation. The relationships between the log of fish standard length and maximum target strength from both the dorsal and lateral aspects were assessed using a linear regression for the frequency-modulated sweep's total energy and four discrete frequencies: 60, 100, 150, and 200 kHz. The relationship between the angle at which the maximum dorsal aspect target strength was measured and the angle of the swimbladder relative to the dorsal aspect of the fish was tested with a linear regression. *F*-tests were used to test the significance of each regression's slope.

The measured maximum dorsal and lateral aspect target strengths were compared to the predictions from various simple models. First, the maximum measured target strength was compared to the target strength of the fish predicted by modeling the swimbladder as a prolate spheroid (Furusawa, 1988) using the equation

$$\mathrm{TS}_{\text{prolate spheroid}} = 10 \log[(bc/2a)^2], \tag{3}$$

where *a*, *b*, and *c* are the axes of the swimbladder measured from the x rays and the source is parallel to axis-*a* (Urick, 1983). The target strength derivation of Furusawa (1988) for the prolate spheroid was used instead of Stanton's (1989), mainly because of its simplicity. The second model predicts target strength of the fish by modeling the swimbladder as a sphere with a volume equivalent to that measured from the plaster cast of the swimbladder using the equation

$$TS_{\text{sphere}} = 10 \log(r^2/4), \tag{4}$$

where *r* is the radius of the sphere (Urick, 1983). The equation for backscatter from a gas-filled sphere is considerably more complicated than Eq. (4), yet for $ka > 1$, the results are similar (Stanton, 1989). While the spherical model is a gross oversimplification of a fish, its use by commercial echosounders and in inverse techniques merits its comparison with other models. The final models for dorsal-aspect target strength are based on fish length using the equation Love (1970) developed for many types of fish and Foote's (1980) equation for swimbladder-bearing fish. Although many models of dorsal-aspect target strength of fish have been developed, those of Love (1970) and Foote (1980) are very consistent with these other models (McClatchie *et al.*, 1996). The differences between the measured and predicted target strength for each model were compared using paired *t*-tests, corrected for multiple comparisons using the Bonferroni method.

III. RESULTS

A. Swimbladders

The shape of the swimbladder of each species is unique and conserved between the various individual specimens of the primary species over a range of sizes (Fig. 3). Comparison of x rays to similar views of plaster casts of the swimbladders reveals that the shape of the swimbladder is conserved by the casts. Axes measurements of the swimbladders taken with vernier calipers were within 3 mm of those taken directly from x rays.

From the lateral view, the pink snapper's swimbladder is shaped like an unbalanced ovoid with the posterior end larger than the anterior. The wider, posterior, end of the swimbladder reaches a small peak near the center of the rounded portion. From the dorsal view, the pink snapper's swimbladder is a nearly perfect ovoid, symmetrical from side to side and from front to back.

The dorsal side of the swimbladder of the red snapper is similar to that of the pink snapper, while its ventral side was much fuller after the anterior quarter. The red snapper's swimbladder, like the pink snapper's, ends in a small peak at the center of the rounded portion. From the dorsal view, the red snapper's swimbladder was nearly ovoid in shape, wider at the posterior than anterior end. Both the anterior and posterior ends of the red snapper's swimbladder taper rapidly to a point.

The long-tailed red snapper has a much more linear swimbladder than either the red snapper or the pink snapper, shaped like a rough parallelogram from the side, similar to a saddle. From the dorsal view, the red snapper's swimbladder is nearly triangular in shape, wide and flat at the anterior end and tapering to the rounded posterior end.

Plaster casts revealed three-dimensional structure that could not be observed in x rays. The pink snapper and the red and long-tailed red snapper's swimbladders have significant rippling on their dorsal and lateral sides. As many as eight bumps on the dorsal side of the swimbladders and eight corresponding diagonal grooves on the sides of the swimbladders are evident. This feature appears to be formed by the swimbladder pushing against muscles that are attached to the spine of the fish. These ripples are not evident in the Brigham's snapper, the blue-striped snapper, or von Siebold's snapper.

The standardized volume of the swimbladder (i.e., the percent of the fish's body volume) showed species-specific differences (Table I). Intraspecific variation was similar to that observed by Ona (1990). The standardized lengths of the axes of the swimbladder (i.e., the percent of the fish's body length) taken together also showed species-specific differences. The swimbladder of the blue-striped snapper and Brigham's snapper were the most distinctive because of their relatively small overall bladder size. Intraspecific variation in the relative sizes of these axes seemed to be related to the fullness of the gut. In particular, the swimbladders of fish with full stomachs were dorso-ventrally flattened, increasing the length of the minor axis and decreasing the length of the vertical axis. This affected the lateral-aspect swimbladder cross-sectional area more than the dorsal-aspect area.

The orientation of the swimbladder relative to the dorsal aspect of the fish was also species-specific. In all six fish species, the swimbladder was tilted backwards from the dorsal aspect of the fish; the anterior end of the swimbladder was higher than the posterior end. Again, the blue-striped snapper and the Brigham's snapper were the most different from the other species; their swimbladders were only slightly angled relative to the fish's dorsal aspect. The three primary

TABLE I. Characteristics of snapper swimbladders.

Species	Local common name	English common name	*n*	% Body length									% Body volume			Angle		
				Major			Vertical			Minor								
				mean	min	max	mean	min	max	mean	min	max	mean	min	max	mean	min	max
Etelis carbunculus	Ehu	Red	10	24.7	20	36	6.7	3.6	10	6.5	4.6	11	3.6	2.7	6.3	12.2	7	14
Etelis coruscans	Onaga	Long-tailed red	10	35.7	32	41	9.7	6.7	14	8.6	6.3	12	5.2	4.6	8.1	8.4	8	9
Pristipomoides filamentosus	Opakapaka	Pink	11	33.5	23	34	9.3	5.6	14	9.5	6.2	13	3.3	2.9	5.2	10.6	8	12
Pristipomoidess zonatus	Gindai	Brigham's	1	30.8			5.2			9			1			1.5		
Pristipomoides sieboldii	KaleKale	von Siebold's	1	23			7			5.9			2.4			14		
Lutjanus kasmira	Taape	Blue-striped	1	23.1			7			5.6			1			3		

species had mean swimbladder angles that varied between 8.4° and 12.2° with the red snapper's swimbladder having the largest angle (Table I).

B. Target strength

The angle at which the maximum dorsal-aspect target strength was measured had a nearly one-to-one relationship (slope=1, $p>0.10$, power=0.93) with the angle of the swimbladder relative to the dorsal aspect of the fish, regardless of species ($r^2=0.88$) (Fig. 4). The points that show the most difference between the two measures are the blue-striped snapper and the Brigham's snapper, both having rounded dorsal surfaces to their swimbladders and a small angle between their swimbladder and their dorsal aspect.

The total energy dorsal aspect target strengths measured with the frequency-modulated and dolphinlike signals had a strong relationship ($r^2=0.89$) the slope of which was not significantly different from 1 ($p>0.10$, power=0.86). The same is true of the lateral aspect target strengths ($r^2=0.84$, $p>0.10$, power=0.83). The echoes resulting from the frequency-modulated signal, which has a broader frequency range, are presented. For each of the three primary species, there was a strong, linear relationship between the log of the standard length of the fish in cm (F_{SL}) and its dorsal aspect target strength [Fig. 5(a)]. Some differences were evident between species. The pink snapper had the steepest slope in the relationship between log of standard length and dorsal target strength ($TS=20.6^*\log(F_{SL})-55.1$, $r^2=0.85$). The red snapper ($TS=13.7^*\log(F_{SL})-46.6$, $r^2=0.54$) and long-tailed red snapper ($TS=12.6^*\log(F_{SL})-42.9$, $r^2=0.80$) had similar slopes but the long-tailed red snapper had target strengths that were about 2 dB higher for equivalently sized fish. The target strengths of the three individual snappers that were not target species fit the same general, length–target strength relationship.

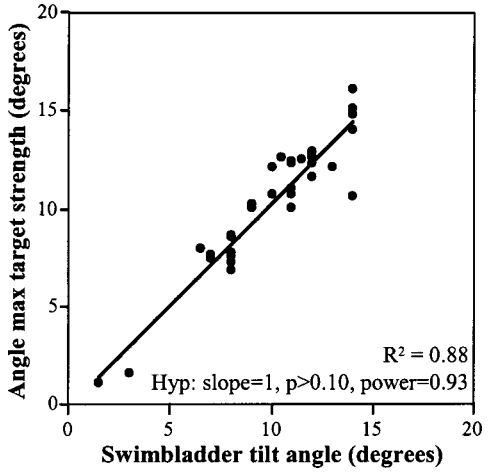


FIG. 4. Relationship between the angle of the swimbladder of each fish relative to its dorsal axis as measured from the lateral x rays, and the angle at which the maximum target strength of the same fish was measured. The slope of the line was not significantly different from 1 (slope=1, $p>0.10$, power=0.93).

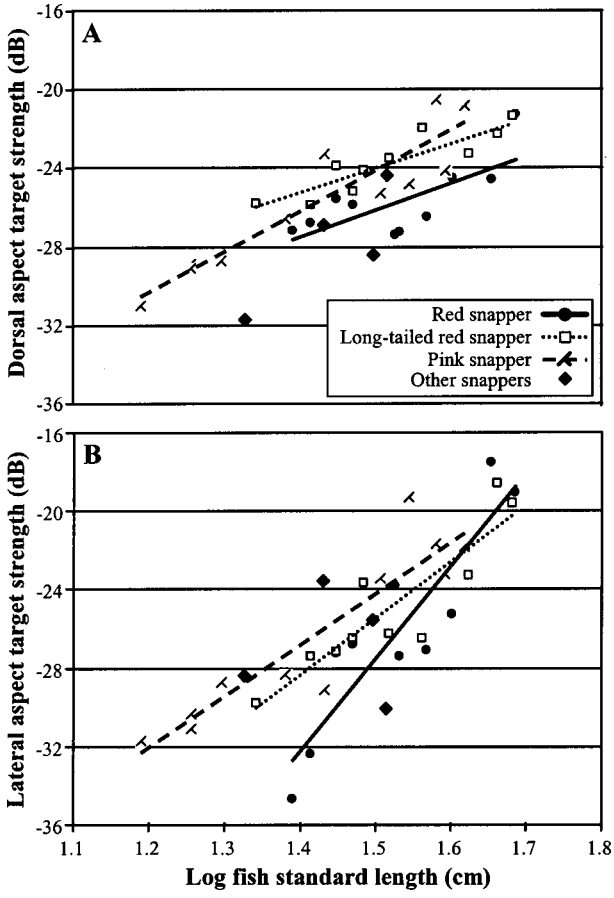


FIG. 5. Relationship between the standard length of each fish species and its dorsal-aspect (a) and lateral-aspect (b) target strength. *F*-tests reveal that the slopes of all lines are significant ($p<0.01$).

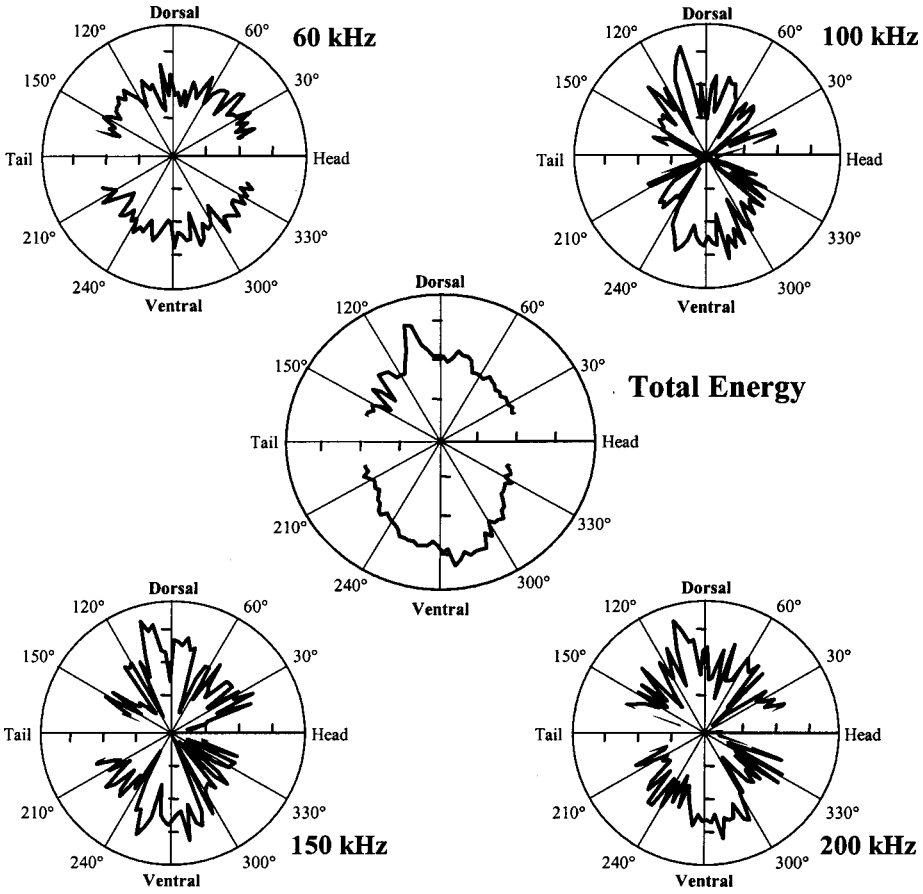


FIG. 6. Target strength as a function of dorsal tilt angle for a 41-cm-long pink snapper. The total energy target strength is shown in the center, surrounded by the target strength at discrete frequencies. The target strengths are scaled to the maximum observed for the fish. Each tick mark represents a loss of 10 dB from the maximum value, up to 35 dB below the maximum value. Echoes taken from within 15° of the head and tail of the fish were dominated by the mounting rig and are not shown.

For the three primary species, there was also a strong, linear relationship between the log of the fish's standard length and its lateral aspect target strength ($p<0.05$ for all comparisons) [Fig. 5(b)]. The relationship between the length and the pink snapper's lateral aspect target strength could be described by the equation $TS=26.2*\log(F_{SL})-63.5$ ($r^2=0.85$). The red snapper had the steepest slope in the relationship between log of standard length and target strength ($TS=47.2*\log(F_{SL})-98.3$, $r^2=0.81$). The long-tailed red snapper's lateral aspect target strength–length relationship can be described as $TS=28.9*\log(F_{SL})-68.8$ ($r^2=0.81$).

Consistent effects of frequency on the target strength of fish were not observed. Other studies have found that target strength decreases with increasing frequency, specifically by 0.9 log frequency (see Love, 1969, 1970; Urick, 1983). While the range of variation in target strength as a function of frequency was approximately the same as that predicted by this equation, no consistent change was observed as a function of frequency. Essentially, the log length–target strength regression lines for each frequency cross both in the dorsal and lateral aspect.

The variance in target strength as a function of the fish's orientation for the entire rotation in each plane increased with increasing frequency for all species (Fig. 6). Variance values increased from 13 to 49 dB^2 with increasing frequency. The total energy target strength variance, with values between 0.1 to 11 dB^2, was lower than the variance of all discrete frequencies, in all three planes. Differences in the magnitude of target strength variance as a function of angle were observed between species (Fig. 7). In the tilt plane, the red snapper had a larger variance than the other species. In the roll plane, the red snapper had a lower variance than the other species, although the overall variance for all three species in this plane was low. The variance in the lateral plane was much higher in the pink snapper than in the other species; both the red and long-tailed red snappers had very low variance in this plane.

The $\pm 15°$ about both the dorsal and lateral aspects of the fish are the most important for utilizing these measures in a field study. The variance over the $\pm 15°$ about each major axis was generally decreased compared with the variance over the entire fish. Species differences in the magnitude of variance in target strength about the tilt and lateral axes were evident. However, there were no differences in variance between species in the roll plane over this limited range of angles; the variance in target strength of all three species was extremely low. The target strength over this range of angles had a range of 8 to 12 dB in the tilt plane, 2.5 to 7 dB in the roll plane, and 4 to 6 dB in the lateral plane. Single factor analysis of variance revealed that there were significant differences in the range of target strengths observed in all three planes as a function of species ($p<0.05$). Long-tailed red snapper had the greatest range of target strength values in the tilt plane and the most limited range of target strength values in the roll and lateral planes. Red snapper and pink snapper had similar ranges in their target strengths over the $\pm 15°$ about each major axis.

Dorsal-aspect acoustic backscattering cross section was linearly related to wet weight in grams, a measure of bio-

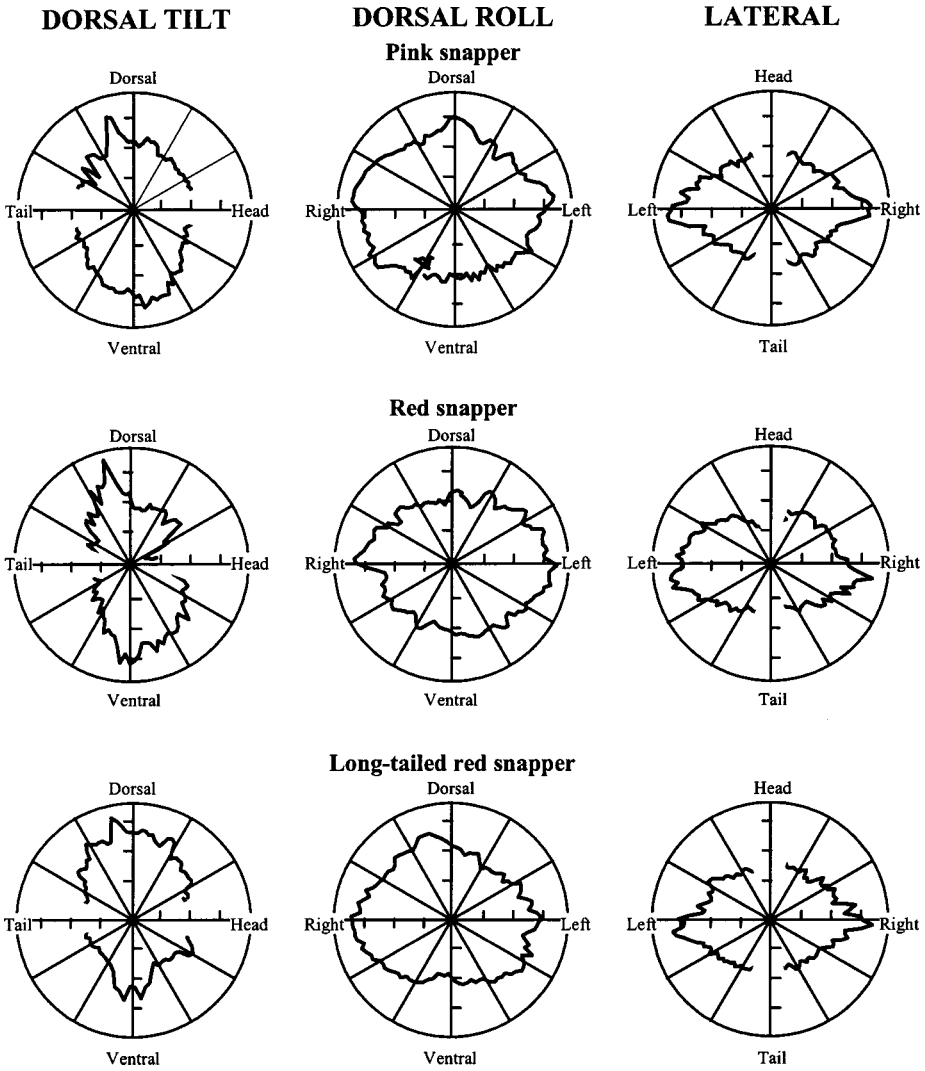


FIG. 7. Total energy target strength of a representative of each species of fish as a function of orientation. Dorsal tilt is shown on the left, dorsal roll in the center, and lateral aspect on the right. Each tick mark represents a loss of 10 dB from the maximum value, up to 35 dB below the maximum value. Echoes taken from within 15° of the head and tail of the fish were dominated by the mounting rig and are not shown.

mass, for each snapper species (Fig. 8) [red snapper wet weight $(g) = 22983^{*}\ \text{sigma}\,(\text{m}^2) - 18.39$ $(r^2 = 0.82,\ p < 0.05)$; long-tailed red snapper wet weight $(g) = 24631^{*}\ \text{sigma}\,(\text{m}^2) - 527.55$ $(r^2 = 0.71,\ p < 0.05)$; pink snapper wet weight $(g) = 12665^{*}\ \text{sigma}\,(\text{m}^2) - 122.91$ $(r^2 = 0.77,\ p < 0.05)$]. Similar relationships were observed between acoustic backscattering cross section and biovolume. Acoustic backscattering cross section did not predict biovolume and wet weight significantly differently than the cube of length, a standard predictor ($p > 0.05$ for both comparisons). These relationships could be helpful in estimating fish biomass in future field studies of these species, without needing length-weight curves for each species (Benoit-Bird and Au, 2002).

C. Models

Comparison between maximum measured target strength values and those predicted by various models for lateral and dorsal aspect target strength showed only the prolate spheroid model was accurate and precise (Fig. 9). The predictions from the sphere model, based on the volume of the swimbladder, were significantly different from both the lateral aspect and dorsal aspect target strength ($p < 0.005$). Both target strength models using fish length, based on the

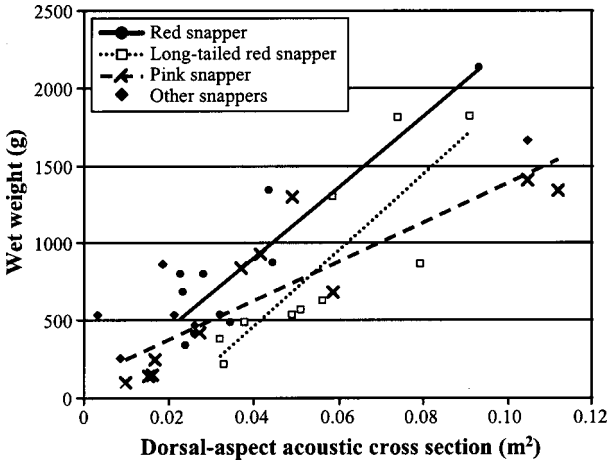


FIG. 8. The relationship between the dorsal-aspect acoustic cross-section and biomass, expressed as wet weight. There is a strong relationship and a significant slope ($p < 0.05$).

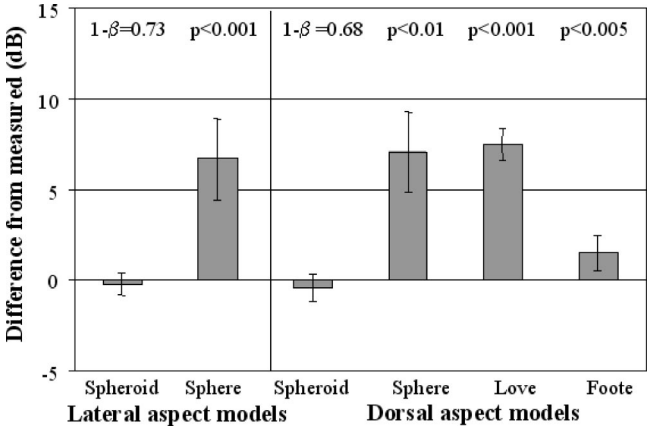


FIG. 9. Comparison between measured target strength values and those predicted by various models. Lateral aspect models, one modeling the swimbladder as a prolate spheroid based on the axes of the swimbladder measured from an x ray, and one modeling the swimbladder as a sphere with a volume equivalent to that measured of a plaster cast of the swimbladder. The dorsal aspect models are shown on the right; again the prolate spheroid and sphere models for target strength based on the swimbladders are included as are the models based on fish length for all types of fish (Love, 1970) and only on fish possessing a swimbladder (Foote, 1980). Positive values indicate that the model, on average, underestimated the fish's target strength and negative values indicate the model overestimated the fish's target strength. Error bars show 95% confidence intervals. The p values are indicated for models that differ significantly from the data, compared with paired t-tests corrected for multiple comparisons using the Bonferroni method. Power values are shown for the models that are not significantly different from the data.

equations of Love (1970) and Foote (1980), predicted values that were significantly different than the maximum dorsal aspect target strength ($p<0.005$). Foote's equation for fish with air-filled swimbladders predicted maximum target strength significantly better than the general fish equation of Love ($p<0.05$). The target strengths predicted by the prolate spheroid model of the swimbladder were not significantly different from the maximum measured target strengths from either the dorsal or lateral aspects (power=0.73). The measured target strengths were all within 3 dB of the predictions of the prolate spheroid model.

On average, the lateral aspect target strength was 0.8 dB higher than the dorsal aspect target strength of the same fish. The relationship is reversed in more than one-third of the individuals, with the dorsal aspect target strength being higher than the lateral, unlike the trend that is commonly observed (Love, 1969). In all fish, the prolate spheroid model correctly predicted which aspect should have the higher target strength, if not the difference between the two. However, it is important to note that the target strength values used here are based on energy over a 60-kHz band while values for the other models were at a specific frequency.

IV. DISCUSSION

X rays and swimbladder casts revealed that the swimbladder shapes among different snapper species varied significantly. This is evidenced in the percent of body length of the three swimbladder axes. The differences are visible in the swimbladder x rays where differences are particularly noticeable in the degree of curvature or linearity of the swimbladders. Interestingly, the fish in the genus *Pristimopoides*, closely related fish, showed remarkable differences in their swimbladder shapes. Likewise, the swimbladder characteristics of two fish species in the genus *Etelis* were very different. The fish within each genus were more different from each other than some of the other, less related fish. This is obvious in the swimbladder volume, expressed as percent of body volume. The differences in volume between fish are greater within each genus than between them.

While swimbladder shape and volume varied between species, target strength values among the three target species did not greatly vary. Target strengths, both lateral and dorsal, of all three target species were correlated with standard length. The lateral aspect target strength was predicted about equally well by standard length in all three species. The dorsal aspect target strength was predicted equally in the long-tailed red snapper and the pink snapper. However, the relationship between length and lateral aspect target strength was weaker in the red snapper. The relationship between fish length and the axes of the fish swimbladders was strong in both the long-tailed red snapper and pink snappers, with r^2 values greater than 0.75 for the major and minor axes of the swimbladders, and greater than 0.44 for the vertical axis. However, the red snapper had a strong correlation between length and only the major axis of the swimbladder ($r^2=0.83$). Their was no significant relationship between red snapper length and the length of the minor or vertical axes of its swimbladders ($r^2<0.14$, slope$\neq 0$ $p>0.15$, power=0.48). Because swimbladder size is a primary factor responsible for the strength of echoes, this lack of a correlation between two of the swimbladder size measures and length translates to a weaker relationship between red snapper length and target strength.

Swimbladder shape in all three target species varied most in the dorso-ventral plane. Changes in swimbladder shape were particularly noticeable in fish with full stomachs. Because the dorsal aspect cross-sections of the swimbladders varied less than the lateral-aspect cross sections, variance in the lateral-aspect target strength between individuals within a species was greater than target strength variance in the dorsal aspect. Variance in target strength within individuals about the lateral axis was also high. Consequently, attempts to estimate snapper size from target strength when fish cannot be directly observed would be more accurate from the dorsal than the lateral aspect, suggesting the most appropriate tool for field surveys would by a downward-looking sonar system or upward-looking, bottom-mounted devices.

The lack of a relationship between frequency and target strength is contrary to previous work. Other studies have found that target strength decreases with increasing frequency (f), specifically by $0.9\log(f)$ (Love, 1969, 1970; Urick, 1983). This is caused by the complex relationship between target strength and frequency within each fish, discussed in Au and Benoit-Bird (2003).

As expected, the directivity of target strength values over each plane of the fish increased with increasing frequency (Urick, 1983) and contained many nulls and local peaks. However, the broadband, total energy target strength had the lowest directivity. The broadband pattern based on

total energy was considerably smoother than the patterns at the various frequencies. Local maxima and minima in the narrow-band pattern are the results of constructive and destructive interferences of the scattered signals. The interference effect is not as strong with broadband signals. Utilizing the broadband target strength of these fish could reduce the error associated with estimating their size when they have an unknown distribution of tilt and roll angles. Differences between species in the directivity of target strength, particularly within 15° of the dorsal axis, can introduce differences in this error, assuming an equivalent distribution of orientations. The use of broadband target strength estimates would reduce the differences.

A strong correlation between the dorsal aspect acoustic backscattering from each of the three target snapper species and their wet weight, a measure of biomass, was observed. This permits a direct conversion of acoustic backscattering to biomass without knowledge of the size distribution of the population (Benoit-Bird and Au, 2002). While some differences in the relationship were observed between species, the combined relationship for all three species was still strong ($r^2=0.65$, $p<0.05$). This could permit a conversion of acoustic scattering to snapper species biomass, even if specific species identification were not possible.

Modeling acoustic backscattering strength using measures of fish physical characteristics can elucidate the factors affecting sound scattering. The prolate spheroid model of Furusawa (1988) accurately predicted the maximum target strengths of all species of snappers, based only on simple size characterization of the swimbladder. The model also accurately predicted the direction of differences between dorsal and lateral-aspect target strength. None of the other models used, based on fish length or swimbladder volume, accurately predicted target strength. It is likely that the Clay and Horne model based on fish swimbladder shape would accurately predict target strength, however the complexities involved in calculating the model were beyond the scope of this work. Models based on fish length varied in their ability to predict target strength in different species because of small species-specific differences in length–target-strength relationship. Both the Love and Foote models use total fish length as the predictive variable. The relationships between fish total length and target strength were significantly weaker for all three species than the relationships between fish standard length, a measure of only the firm tissue of the fish ignoring the tail, and target strength ($p<0.05$ for all comparisons). This indicates that variance in fish target-strength–length relationships could be reduced by utilizing standard length instead of total length. The prolate spheroid model showed no differences in its ability to predict target strength. Unlike the other models of target strength used, the prolate spheroid model overestimated the target strength of the fish. This is likely because of the simple swimbladder shape assumed by the model and the contribution of the rest of the fish's body to the backscattering. The accuracy and generality of the predictions of target strength based on these simple measures of swimbladder size, however, indicate the importance of the swimbladder in the echo strength of these fish. This is confirmed by the correlation between swimbladder tilt angle and the angle of maximum reflection.

Few detailed studies of acoustic backscattering strength by a group of closely related fish have been reported. The questions of frequency, orientation, biomass, and the relationship between swimbladder shape, orientation, and size were investigated utilizing a combination of techniques. Specific target-strength–length and backscattering–biomass relationships were determined for the three most abundant snapper species. The effect of frequency on scattering strength was unpredictable, unlike previous results. Orientation effects on backscattering strength show the potential of broadband target strength measures to reduce errors associated with an unknown distribution of fish orientation in the wild. The results of target strength models based on x ray measures of swimbladder characteristics indicate the importance of the shape of a snapper's swimbladder size on its backscattering strength. These results also provide an important base for the utilization of sonar techniques for field studies of snappers in the Hawaiian Islands.

ACKNOWLEDGMENTS

Aaron and Virginia Moriwake maintained the fish while in the hatchery, and assisted in the capture of the fish from the hatchery, design of the fish mounting setup, and data collection. Bo Alexander caught the fish from the wild, assisted in their maintenance, and captured fish from the hatchery for study. More than this, however, they shared their knowledge of the fish, anesthesia techniques, and their enthusiasm for the Bottomfish Project as a whole. Romeo Estabilio and the Queen's Hospital Imaging Department generously donated their time, expertise, and materials to provide all of the fish x rays. Kristine Hiltunen, a student in the National Science Foundation's Research Experience for Undergraduates Program, assisted in data collection and processing. Kimberly Andrews and Caroline DeLong assisted in data collection. Jeff Pawloski provided valuable information regarding x-ray facilities. This work was supported by the Hawaii State Department of Land and Natural Resources. This is HIMB contribution 1166.

Au, W. W. L., and Benoit-Bird, K. J. (**2003**). ''Acoustic backscattering by Hawaiin lutjanid snappers. II. Broadband temporal and spectral structure,'' J. Acoust. Soc. Am. **114**, 2767–2774.

Benoit-Bird, K. J., and Au, W. W. L. (**2002**). ''Energy: Converting from acoustic to biological resource units,'' J. Acoust. Soc. Am. **111**, 2070–2075.

Benoit-Bird, K. J., Au, W. W. L., Kelley, C. D., and Taylor, C. (**2003**). ''Acoustic backscattering from deepwater measured in situ from a manned submersible,'' Deep-Sea Res., Part I **50**, 221–229.

Clay, C. S., and Horne, J. K. (**1994**). ''Acoustic models of fish: The Atlantic cod (*Gadus morhua*),'' J. Acoust. Soc. Am. **96**, 1661–1668.

Do, M. A., and Surti, A. M. (**1990**). ''Estimation of dorsal aspect target strength of deep-water fish using a simple model of swimbladder backscattering,'' J. Acoust. Soc. Am. **87**, 1588–1596.

Foote, K. G. (**1980**). ''Importance of the swimbladder in acoustic scattering by fish: A comparison of gadoid and mackeral target strengths,'' J. Acoust. Soc. Am. **67**, 2084–2089.

Foote, K. G., and Ona, E. (**1985**). ''Swimbladder cross secitons and acoustic target strengths of 13 pollack and 2 saithe,'' Fiskeridirektoratetes **18**, 1–57.

Furusawa, M. (**1988**). ''Prolate spheroidal models for predicting general trends of fish target strength,'' J. Acoust. Soc. U. K. **9**, 13–24.

Love, R. H. (**1969**). ''Maximum side-aspect target strength of an individual fish,'' J. Acoust. Soc. Am. **3**, 746–752.
Love, R. H. (**1970**). ''Dorsal-aspect target strength of an individual fish,'' J. Acoust. Soc. Am. **49**, 816–823.
MacLennan, D. N. (**1990**). ''Acoustical measurement of fish abundance,'' J. Acoust. Soc. Am. **87**, 1–15.
McClatchie, S., Alsop, J., and Coombs, R. F. (**1996**). ''A re-evaluation of relationships between fish size, acoustic frequency, and target strength,'' ICES J. Mar. Sci. **53**, 780–791.
Ona, E. (**1990**). ''Physiological factors causing natural variations in acoustic target strength of fish,'' J. Mar. Biol. Assoc. U.K. **70**, 107–127.
Stanton, T. K. (**1989**). ''Simple approximate formulas for backscattering of sound by spherical and elongated objects,'' J. Acoust. Soc. Am. **86**, 1499–1510.
Thiebaux, M. L., Boudreau, P. R., and Dickie, L. M. (**1991**). ''An analytical model of acoustic fish reflection for estimation of maximum dorsal aspect target strength,'' Can. J. Fish. Aquat. Sci. **48**, 1772–1782.
Urick, R. J. (**1983**). *Principles of Underwater Sound* (McGraw–Hill, New York).
Western Pacific Regional Fisheries Management Council (**1999**). ''Bottomfish and Seamount Ground fish of the Western Pacific Region, 1998,'' Annual Report.

Acoustic backscattering by Hawaiian lutjanid snappers. II. Broadband temporal and spectral structure

Whitlow W. L. Au[a] and Kelly J. Benoit-Bird
Hawaii Institute of Marine Biology, P.O. Box 1106, Kailua, Hawaii 96734

(Received 13 September 2002; accepted for publication 11 August 2003)

The characteristics of acoustic echoes from six species of deep-dwelling (up to 400 m) Hawaiian Lujanid snappers were determined by backscatter measurements at the surface. A broadband linear frequency-modulated signal and a short dolphinlike sonar signal were used as the incident signals. The fish were anesthetized and attached to a monofilament net that was attached to a rotor so echoes could be collected along the roll, tilt, and lateral axes of each fish. The temporal highlight structure of broadband echoes was determined by calculating the envelope of the cross-correlation function between the incident signal and the echoes. The echo waveforms were complex with many highlights and varied with the orientation of the fish. In the tilt plane, the strongest echoes occurred when the incident signal was perpendicular to the long axis of the swimbladder. The number of highlights was the fewest at this orientation. The number of echo highlights and the length of echoes increased as the fish was tilted from this orientation. The highlight structure of the echoes resulted in the transfer function being rippled, with local maxima and minima that changed with fish size and species. The echo structures in both the time and frequency domains were generally consistent within species and were easily distinguishable between species. [DOI: 10.1121/1.1614257]

PACS numbers: 43.30.Sf [WMC] Pages: 2767–2774

I. INTRODUCTION

Information on the behavior, movement patterns, habitat utilization, and abundance of deepwater lutjanid snappers in Hawaii, an important and threatened fishery, is extremely limited (Haight *et al.*, 1993). Various acoustic techniques have the potential to provide important information to fill these gaps (MacLennan, 1990; MacLennan and Holliday, 1996; MacLennan and Simmonds, 1992; Simmonds and MacLennan, 1996). However, the difficulties identifying and estimating the abundance of species with acoustics limits the utilization of these techniques.

In order to identify species using acoustics, more information is required than can be obtained from a single frequency. Multiple-frequency techniques have been suggested as an effective way to estimate the size distributions and total abundance of many classes of organisms in the ocean (see a review in Greenlaw and Johnson, 1983). These methods have primarily been limited to discrete frequencies that must be carefully selected with knowledge of the scattering characteristics of potential targets.

Broadband acoustic signals, those that contain a continuous, wide range of frequencies rather than a few, discrete frequencies, have the potential to provide significant information about targets (Barr, 2001; Burdic, 1968). Species may reflect a broadband acoustic signal differently and these echo differences may be obvious in the time or frequency domains (Zakharia *et al.*, 1996). For example, differences may result in target strengths at specific frequencies; the number, position, and amplitude of echo highlights or spectral ripples as well as changes in these characteristics as a function of the orientation of the target (Au and Snyder, 1980; Urick, 1983). Using a sonar with a broadband signal that has a good frequency and temporal resolution, the backscatter characteristics of different species may be resolvable.

In order to identify fish in the wild, however, information is needed on the characteristics of the population of interest and the relationship of these characteristics to the acoustic properties of the target populations. A mathematical or theoretical solution to this problem, particularly for complicated, acoustically understudied targets such as fish, is not currently available. The objectives of Part II of this work were to (1) determine if species-specific differences in the broadband characteristics of closely related Hawaiian lutjanid snappers exist; (2) quantify these differences; and (3) determine how these differences are affected by fish size. Three specific species of lutjanid snappers are of prime interest, the onaga or long-talied red snapper (*Etelis coruscans*), the ehu or red snapper (*Etelis carbuncuhus*), and the opakapaka or pink snapper (*Pristipomoides filamentosus*). These three species are the most critical to the lutjanid fishery (Western Pacific Regional Fisheries Management Council, 1999) and are the most depleted in the main Hawaiian Islands.

II. METHODS

Acoustic data were collected as in Part I of this work (Benoit-Bird *et al.*, 2003a). Lutjanid snappers caught off the coasts of the Hawaiian Islands were allowed to acclimate to ambient conditions for at least eight days before their backscattering properties were measured. Although these lutjanid snappers are deep dwelling (up to 400 m), backscatter measurements done both at the surface and *in situ* at 250 m indicate that once the fishes acclimatize to surface condi-

[a]Electronic mail: wau@hawaii.edu

tions, their swimbladders retain a similar shape and volume as in deep waters (Benoit-Bird *et al.*, 2002). Live individual fish were anesthetized with 1 mL of 2-phenoxy-ethanol per 10 L of seawater and enclosed in a fitted bag made of monofilament netting. The net bag was mounted to a large, weighted, monofilament net that was turned by a rotor 360°.

Two broadband signals were used to measure backscattering; a frequency-modulated sweep with a frequency range of 60 to 200 kHz, and a dolphinlike click with a peak frequency of 120 kHz and a 60-kHz bandwidth. The ratio of fish length to acoustic wavelength at the peak frequency of the dolphinlike signal varied from 12 to 38, placing the results in the intermediate frequency range, where the swimbladder accounts for 90% to 95% of the reflected energy (Foote and Ona, 1985). The waveform and frequency spectrum of the signals are shown in Fig. 2 of Benoit-Bird *et al.* (2003a). The results from the dolphinlike signal were considered for time-domain analyses. The short duration of this signal (80 μs) compared to the FM signal (500 μs) made the highlight or echo structure more apparent without any special processing; however, the results with both signals are very similar after cross-correlating the echoes with the transmitted signal. For spectral domain analyses, the results from the frequency-modulated sweep were utilized because they included a slightly wider frequency range. Again, the results from the transfer functions of the two signals were similar above 75 kHz.

The outgoing signals were produced using a function generator computer plug-in board. The function generator also produced a trigger signal for each transmission. After a delay related to the two-way travel time from the signal to the target, a trigger prompted a Rapid System R1200 analog-to-digital (A/D) converter to digitize and store a block of 1024 samples. Sampling rates of 1 MHz were used for the function generator and the A/D converter. The delayed trigger also caused the rotor and net to advance by an incremental angle. Echoes were collected in 1.5°–2.5° increments about each of the fish's three axes for both source signals. The transmit and receive transducers, with an effective 12° 3-dB beamwidth at the center frequency of the signals, were set up 2.2 m deep, the same depth as the mounted fish, approximately 6 m from the fish. The use of broadband signals virtually eliminates the presence of side lobes (Au, 1993). After acoustic measurement, fish were sacrificed with a 2-mL/10-L dose of 2-phenoxy-ethanol. The standard length, total length, displacement volume, and wet weight of each fish were measured and the fish was immediately frozen.

A. Data analysis

The incident signals were measured and digitized with the receiving hydrophone located at the position of a target fish, directly facing the projecting transducer. Reflected signals were compared with the incident signals and corrected for gain. The envelope of the cross correlation between each echo and the incident signal were examined to determine their basic time-domain characteristics. The cross-correlation function was determined by the Fourier transform technique using the equation

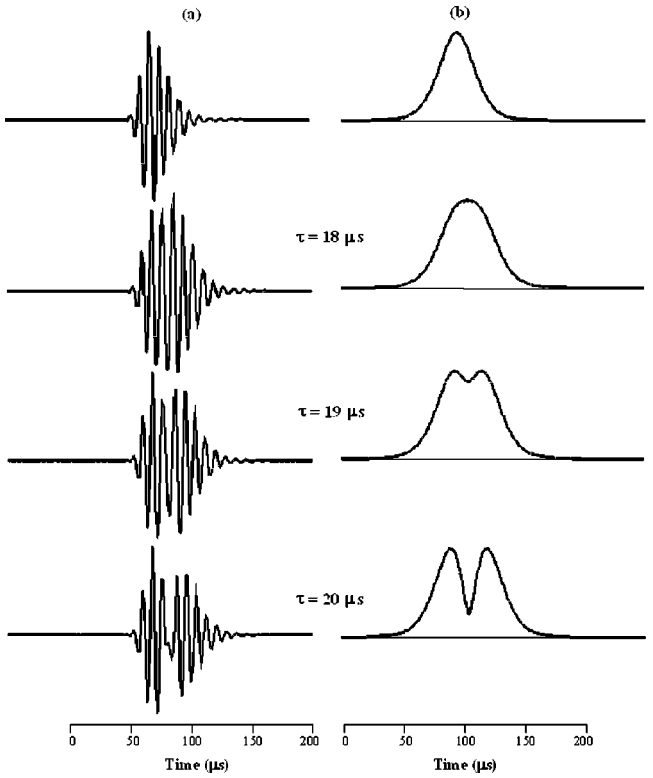


FIG. 1. The top left-hand panel is the waveform of the transmitted dolphinlike signal with the envelope of its auto-correlation function to the right. The remaining panels are simulated echoes consisting of the sum of two of the transmitted signal separated by a time τ (left) along with the envelopes of the cross-correlation functions between the transmitted signal and the simulated echoes (right). This figure shows the time-resolution property of the transmitted signal.

$$c(t)=\Im^{-1}[E(f)U(f)] \tag{1}$$

(Brigham, 1988) where $E(f)$ and $U(f)$ are the Fourier transform of the echo and incident signals, respectively, and $\Im^{-1}$ denotes the inverse Fourier transform of the terms in the brackets. The envelope of the cross-correlation function was calculated by converting $c(t)$ into an analytic signal using the Hilbert transform method where the absolute value of the analytic signal represents the envelope of the signal (Barr, 2001; Burdic, 1968).

The time resolution capability of the dolphinlike signal can be determined by simulating reflections from two discrete point targets separated by a travel time of τ as expressed by the equation

$$e(t)=s(t)+s(t-\tau), \tag{2}$$

where $s(t)$ is the incident signal that is reflected perfectly and $s(t-\tau)$ is another reflection delayed by a time τ. The result of this simulation is in Fig. 1, where the waveforms are under column **a**, and under column **b** are the envelopes of the cross-correlation functions. The first signal in column **a** is the incident signal, followed by signals described by Eq. (2) with various values of τ. The envelopes under column **b** in Fig. 1 indicate that highlights must separated by at least 19 μs before they are resolvable. Note that for τ=18 the two highlights are not resolvable and the envelope of the cross-correlation function is wider than for the incident signal. This property and the 3-dB width of the correlation function (22 μs) can be used to indicate whether a second highlight might

be hidden or unresolvable at different portions of an echo. The number of highlights, or glints, in each echo waveform was analyzed for each angle within 15° of the dorsal aspect, in both the tilt and roll planes.

Other characteristics of the waveforms were analyzed for echoes obtained from the dorsal aspect of each fish and at the tilt angle where maximum echo strength was measured. These characteristics, the relative amplitude of each highlight, the distance between highlight peaks, and the 3-dB width of each highlight, were compared between species, sizes, and angles.

The transfer function of dorsal aspect frequency-modulated echoes was determined using the equation

$$H(f) = 20 \log \left| \frac{E(f)}{U(f)} \right|. \quad (3)$$

The transfer function is characterized by an intricate spectrum with many peaks and nulls. Nulls were defined as decreases in echo intensity from the intensity at surrounding frequencies by at least 10 dB. Peaks were put into two categories, frequencies at which the intensity was 5 dB greater than surrounding intensities and those with 10 dB greater intensities.

III. RESULTS

The results are presented from the perspective of a broadband sonar pointed vertically downward. The waveforms of dorsal aspect echoes for the different species are shown in Fig. 2. The envelope of the cross-correlation function for each echo is represented by the dashed curve. From the time representation of the echoes, it is obvious that echo structures are very different between the species and that many highlights exist for most of the echoes. The echo structure is very complex with portions within echoes where two or more unresolvable highlights are apparently present. These are indicated by the width of the cross correlation function and the many oscillations of the signal within the wide correlation peaks. The echo structure obtained from 10–11 of each of the targeted or primary species were similar except for the highlight intervals which were somewhat related to fish size. The relationship between echo highlight intervals for highlights 2–8 and fish size for dorsal aspect echoes are plotted in Fig. 3. Some of the intervals changed significantly with fish size while others showed little change.

An example of the backscattered waveforms from a pink snapper as a function of the fish tilt angle is shown in Fig. 4. The lateral aspect x ray of the fish is shown with the tilt angles above each waveform indicating the orientation of the fish with respect to the direction of the sonar signal. At the 0° tilt angle, the longitudinal axis of the fish is perpendicular to the direction of the sonar signal. Each waveform is 500 μs in duration. The shape and orientation of the swimbladder with respect to the direction of the incident signal were the most important factors influencing the backscattered waveforms. The waveforms varied considerably as a function of the tilt angle with the most highlights present at the 15° and 10°

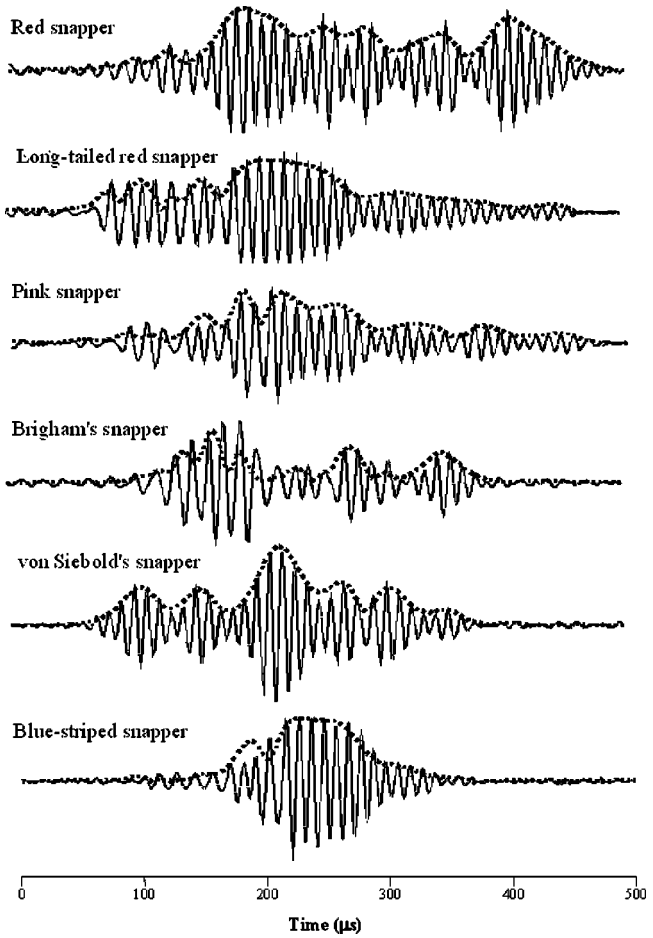


FIG. 2. Representative echoes from the dorsal aspect of six species of lutjanid snappers. The dashed line plotted with each echo is the envelopes of the cross-correlation function of the echo with the transmitted signal.

angles. The backscatter has the highest amplitude when the longitudinal axis of the swimbladder was approximately perpendicular to the sonar beam.

The number of echo highlights as a function of the tilt angle is shown in Fig. 5. The number of highlights varies with the tilt angle of the fish and is the least when the echo is the strongest. The echo is the strongest when the incident signal is perpendicular to the longitudinal axis of the swimbladder (Benoit-Bird *et al.*, 2003a). In all species, the number of highlights generally increases as the fish is tilted further from this point with the number of highlights gradually increasing towards increasing head-up aspect angle. The relative amplitude, interhighlight interval, and 3-dB width of the highlights for the three target species at the dorsal aspect and at the aspect at which the longitudinal axis of the swimbladder is perpendicular to the incident signal are shown in Fig. 6. All three of the parameters of Fig. 6 differed between different species.

It was not within the scope of this study to determine possible sources of reflections for the different highlights in a species. Considerably more knowledge on the scattering processes involved with a fish body, bony structures, and swimbladder shape and volume are required in order to determine the origin of the various highlights in an echo. Figure 7 illustrates the complexity of the task of identifying the sources

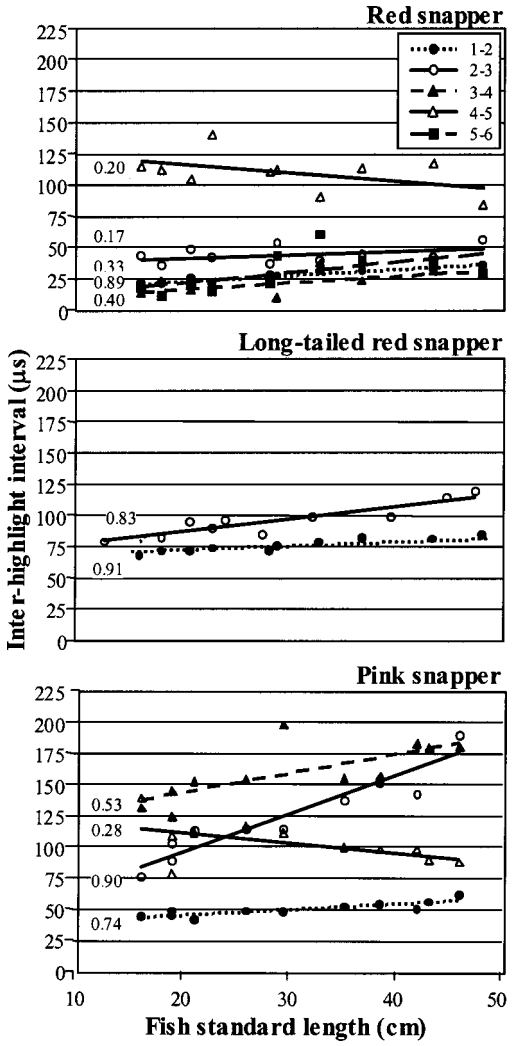


FIG. 3. The interval between each echo highlight as a function of fish length. Regression coefficients are shown for each relationship. Most intervals increased with increasing fish size, although some remained constant or decreased.

of highlights, showing the envelope of the cross-correlation function and an x ray of the corresponding fish (pink snapper). The relative time of occurrence of each highlight with respect to the first reflected component of the echo is shown in a table within the figure. Also shown is the two-way distance in cm that an acoustic signal would travel for the corresponding delay time. If we assume that the first highlight originated at the fish body, then the second highlight at 22 μs might have been from the forward tip of the swimbladder. After this simple explanation for the first two highlights, the situation becomes extremely complex, with two-way delay distances as large as 27.5 cm. The interhighlight interval plot of Fig. 6 also suggests relatively long echo structures that are not easily explainable. Furthermore, the width of most of the peaks of the cross-correlation function shown in Fig. 6 are greater than 22 μs, suggesting that these highlights are composed of several unresolvable highlights.

As with target strength (Benoit-Bird *et al.*, 2003a), changes in echo structure were significantly reduced in the roll plane compared with the tilt plane. There were no significant changes in the number of echo highlights within 10°

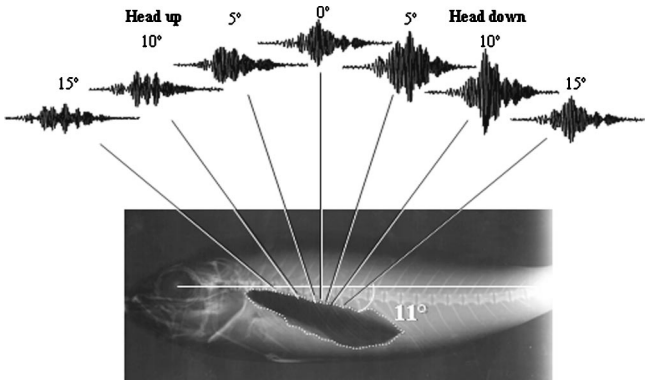


FIG. 4. An example of the echo waveforms from different tilt angles about the dorsal aspect of a pink snapper. The duration of each waveform is 500 μs.

of fish dorsal aspect in the roll plane. The number of highlights increased by one in the long-tailed red snapper and pink snapper at 15° from dorsal and two in the red snapper. There were also few differences in the highlight structure, in terms of relative amplitude and width, over this range of roll angles. The 95% confidence intervals of all highlight characteristics measured at dorsal and 15° in the roll plane overlapped. Consequently, few differences were observed in the transfer functions (the spectral domain) as a function of roll angle.

The transfer functions of the dorsal aspect echoes were similar from both incident signal types. Because the frequency-modulated signal included a slightly wider frequency range, the echoes from this signal are presented. Species-specific differences in the spectral structure of broadband echoes are evident in Fig. 8. Because spectral and temporal structure are related, differences observed in the highlight structure as a function of orientation were mirrored in the spectra. The most conserved feature of spectral structure is the frequency position of nulls, sharp decreases in intensity of at least 10 dB, whose positions are associated with the interhighlight intervals in the time domain echoes.

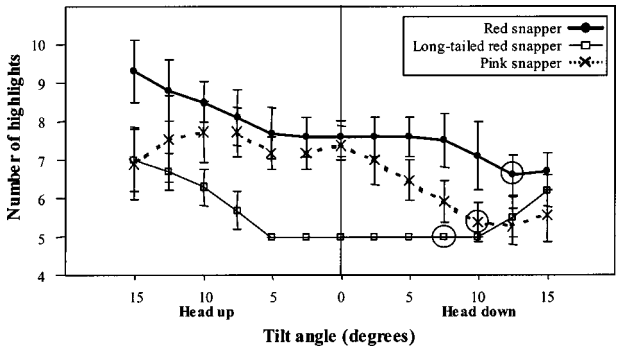


FIG. 5. The number of echo highlights, measured from the envelope of the cross-correlation between the echo and the incident signal as a function of fish tilt angle. The angle at which the strongest echo was obtained is circled. For each species, the number of highlights was the fewest at and around the angle at which the strongest echo was obtained. Error bars indicate one standard deviation.

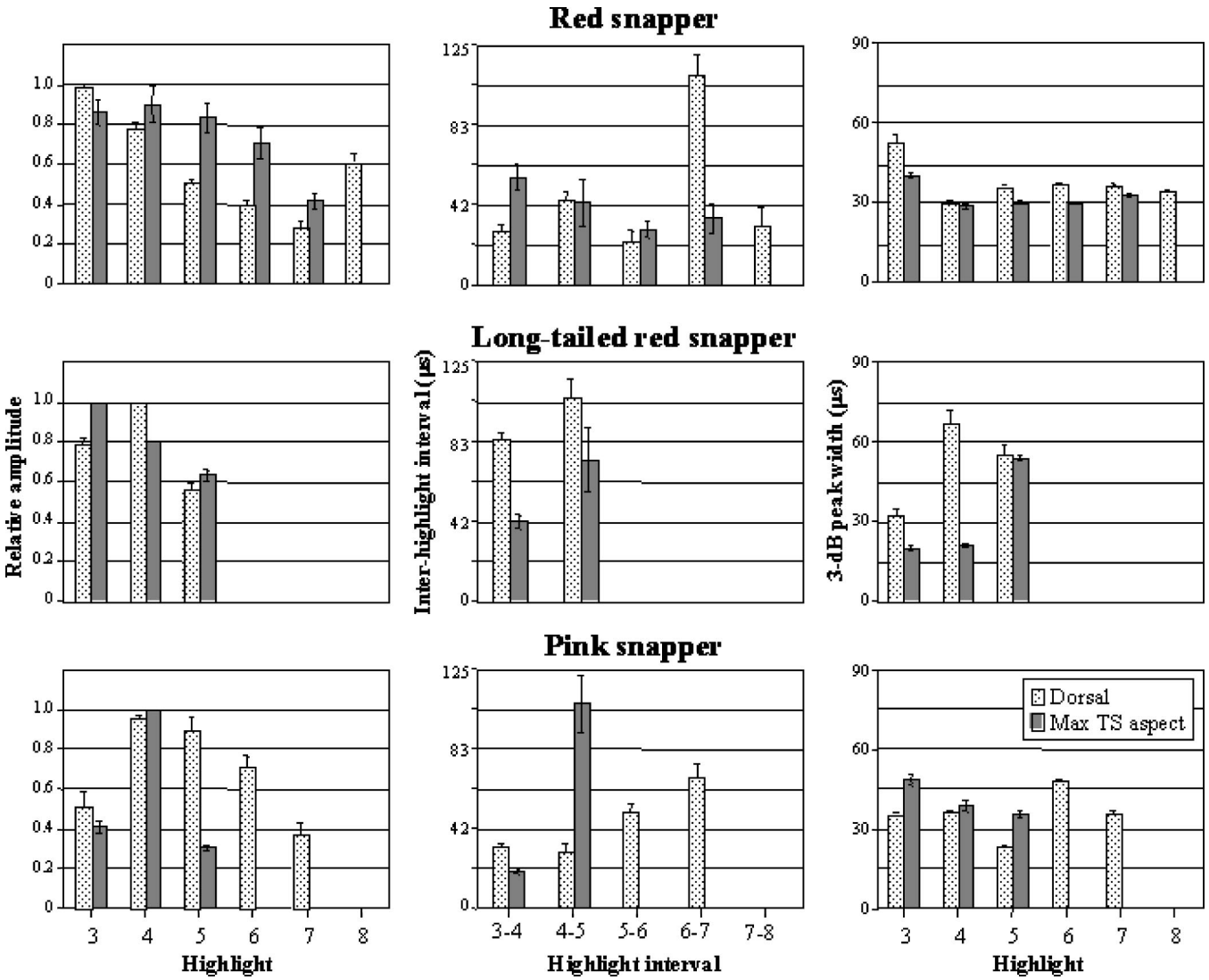


FIG. 6. Relative amplitude of each highlight (left), interval between each highlight peak (center), and the 3-dB width of each highlight (right) for each of the three primary species. Each characteristic is shown for dorsal aspect echoes, light gray, and the maximum amplitude dorsal echoes, dark gray. Error bars show 95% confidence intervals.

The position of spectral nulls was strongly related to fish size as seen in Fig. 9. The shift of the frequencies of spectral peaks and nulls as a function of fish length (L_F) was not constant as L_F/λ, where λ is wavelength. An analysis of variance revealed that there was still a significant effect of length on the position of spectral features after L_F/λ was considered ($p<0.05$). However, when the length of the axes of the swimbladder (L_{SB}) was used, the position of spectral features that varied significantly as a function of length did not vary significantly as a function of L_{SB}/λ ($p>0.05$ for all comparisons). This was particularly evident in the red snapper where swimbladder size and fish length are not as strongly correlated as in other species.

IV. DISCUSSION AND CONCLUSIONS

Echo highlight structure varied between fish species. While there was strong overlap in some highlight characters, in particular, the relative amplitude of highlights, when all three characters measured are utilized in concert, species differentiation is possible from echo highlight characters. Within each species, the number of highlights in echo waveforms increased off-axis. The further each fish was tilted from its dorsal maximum target strength aspect, the greater the number of echo highlights. The length of the echo also increased off-axis. The shortest echo was observed in the highest amplitude echo. Roll plane changes in echo highlights were much smaller than those in the tilt plane. While the time-domain characteristics of echoes changed with aspect, the variance in these characteristics was much less within species than between them within $\pm 15°$ about the dorsal aspect.

Species-specific differences in broadband echo characteristics were also visible in the spectral domain. The number of peaks and nulls, their position, relative amplitude, and width varied strongly between species. The position of peaks and nulls changed with fish size but the relative position and width of features did not change with fish size within a species. Few spectral characters were observed below about 100 kHz, regardless of fish size. This indicates that even low-frequency broadband echoes do not have the resolution to

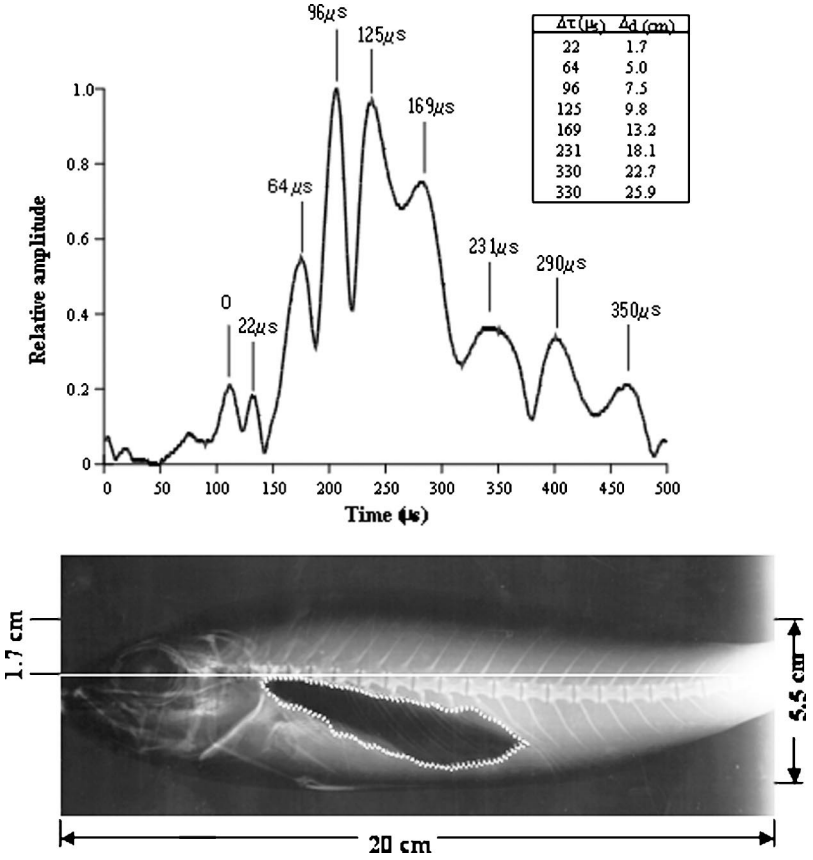


FIG. 7. An expanded plot of the envelope of the cross-correlation function of Fig. 4 for the pink snapper, along with the time delay between each highlight and the first highlight at 0°. The inserted table indicates the delay time and the corresponding two travel distances from the first highlight at 0°. The large differences between the two travel distances and the dimensions of the fish in the x ray indicate a very complex backscattering process.

distinguish differences in spectral structure that are caused by species differences in the shape of the swimbladder, a relatively small structure.

In order to observe the relationship between highlight spacing and the ripple structure in the spectral domain, consider a target that produces an echo, $e(t)$ with n distinct and separable highlights that can be expressed as

$$e(t)=a_0 s(t)+a_1 s(t-\tau_1)+\cdots+a_n s(t-\tau_n), \tag{4}$$

where a_n is the amplitude of the nth highlight and τ_n is the delay time between the nth highlight and the first highlight. The spectrum of Eq. (4) can be expressed as

$$|E(f)|=|S(f)|\{[a_0+a_1\cos(2\pi f\tau_1)+\cdots+\cos(2\pi f\tau_n)]+[a_1\sin(2\pi f\tau_1)+\cdots+\sin(2\pi f\tau_n)]\}^{1/2}. \tag{5}$$

The cos and sin terms are responsible for the ripple pattern and the τ term specifies the position of the nulls in the spectrum. From Eq. (5) we can obtain an insight between the relationship of the spectra shape and highlight intervals.

Fish size is traditionally related to target strength. The relationship between fish length and target strength for the snapper species was not particularly strong (Benoit-Bird *et al.*, 2003a). Fish length was also indicated by the time and frequency characteristics of broadband echoes. The distance between highlights in each species generally increased with fish standard length. The frequency of spectral peaks and nulls generally decreased with increasing fish length. Highlight width was not correlated with fish size. The use of the combination of these factors correlated with fish length, including temporal and spectral structure and target strength, could provide a more accurate estimate of fish length. This combination of factors could also provide an error term for each individual fish instead of one error estimate for all fish measured.

The position of spectral characters was correlated with fish length within each species. Dividing fish length by the frequency of the individual character should remove the effect of length. In other words, the length of the fish divided by the wavelength of the spectral character should be a constant with no correlation with fish length. However, this does not occur for any spectral characteristic in any of the three species. Utilizing the size characteristics of the swimbladder instead of fish length did remove the effect of fish size, however. This indicates the importance of the swimbladder not only in the amplitude of the echo, but in its spectral characteristics. Interestingly, none of the spectral characteristics of the echoes appear to be caused by the remainder of the body of the fish.

The structure of the broadband echoes cannot be solely explained by specular reflections off different parts of the swimbladder. That the echo structure is related to the swimbladder shape can be surmised by considering the shapes shown in Fig. 4 of Benoit-Bird *et al.* (2003a) and the time waveforms in Fig. 2. Each of the snapper species had different swimbladder shapes and, subsequently, differences in the backscatter waveforms. Some of the echoes had durations of over 400 μs, considerably longer than the 80-μs duration of the incident signal. These echoes suggest the presence of some type of high-frequency resonance associated with the backscatter from the swimbladders of snappers, and perhaps

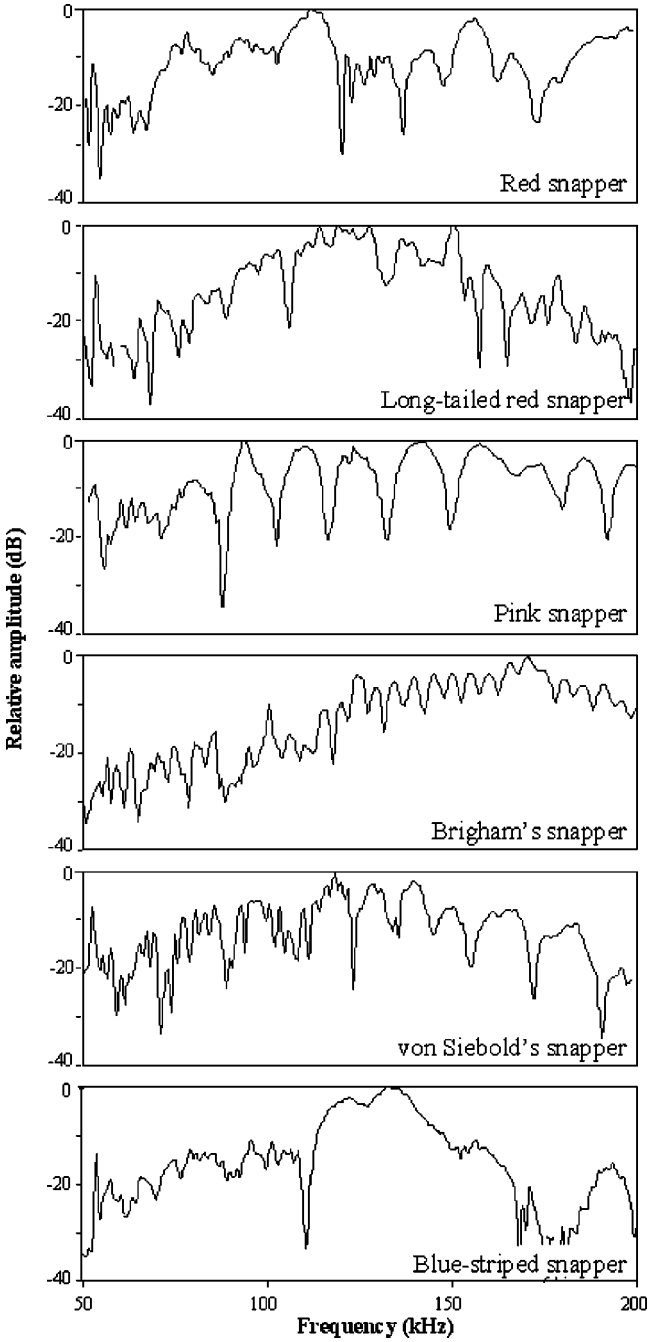


FIG. 8. Representative normalized transfer functions for each of six species of snapper. Echoes are from the broadband, frequency-modulated signal from the dorsal aspect of each fish. Patterns of frequency peaks and nulls were conserved between individuals within each of the three primary species, red, long-tailed red, and pink snapper, regardless of size.

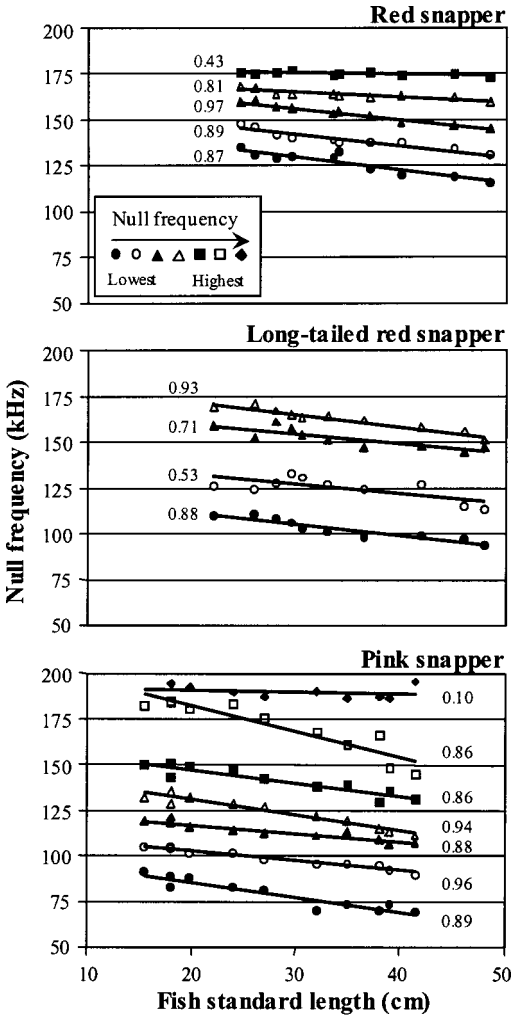


FIG. 9. Frequency of spectral nulls, sharp decreases in amplitude by at least 10 dB, as a function of fish length. Regression coefficients are shown for each relationship. The frequency for each null, except the high-frequency pink snapper null, significantly decreased as fish size increased (F-tests, $p < 0.05$). Similar patterns were observed in frequency peaks.

other types of fishes with swimbladders. Echoes from the lateral aspect with the incident signal being perpendicular to the longitudinal axis of the fishes also have relatively long echoes. Examples of echoes from a lateral aspect taken at the surface and *in situ* at 250-m depth using a manned submersible are reproduced from Benoit-Bird *et al.* (2003b) in Fig. 10. The transmitted signal from the sonar on the submersible was the same as the signal used in this study. These echoes illustrate further the complex backscattering process involved with these lutjanid snappers that are not simply explainable at our current level of understanding. The idea of reflecting waves interacting with various parts of a fish body causing multiple reflections has been considered (Clay, 1991, 1992); however, whether the model used can explain broadband echo durations that are as much as five times longer than the incident signal is questionable. Certainly more physical and mathematical modeling research needs to be done in order to understand the complicated backscatter processes evident in our data.

Fish and their swimbladders are complicated structures that do not lend themselves to simple geometric description. In order to gain a deeper understanding of the acoustic scattering processes in fish, a detailed numerical technique might be necessary. A possible approach is to obtain the three-dimensional geometry of an entire fish body, including the flesh, bones, and swimbladder using x-ray computed tomographic (CT) scans and applying the wave equation to the situation. Jech and Horne (2002) digitized lateral and dorsal radiographs of a fish to construct a three-dimensional representation of the fish body and swimbladder and applied the Kirchhoff-ray mode model (Clay and Horne, 1994) to calculate the acoustic backscatter in three dimensions. The fish

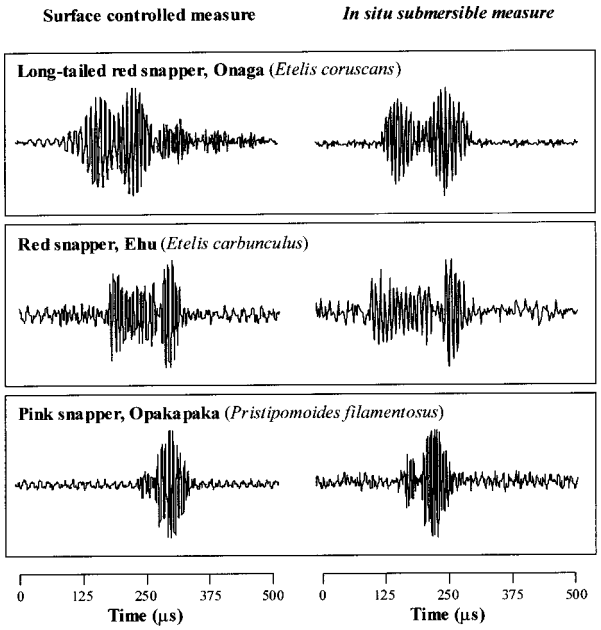


FIG. 10. Representative echoes from the lateral aspect of each Hawaiian snapper species. Echoes on the left were taken from anesthetized fish under controlled conditions at the surface. Echoes on the right were taken from the submersible of free-swimming fish at 250 m.

body and swimbladder were modeled by a series of finite cylinders and the total backscatter was estimated by summing the backscatter overall all the cylinders. Aroyan (2001) has considered the problem of sound propagation in a dolphin's head, modeling the tissue and bones as inhomogeneous fluids so that shear waves could be ignored. He then numerically solved the linearized wave equation using a 3-D finite difference approach. A numerical approach can model the geometry of the backscatter problem more accurately than other approximations using geometry shapes and could also provide a time-history visualization of the scattering process to provide deeper insight into a complex problem. The results obtained by Aroyan were very useful in visualizing in space and time the process of sound propagation and scattering within a complex structure such as a dolphin's head. Whereas Aroyan considered only monocromatic waves, the propagation of a broadband wave such as a pulse would need to be used in order to resolve closely spaced highlights created by fish with many, small internal structures. Much could be learned for such numerical simulation, particularly if species with different swimbladder characteristics are modeled.

ACKNOWLEDGMENTS

Christopher Kelley, Director of the Bottomfish Project, provided support and guidance from the beginning of this work. Aaron Moriwake, Virginia Moriwake, and Bo Alexander maintained the fish while in the hatchery; assisted in the capture of the fish, design of the fish mounting setup, and data collection; and shared their knowledge and expertise. Kristine Hiltunen, a student in the National Science Foundation's Research Experience for Undergraduates Program, assisted in data collection and processing. Kimberly Andrews assisted in data collection. Christopher Kelley provided helpful comments on earlier versions of this manuscript. This work was supported by the Hawaii State Department of Land and Natural Resources. This is HIMB contribution 1167.

Aroyan, J. (**2001**). ''Three dimensional modeling of hearing in *Delphinus delphis*,'' J. Acoust. Soc. Am. **56**, 3305–3318.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).

Au, W. W. L., and Snyder, K. J. (**1980**). ''Long-range target detection in open waters by an echolocating Atlantic bottlenose dolphin,'' J. Acoust. Soc. Am. **56**, 1280–1290.

Barr, R. (**2001**). ''A design study of an acoustic system suitable for differentiating between orange roughy and other New Zealand deep-water species,'' J. Acoust. Soc. Am. **109**, 164–178.

Benoit-Bird, K. J., and Au, W. W. L. (**2002**). ''Enersy: Converting from acoustic to biological resource units,'' J. Acoust. Soc. Am. **96**, 1661–1668.

Benoit-Bird, K. J., Au, W. W. L., and Kelley, C. D. (**2003a**). ''Acoustic backscattering by Hawaiian lutjanid snappers. I. Target strength and swimbladder characteristics,'' J. Acoust. Soc. Am. **114**, 2757–2766.

Benoit-Bird, K. J., Au, W. W. L., Kelley, C. D., and Taylor, C. (**2003b**). ''Acoustic backscattering from deepwater measured in situ from a manned submersible,'' Deep-Sea Res., Part I **50**, 221–229.

Brigham, E. O. (**1988**). *The Fast Fourier Transform and its Applications* (Prentice–Hall, Englewood Cliffs, NJ).

Burdic, W. S. (**1968**). *Radar Signal Analysis* (Prentice–Hall, Englewood Cliffs, NJ).

Clay, C. (**1991**). ''Low-resolution acoustic scattering models: Fluid-filled cylinders and fish with swim bladders,'' J. Acoust. Soc. Am. **89**, 2168–2179.

Clay, C. S. (**1992**). ''Composite ray-mode approximations for backscattered sound from gas-filled cylinders and swimbladders,'' J. Acoust. Soc. Am. **94**, 2173–2180.

Clay, C. S., and Horne, J. K. (**1994**). ''Acoustic models of fish: The Atlantic cod (*Gadus morhua*),'' J. Acoust. Soc. Am. **96**, 1661–1668.

Foote, K. G., and Ona, E. (**1985**). ''Swimbladder cross secitons and acoustic target strengths of 13 pollack and 2 saithe,'' Fiskeridirektoratetes **18**, 1–57.

Greenlaw, C. F., and Johnson, R. K. (**1983**). ''Multiple-frequency acoustical estimation,'' Biol. Oceanogr. **2**, 227–252.

Haight, W. R., Parrish, J. D., and Hayes, T. A. (**1993**). ''Feeding ecology of deepwater lutjanid snappers at Penguin Bank, Hawaii,'' Proc. Am. Fish. Soc. **122**, 328–347.

Jech, J. M., and Horne, J. K. (**2002**). ''Three-dimensional visualization of fish morphometry and acoustic backscatter,'' ARLO **3**, 35–40.

MacLennan, D. N. (**1990**). ''Acoustical measurement of fish abundance,'' J. Acoust. Soc. Am. **87**, 1–15.

MacLennan, D. N., and Holliday, D. V. (**1996**). ''Fisheries and plankton acoustics: past, present, and future,'' ICES J. Mar. Sci. **53**, 513–516.

MacLennan, D. N., and Simmonds, E. J. (**1992**). *Fisheries Acoustics* (Chapman and Hall, New York)

Simmonds, E. J., and MacLennan, D. N. (**1996**). ''Fisheries and Plankton Acoustics,'' paper presented at the ICES Marine Science Symposia, London.

Urick, R. J. (**1983**). *Principles of Underwater Sound* (McGraw–Hill, New York).

Western Pacific Regional Fisheries Management Council (**1999**). ''Bottomfish and Seamount Groundfish of the Western Pacific Region, 1998,'' Annual Report.

Zakharia, M., Magand, F., Hetroit, F., and Diner, N. (**1996**). ''Wideband sounder for fish species identification at sea,'' ICES J. Mar. Sci. **53**, 203–208.

Theory of explosive beam spreading due to ray chaos[a)]

F. D. Tappert
Applied Marine Physics, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149

(Received 5 January 1998; revised 19 June 2003; accepted 28 July 2003)

In an ocean acoustic situation where rays are known to be chaotic, a theory based on a Gaussian beam approximation is used to show that the width of a narrow-angle beam increases exponentially (explosively) with range at a rate that is given by the Lyapunov exponent that is calculated in the geometric limit. This finite frequency result shows that the phenomenon of ''wave chaos'' is real, and might be observable experimentally. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1612496]

PACS numbers: 43.30.Qd, 43.30.Ft, 43.30.Dr [DLB] Pages: 2775–2781

I. INTRODUCTION

The phenomenon of ray chaos in ocean acoustics has been studied for many years.[1–6] Although the reality of this phenomenon in ray-based acoustic models has been well established, the issue remains whether the exponential sensitivity of ray trajectories has a finite-frequency counterpart that can be observed experimentally. To date, numerical simulations with full-wave acoustic models have not revealed a clear exponential increase of some quantity, such as the complexity of the wave field, that corresponds to ray chaos. In this paper, theoretical calculations of the spreading of a Gaussian beam in a random medium are used to demonstrate that the main effect of ray chaos, namely, exponential spreading, persists at finite frequency as a transient effect. This transient exponential spreading saturates at long range, but may nevertheless be observed experimentally under carefully controlled conditions.

In this paper, the acoustic model pertains to underwater sound propagation in the two-dimensional horizontal plane with mesoscale structure modeled as a random function of x and y.[5,6] The corresponding vertical plane problem is also known to exhibit ray chaos.[7] The main results found here apply also to sound propagation in the vertical plane if account is taken of the effects of a background sound speed profile and the anisotropy of the mesoscale structure in the vertical plane. The horizontal plane problem is preferred because it is simpler, and because the Lyapunov exponent for this problem has recently been calculated analytically.[5,6]

In Sec. II, the acoustic model and the Gaussian beam initial condition is introduced, and properties of the solution, such as the rms beam width and the rms angular width, are defined. Then in Sec. III, the geometric approximation to the solution is derived in detail. This solution exhibits ray chaos, and the Lyapunov exponent that characterizes the chaos is obtained. Although there are several ways to develop the connection between ray and wave theory, we choose to establish this connection by revisiting an exact hybrid PE-ray model called HYPER.[8,9] This model is formally derived in Sec. IV. A full-wave approximation to this hybrid model, called the Gaussian beam approximation, is obtained in Sec. V; the exact solution of this approximate wave equation is derived and its properties are studied. The most important property derived in Sec. V is the exponential (explosive) increase of the beam width with increasing range that is due to ray chaos. The results are discussed in Sec. VI.

II. ACOUSTIC MODEL

In Refs. 5 and 6, a parabolic wave equation[10] is derived that describes underwater sound propagation in the horizontal plane at global ranges including the effects of ocean mesoscale structure. Mode coupling is neglected, and the basic equation that determines the reduced acoustic pressure ψ is the following:

$$ik_0^{-1}\frac{\partial\psi}{\partial x}=-\frac{k_0^{-2}}{2}\frac{\partial^2\psi}{\partial y^2}+\varepsilon\mu(x,y)\psi, \tag{1}$$

with the boundary condition

$$\lim_{|y|\to\infty}\psi(x,y)=0. \tag{2}$$

Here x is range and y is cross range. The mesoscale structure near the axis of the sound channel is modeled by the dimensionless function $\mu(x,y)$ and its strength is measured by the dimensionless parameter ε. It is assumed that the potential $\mu(x,y)$ has unit-variance and zero-mean, is spatially stationary, isotropic, and has a single scale length denoted by L. Typical numerical values are $L\sim 100$ km, and $10^{-3}\lesssim\varepsilon\lesssim 10^{-2}$. The correlation function of the mesoscale structure is related to its power spectrum by

$$\langle\mu(x,y)\mu(x+\overline{x},y+\overline{y})\rangle = \int\int_{-\infty}^{\infty}W(k,l)\exp[i(k\overline{x}+l\overline{y})]dk\,dl. \tag{3}$$

Following Refs. 5 and 6, the spectrum is assumed to be regular, i.e., not a power-law spectrum, as is typically observed in the real ocean. We have seen from simulations through a power law media that the behavior described in Refs. 5 and 6 still holds.

[a)]This manuscript was completed posthumously by Michael A. Wolfson, and care was taken to complete it in a manner consistent with Professor Tappert's passionate view regarding the impact of ''ray chaos'' on the fundamental study of wave propagation in random media.

The initial condition used in this paper, in contrast to the plane wave initial condition in Refs. 5 and 6 is a Gaussian beam centered at $y=y_0$ that has width w_0 and launch angle p_0:

$$\psi(y,0)=\exp[-(y-y_0)^2/4w_0^2]\exp[ik_0p_0(y-y_0)]. \quad (4)$$

Here, p_0 represents the angle with respect to the x axis of the direction of the Gaussian beam. As a consequence of the assumption of horizontal isotropy of the mesoscale structure $\mu(x,y)$, if $|p_0|\ll 1$ there is no loss of generality in taking $p_0=0$, and that is done in the remainder of this paper. Likewise, as a consequence of the assumption of cross-range stationarity of $\mu(x,y)$, there is no loss of generality in taking $y_0=0$, and that is done also. Quite generally, the power in the beam is conserved:

$$P=\int |\psi(y,x)|^2dy=\text{const}. \quad (5)$$

The centroid of the beam at range x is defined as

$$y_c(x)=P^{-1}\int y|\psi(y,x)|^2dy, \quad (6)$$

and the rms width of the beam at range x is

$$w(x)=\left[P^{-1}\int [y-y_c(x)]^2|\psi(y,x)|^2dy\right]^{1/2}. \quad (7)$$

Initially, it is found that $y_c(0)=y_0=0$ and $w(0)=w_0$. To obtain the angle spread of the beam, the Fourier transformation from y to θ is defined as

$$\hat{\psi}(\theta,x)=\int \psi(y,x)e^{-ik_0\theta[y-y_c(x)]}dy. \quad (8)$$

After setting

$$\hat{P}=\int |\hat{\psi}(\theta,x)|^2d\theta=(2\pi/k_0)P=\text{const}, \quad (9)$$

the center angle is defined as

$$\theta_c(x)=\hat{P}^{-1}\int \theta|\hat{\psi}(\theta,x)|^2d\theta, \quad (10)$$

and the rms angle spread at range x is

$$\vartheta(x)=\left[\hat{P}^{-1}\int [\theta-\theta_c(x)]^2|\hat{\psi}(\theta,x)|^2d\theta\right]^{1/2}. \quad (11)$$

Initially, it is found that $\theta_c(0)=p_0=0$, and $\vartheta(0)=1/2k_0w_0$. With this notation, the uncertainty principle for beams (actually a mathematical theorem) may be stated as

$$w(x)\vartheta(x)\geqslant(2k_0)^{-1}. \quad (12)$$

Initially, equality is found to hold because the Gaussian starting field is known to have the minimum uncertainty.

It happens that the Gaussian function in Eq. (4) is the traditional starting field for the PE acoustic model.[10] For modeling omnidirectional sources, it was replaced in the early 1980s by high-angle starting fields. However, for modeling highly directional sources this starting field has been retained in some PE models. In any case, the solution of Eq. (1) for $\varepsilon=0$ (and $y_0=p_0=0$) is known to be[10]

$$\psi(y,x)=w_0(w_0^2+ix/2k_0)^{-1/2}\times\exp[-y^2/4(w_0^2+ix/2k_0)]. \quad (13)$$

Then it is found that

$$w(x)=w_0(1+x^2/4k_0^2w_0^4)^{1/2}, \quad (14)$$

and $\vartheta(x)=\text{const}=1/2k_0w_0$. This result displays the diffractive spreading of the beam, and it shows that the transition from the near-field to the far-field zone (Fraunhofer zone), where the beam width on average increases linearly with range, occurs at the range $x_f\sim k_0w_0^2$. Even when $\varepsilon\neq 0$, the same transition to the far-field zone takes place at the same relatively small range, as shown in the following.

In the following, it is assumed that $k_0w_0\gg 1$ so that the rms angle of the beam is initially small: $\vartheta(0)=1/2k_0w_0\ll 1$. It is also assumed that $w_0/L\ll 1$, where L is the horizontal length scale of $\mu(x,y)$. As is shown in Sec. V, this assumption allows many e-folding growths of the beam width before saturation. In principle, this initial condition could be realized experimentally with a long horizontal array that is shaded to reduce sidelobe levels. It is not necessary that the beam pattern be exactly Gaussian; that particular function is used to make closed-form analytical solutions possible.

III. GEOMETRIC APPROXIMATION

In the geometric approximation, the Gaussian beam is represented by a single ray. The initial condition for this ray has both a definite position, $y=y_0=0$, and a definite momentum, $p=p_0=0$. To derive the appropriate equations, which are also needed in the following sections, a solution of Eq. (1) is sought in the form

$$\psi_g=a\exp(ik_0s). \quad (15)$$

This form is put into Eq. (1), and the resulting equation is expanded in powers of k_0^{-1}. To leading order, the parabolic eikonal equation is obtained:

$$\frac{\partial s}{\partial x}+\frac{1}{2}\left(\frac{\partial s}{\partial y}\right)^2+\varepsilon\mu(x,y)=0. \quad (16)$$

Next let $p=\partial s/\partial y$, and take the partial derivative of the parabolic eikonal equation with respect to y to obtain

$$\frac{\partial p}{\partial x}+p\frac{\partial p}{\partial y}=-\varepsilon\mu_y(x,y), \quad (17)$$

where the subscript y denotes the partial derivative with respect to y. This nonlinear equation is solved by the method of characteristics to obtain the parabolic ray equations in Hamiltonian form:

$$\dot{y}=p, \quad \dot{p}=-\varepsilon\mu_y(x,y), \quad (18)$$

where an overdot denotes the total derivative with respect to range x. The cross-range variable is y, and the variable p is interpreted as the angle of a ray with respect to the x axis, and is necessarily small due to the parabolic approximation.[10] The initial condition is expressed as $y(0)=y_0=0$, $p(0)=p_0=0$. The solutions of the ray equations with these initial conditions are denoted as $y_r(x)$ and $p_r(x)$.

Here and in the following, the subscript r distinguishes a ray quantity from a general coordinate.

From the relation

$$ds = \frac{\partial s}{\partial x}dx + \frac{\partial s}{\partial y}dy, \tag{19}$$

it is easily shown that

$$ds = (p^2/2 - \varepsilon\mu)dx. \tag{20}$$

Thus the eikonal function along the ray is found to be

$$s_r(x) = \int_0^x \left[\frac{1}{2}p_r^2(x') - \varepsilon\mu(x', y_r(x'))\right]dx'. \tag{21}$$

The travel time along the ray is then

$$t_r(x) = [x + s_r(x)]/c_0. \tag{22}$$

The variational equations that correspond to this ray dynamical system describe the local behavior of rays that remain infinitesimally close to a chosen reference ray. A linearization of the dynamics about a reference ray yields

$$\begin{pmatrix} \delta p_r \\ \delta y_r \end{pmatrix} = J \begin{pmatrix} \delta p_0 \\ \delta y_0 \end{pmatrix}, \tag{23}$$

with the stability (Jacobi) matrix being given by the partial derivatives

$$J = \begin{pmatrix} \eta_1 & \xi_1 \\ \eta_2 & \xi_2 \end{pmatrix} = \begin{pmatrix} \left.\frac{\partial p_r}{\partial p_0}\right|_{y_0} & \left.\frac{\partial p_r}{\partial y_0}\right|_{p_0} \\ \left.\frac{\partial y_r}{\partial p_0}\right|_{y_0} & \left.\frac{\partial y_r}{\partial y_0}\right|_{p_0} \end{pmatrix}. \tag{24}$$

The evolution of J is governed by the set of variational equations

$$\frac{d}{dr}J = KJ, \tag{25}$$

with the initial condition $J(x=0)$ being the identity matrix, and

$$K = \begin{pmatrix} 0 & -\frac{\partial^2\mu}{\partial y^2} \\ 1 & 0 \end{pmatrix}. \tag{26}$$

The determinant of the Jacobi matrix, $\det J$, is found to be unity, and this is an expression of Liouville's theorem that states that the phase-space area for the Hamiltonian dynamical system in Eq. (18) is conserved:

$$dy_r dp_r = dy_0 dp_0. \tag{27}$$

Returning to the geometric approximation, it is found to next order in the k_0^{-1} expansion that the ray amplitude a satisfies the conservation equation

$$\frac{\partial a^2}{\partial x} + \frac{\partial}{\partial y}(pa^2) = 0. \tag{28}$$

In characteristic coordinates, this becomes

$$\frac{da^2}{dx} + a^2\frac{\partial p}{\partial y} = 0. \tag{29}$$

Using the relations

$$\frac{\partial p}{\partial y} = \frac{\partial p}{\partial y_0}\Big/\frac{\partial y}{\partial y_0} = \eta_2(x)/\xi_2(x) = \frac{d}{dx}\ln\xi_2(x), \tag{30}$$

it follows that the amplitude of the ray is given by

$$a_r(x) = [\xi_2(x)]^{-1/2}. \tag{31}$$

The correct branch of this square-root function is determined by the well-known Keller–Maslov index. Ranges at which $\xi_2(x) = 0$ are caustic contact points along the ray, and the ray amplitude $a_r(x)$ diverges at these points. The full geometric approximation to ψ along the ray at the center of the beam is

$$\psi_g = [\xi_2(x)]^{-1/2}\exp[ik_0 s_r(x)]. \tag{32}$$

In order to calculate $s_r(x)$ and $\xi_2(x)$, the ray trajectories $y_r(x)$ and $p_r(x)$ must first be calculated. Thus this expression is actually quite formidable for this chaotic ray situation.

In Refs. 5, 6, and 11, a statistical analysis of the above-given ray equations has been made. It was shown that, since $\mu(x,y)$ has no symmetries or periodicities whatsoever, the dynamical system described by Eq. (18) is "completely chaotic" in the sense of Percival[12] and Chirikov.[13] It is believed that this system is an Anosov dynamical system—ergodic and strongly mixing. The main results of the analysis can be summarized as follows: It is found that $p_r(x)$ and $y_r(x)$ are Gaussian random variables. The mean ray angle is $\langle p_r(x)\rangle = p_0 = 0$, and the standard deviation of the ray angle angle spread is $\sigma_p(x) = \varepsilon(Dx)^{1/2}$, where the momentum diffusion coefficient is found in terms of the spectrum of the mesoscale fluctuations to be

$$D = \pi\int_0^\infty l^2 W(0,l)dl. \tag{33}$$

In order of magnitude, $D \sim 1/L$, and thus $\sigma_p(x) \sim \varepsilon(x/L)^{1/2}$. The mean ray position is $\langle y_r(x)\rangle = y_0 = 0$, and the standard deviation of the ray position spread is $\sigma_y(x) = \varepsilon D^{1/2}x^{3/2}/\sqrt{3}$. In order of magnitude, $\sigma_y(x) \sim \varepsilon L(x/L)^{3/2}$. Although $s_r(x)$ is not a Gaussian random variable, its mean value is found to be $\langle s_r(x)\rangle = \varepsilon^2 Dx^2/4$. In order of magnitude, $\langle s_r(x)\rangle \sim \varepsilon^2 L(x/L)^2$. From Eq. (22), it follows that the ray travel time bias is positive (delay), and is given by $\langle t_r(x)\rangle \sim [x + \varepsilon^2 L(x/L)^2]/c_0$. These results show that ray chaos cannot be discerned at the level of one-ray statistics.

The important statistical results this paper is concerned with have been derived in Refs. 5, 6, and 14, and concern the four variational quantities: $\eta_1(x), \xi_1(x), \eta_2(x), \xi_2(x)$. It is found that all four quantities are lognormal random variables, and that their mean values are the same as their initial values: $\langle\eta_1(x)\rangle = 1$, $\langle\xi_1(x)\rangle = 0$, $\langle\eta_2(x)\rangle = 0$, $\langle\xi_2(x)\rangle = 1$. The variance of $\xi_2(x)$ is found to be

$$\langle\xi_2^2(x)\rangle = [\exp(2\nu_0 x) + 2\exp(-\nu_0 x)\cos(\sqrt{3}\nu_0 x)]/3. \tag{34}$$

Here ν_0 is the mean Lyapunov exponent that is given in terms of the spectrum of mesoscale fluctuations as

$$\nu_0 = \varepsilon^{2/3}(\alpha_0/2)^{1/3}, \quad \alpha_0 = \pi\int_0^\infty l^4 W(0,l)dl. \tag{35}$$

In order of magnitude, $\alpha_0 \sim 1/L^3$, and thus $\nu_0 \sim \varepsilon^{2/3}/L$. The inverse Lyapunov exponent is the *e*-folding range, and is also the order of magnitude of the "predictability horizon," called x_p, that is familiar from ray chaos theory. This is the expected range where one loses the ability to predict with any reasonable accuracy the location and reception angle of a ray. For our system it is $x_p \sim L\varepsilon^{-2/3}$. Further, it is found that

$$\langle \eta_1^2(x) \rangle = \langle \xi_2^2(x) \rangle, \tag{36}$$

$$\langle \eta_2^2(x) \rangle = (2\nu_0^2/3)[\exp(2\nu_0 x) - \exp(-\nu_0 x) \times \{\cos(\sqrt{3}\nu_0 x) - \sqrt{3}\sin(\sqrt{3}\nu_0 x)\}], \tag{37}$$

$$\langle \xi_1^2(x) \rangle = (1/6\nu_0^2)[\exp(2\nu_0 x) - \exp(-\nu_0 x) \times \{\cos(\sqrt{3}\nu_0 x) + \sqrt{3}\sin(\sqrt{3}\nu_0 x)\}]. \tag{38}$$

Thus all of the variational quantities increase exponentially, on average, if $\nu_0 x \sim 1$, or $x \sim x_p$. This exponential effect holds for all initial values y_0 and p_0, and that is why this system of ray equations is called completely chaotic.

Since $\xi_2(x) = \partial y/\partial y_0$, the rms separation between two rays at range x that are initially separated by the infinitesimal amount dy_0 is

$$dy(x) = \xi_2(x) dy_0. \tag{39}$$

Roughly speaking, $|dy|$ can be taken to be the rms width of the Gaussian beam in the geometric limit, as long as the width remains small enough for the analysis to remain valid, i.e., $|dy| \ll L$. The classical rms beam width is then

$$w_g(x) = w_0 |\xi_2(x)|. \tag{40}$$

Thus Eq. (34) demonstrates that the rms value of $w_g(x)$ grows exponentially with range. The exponential divergence between neighboring rays is the distinguishing feature of ray chaos. This effect anticipates the main result of this paper, namely, that the width of a Gaussian beam diverges exponentially on average even at finite frequency.

IV. HYBRID PE-RAY THEORY

A hybrid model that combines the parabolic equation (PE) model and the corresponding ray model is called HYPER,[8,9] which is an acronym for "Hybrid PE-Ray." The original motivation for developing the HYPER model was to extend the PE model to higher acoustic frequencies by using a ray-based coordinate system that facilitates numerical calculations.[15] It was also appreciated[9] that the HYPER model could be used as the analytic basis for a Gaussian beam approximation method that is discussed in the next section. See also Ref. 16, Sec. 3.7, p. 191, for a brief description of the HYPER model. The transformations that lead from the parabolic wave equation, Eq. (1), to the HYPER wave equation, Eq. (46), are exact.

Starting from Eq. (1), the solution is represented as

$$\psi(y,x) = \phi(y,x)\exp[ik_0 S(y,x)], \tag{41}$$

where $S(y,x)$ is the Maslov phase function:

$$S(y,x) = s_r(x) + p_r(x)[y - y_r(x)]. \tag{42}$$

Substitution of Eq. (41) into Eq. (1) yields the following equation for $\phi(y,x)$:

$$ik_0^{-1}\left(\frac{\partial \phi}{\partial x} + p\frac{\partial \phi}{\partial y}\right) = -\frac{k_0^{-2}}{2}\frac{\partial^2 \phi}{\partial y^2} + \left[\frac{\partial S}{\partial x} + \frac{1}{2}\left(\frac{\partial S}{\partial y}\right)^2 + \varepsilon\mu(x,y)\right]\phi. \tag{43}$$

Differentiation of Eq. (42) yields

$$\frac{\partial S}{\partial x} + \frac{1}{2}\left(\frac{\partial S}{\partial y}\right)^2 = -\varepsilon[\mu(x,y_r(x)) + (y - y_r(x))\mu_y(x,y_r(x))]. \tag{44}$$

Next the variable y is replaced by the local coordinate

$$z = y - y_r(x). \tag{45}$$

The HYPER wave equation then takes the form of the original parabolic wave equation in Eq. (1):

$$ik_0^{-1}\frac{\partial \phi}{\partial x} = -\frac{k_0^{-2}}{2}\frac{\partial^2 \phi}{\partial z^2} + \varepsilon u(x,z)\phi, \tag{46}$$

where the potential in these new variables is

$$u(x,z) = \mu(x,y_r(x)+z) - \mu(x,y_r(x)) - z\mu_y(x,y_r(x)). \tag{47}$$

The initial condition obtained from Eq. (4) is

$$\phi(z,0) = \exp(-z^2/4w_0^2). \tag{48}$$

Since $|\psi|^2 = |\phi|^2$, the power in the beam is

$$P = \int |\phi(z,x)|^2 dz, \tag{49}$$

and unitarity of Eq. (46) shows that P is indeed constant.

Since the HYPER wave equation is exactly equivalent to the parabolic wave equation, no information has been lost. For analytical purposes nothing has been gained because Eq. (46) is just as intractable as Eq. (1). For numerical purposes, however, the HYPER wave equation has the advantage at high acoustic frequencies that use of the local coordinate z allows fewer mesh points in both the range and cross-range variables, and thus provides greater numerical efficiency. This advantage extends to the vertical plane problem and to omnidirectional sources, that are decomposed into a large number of Gaussian beams.[15]

In range-dependent (laterally inhomogeneous) environments such as the one considered in this paper, the underlying ray trajectories are generically chaotic, and this fact has severe consequences for numerical calculations that use ray-based models such as the HYPER model. The ray trajectory $y_r(x)$, for example, is extremely sensitive to the initial value y_0. Thus a small measurement error in y_0 is amplified by an exponentially large factor, $\exp(\nu_0 x)$, at the predictability horizon range $x_p = O(\nu_0^{-1})$. Therefore the HYPER wave equation is used only for theoretical analysis in Sec. V, and no attempt is made to obtain numerical solutions, nor is such advocated.

V. GAUSSIAN BEAM APPROXIMATION

The Gaussian beam technique has a long history of application in seismology, where it was first developed, and it was introduced into underwater acoustics by Porter and Bucker,[17] whose paper contains numerous citations to previous work in seismology. In essence, the Gaussian beam technique uses a local parabolic approximation along a ray. If a sufficiently large number of rays are used to represent an omnidirectional source, then in principle an accurate solution to the Helmholtz equation can be obtained.[17] A known deficiency of the Gaussian beam technique is that diffractive leakage from a duct (quantum mechanical tunneling) is not included. An even greater deficiency, in the author's opinion, is that the phenomenon of ray chaos prevents accurate prediction of ray trajectories at sufficiently long ranges. This last deficiency is shared by all ray-based acoustic models.

In the present context, the Gaussian beam approximation replaces the potential $u(x,z)$, defined in the previous section in Eq. (47), with the first term of the expansion in powers of z, which is a quadratic function:

$$u(x,z) \approx \tfrac{1}{2} z^2 \mu_{yy}(x, y_r(x)). \tag{50}$$

Note the term μ_{yy}—relating to the change in cross-range curvature of sound speed—is the same term that played a key role in the variational analysis of Sec. III. Then the exact HYPER wave equation becomes the approximate Gaussian beam wave equation:

$$ik_0^{-1} \frac{\partial \phi_b}{\partial x} = -\frac{k_0^{-2}}{2} \frac{\partial^2 \phi_b}{\partial z^2} + \frac{1}{2} \varepsilon z^2 \mu_{yy} \phi_b . \tag{51}$$

The initial condition for ϕ_b is the same as above in Eq. (48), and the beam power given by Eq. (49) is conserved. This equation is solved in the following to obtain the beam width $w_b(x)$. The condition for the Gaussian beam approximation to be valid is that $w_b(x) \ll L$, and this approximation breaks down if $w_b(x) \sim L$. Stated crudely, rays making up the Gaussian beam become uncorrelated when this condition is not satisfied. With this approximation, the solution of Eq. (1) becomes

$$\psi(y,x) \approx \phi_b(z,x) \exp[ik_0(s_r(x) + z p_r(x))]. \tag{52}$$

The parabolic wave equation in the Gaussian beam approximation, Eq. (51), has the same mathematical form as the time-dependent Schrödinger equation for a harmonic oscillator that has a random (zero-mean) time-dependent "spring constant." Although the solution could be taken from the quantum mechanical literature, it is more appropriate to derive the solution here. The solution is known to have the form of a Gaussian function if the initial condition is Gaussian. Thus a solution of Eq. (51) is sought that has the form

$$\phi_b(z,x) = A(x) \exp[-B(x) z^2]. \tag{53}$$

Substitution of this expression into Eq. (51) yields

$$\frac{dA}{dx} = -iA(x)B(x)/k_0, \tag{54}$$

$$\frac{dB}{dx} = i[-2B^2(x)/k_0 + \varepsilon k_0 \mu_{yy}/2]. \tag{55}$$

Equation (55) contains $B(x)$ only, and has the form of a Ricatti equation. This equation can be solved by setting

$$B(x) = -(ik_0/2) \frac{d}{dx} \ln q(x). \tag{56}$$

Then $q(x)$ satisfies the linear equation

$$\ddot{q} = -\varepsilon \mu_{yy} q. \tag{57}$$

This is the same as one of the ray variational equations in Eq. (25). The fact that the solution of the Gaussian beam equation involves solutions of the ray variational equation is not new.[17] It is also known that solutions of the ray variational equation tend to increase exponentially in a chaotic situation, as was discussed in Sec. III. The combination of these two known results provides the insight that leads to the main result of this paper.

The solution of Eq. (57) is a linear combination of $\xi_1(x)$ and $\xi_2(x)$. Use of the initial condition $B(0) = 1/4w_0^2$ then yields

$$q(x) = \text{const}[\xi_2(x) + i\beta \xi_1(x)], \tag{58}$$

where the constant parameter β is

$$\beta = (2k_0 w_0^2)^{-1}. \tag{59}$$

The solution for $B(x)$ is therefore

$$B(x) = (2w_0)^{-2} \left[\frac{\eta_1(x) - i\beta^{-1} \eta_2(x)}{\xi_2(x) + i\beta \xi_1(x)} \right]. \tag{60}$$

The amplitude function $A(x)$ is then found to be

$$A(x) = [\xi_2(x) + i\beta \xi_1(x)]^{-1/2}. \tag{61}$$

The solution of the Gaussian beam wave equation has been obtained, and the approximate solution of Eq. (1) is

$$\psi(y,x) \approx A(x) \exp[ik_0 S(y,x) - B(x)(y - y_r(x))^2]. \tag{62}$$

Clearly, at $x = 0$ this expression reduces to the initial condition specified in Eq. (4).

Before this Gaussian beam solution is fully interpreted and explained, it is examined in the near-field $x \ll x_f \sim \beta^{-1}$. Then $\xi_1(x) \approx x$, $\eta_1(x) \approx 1$, $\xi_2(x) \approx 1$, and $\eta_2(x) \approx 0$. It is easily seen that the above-given expressions for $B(x)$ and $A(x)$ simplify, and that the resulting Gaussian beam solution reduces to the solution found in Sec. II, Eq. (13), in the limit $\varepsilon = 0$. Thus the Gaussian beam solution is exact in this limit, and it correctly describes diffractive spreading of the beam. Another limit that is easily examined is the geometric limit, $k_0^{-1} \to 0$, or $\beta \to 0$. Then $A(x) \to a_r(x)$, where $a_r(x)$ is the ray amplitude specified in Eq. (31), and at the center of the beam, $y = y_r(x)$, Eq. (62) reduces to the geometric solution specified in Eq. (32). Note that the Gaussian beam solution is not singular at the caustic contact points where $\xi_2(x) = 0$ because the interleaving property of the two independent solutions of the linear variational equations prevents $\xi_1(x)$ from vanishing at the same points. This is a good reason for preferring the Gaussian beam approximation over the geometric approximation.[17]

Next the intensity of the Gaussian beam solution is calculated. Using the fact that the determinant of the Jacobi matrix is unity, it can be shown that the real part of $B(x)$ is

$$\Re B(x) = (2w_0)^{-2}[\xi_2^2(x) + \beta^2 \xi_1^2(x)]^{-1}. \tag{63}$$

Thus it is found that

$$|\phi_b(z,x)|^2 = [w_0/w_b(x)]^{-1} \exp[-z^2/2w_b^2(x)], \tag{64}$$

where the width of the Gaussian beam is

$$w_b(x) = w_0[\xi_2^2(x) + \beta^2 \xi_1^2(x)]^{-1/2}. \tag{65}$$

This solution obviously conserves the beam power, and $w_b(x)$ reduces to $w_g(x)$, see Eq. (40), in the geometric limit $k_0^{-1} \to 0$, or $\beta \to 0$.

The simple appearing expression in Eq. (64) is actually quite complicated. The center of the beam is at $z=0$, or $y = y_r(x)$. Since the ray trajectory $y_r(x)$ is chaotic, it is effectively a random variable and is unpredictable at long ranges due to its extreme sensitivity to measurement errors of the initial value y_0. In Sec. III, it was shown that the center of the beam has mean $\langle y_r(x) \rangle = y_0$, and standard deviation $\sigma_y \sim \varepsilon L (x/L)^{3/2}$. Thus at the predictability horizon, $x \sim x_p \sim L\varepsilon^{-2/3}$, it is seen that $\sigma_y \sim L$ and the center of the beam has wandered to the length scale of the mesoscale structure. Nevertheless, the statistical uncertainty of the center of the beam increases with range only as a power law, and not exponentially.

The rms width of the beam $w_b(x)$ is also a random variable since both $\xi_1(x)$ and $\xi_2(x)$ are random variables. In the near field, $\xi_1(x) \approx x$ and $\xi_2(x) \approx 1$, and then $w_b(x)$ exhibits diffractive spreading according to Eq. (14) in Sec. II. According to the ray chaos theory presented in Sec. III, if $\nu_0 x \sim 1$ then $w_b(x)$ increases exponentially with range:

$$\langle w_b^2(x) \rangle^{1/2} \approx (w_0/\sqrt{3})(1+\gamma^2)^{1/2} \exp(\nu_0 x), \tag{66}$$

where γ is the dimensionless parameter

$$\gamma = \sqrt{2}/(4k_0 w_0^2 \nu_0) \sim (L/k_0 w_0^2)\varepsilon^{-2/3}. \tag{67}$$

This predicted exponential spreading of the beam is called "explosive beam spreading," and it is the main result of this paper.

The Lyapunov exponent ν_0 that appears in Eq. (66) does not depend on the acoustic frequency, and it characterizes ray chaos for this problem. The parameter γ that appears in Eq. (66) is inversely proportional to the acoustic frequency, and it characterizes the finite frequency effect on beam spreading. The beam spreading result in Eq. (66) shows that the spreading is exponentially fast in range whatever the frequency, and in fact it is larger at finite frequencies than in the geometrical limit where $\gamma \to 0$. This is due to the more rapid diffractive spreading of a beam at finite frequency that was found when $\varepsilon = 0$ in Eq. (14) in Sec. II. In summary, a finite frequency counterpart of ray chaos has been discovered that rightfully deserves to be called "wave chaos."

Since the beam width is a random variable and ν_0 is the mean Lyapunov exponent, there is actually considerable variability in the growth with range of the beam width. For some values of y_0 and p_0, $w_b(x)$ increases more rapidly than its rms value in Eq. (66), and such center rays are called "super-chaotic," because their Lyapunov exponent is greater than its mean value. For other values of y_0 and p_0, $w_b(x)$ increases less rapidly than its rms value, and such center rays are called "sticky," because their Lyapunov exponent is less than its mean value and neighboring rays tend to stick together more than the average. The concept of "stickiness" has been discussed, for example, by Chirikov.[13]

Since the width of the Gaussian beam tends to increase exponentially at long range, it is obvious that the intensity on the axis of the beam at $z=0$ tends to decrease exponentially at long range. Indeed, from Eq. (64) it follows that

$$\langle |\phi_b(0,x)|^2 \rangle = \langle [w_0/w_b(x)]^{-1} \rangle \sim \exp(-\nu_0 x). \tag{68}$$

This estimate, however, is not firmly based on statistical analysis. Instead, it is based on power conservation:

$$P \sim \langle |\phi_b(0,x)|^2 \rangle \langle w_b^2(x) \rangle^{1/2}/w_0 = \text{const}. \tag{69}$$

More statistical analysis needs to be done on the beam intensity and on the intensity fluctuations.

At the beginning of this section it was stated that the Gaussian beam approximation breaks down when $w_b(x) \sim L$. From Eq. (66), the breakdown condition on average is $w_0 \exp(\nu_0 x_b) \sim L$, or

$$x_b \sim \nu_0^{-1} \ln(L/w_0) \sim x_p \ln(L/w_0). \tag{70}$$

Thus in order to observe many e-folding growths of the beam width before the Gaussian beam approximation breaks down, it is necessary that $w_0 \ll L$. This assumption was stated at the end of Sec. II. What happens beyond the range x_b where the Gaussian beam approximation breaks down? The HYPER wave equation, Eq. (46), holds the answer to this question, but it is too difficult to solve analytically. It is reasonable to suppose that $w(x)$ does not continue to grow exponentially, and instead it saturates to values that grow more slowly, probably proportional to the rms beam center growth that goes as $\sigma_y(x) \sim \varepsilon L (x/L)^{3/2}$. Thus the exponential growth of the beam width is a transient phenomenon that saturates at long range and becomes a power-law growth.

To experimentally test the above-discussed theory, one could set up a line array of acoustic sources on the axis of the sound channel, each transmitting at 100 Hz and propagate an approximate Gaussian beam of width 1 km. If mesoscale activity were the dominant sound speed fluctuations ($\varepsilon = 5 \times 10^{-3}, L = 100$ km), then there would be a range domain where the exponential growth of the beam width would overtake the linear growth predicted by diffraction theory. From Eq. (14), linear growth of the beam width would occur at ranges significantly greater than 1000 km, and exponential growth would occur at ranges of $O(2500\ \text{km})$. From Eq. (70), $x_b = O(5\nu_0^{-1})$, so with $\nu_0 = O(2500\ \text{km})$, the saturation range is beyond the scale of the ocean basin. The receiving array location would have to consider the randomness with range of the mean position of the beam.

Lastly, for the sake of completeness, the results in angle space are presented. The Gaussian beam solution given in Eq. (62) is easily Fourier transformed according to Eq. (8) to obtain $\hat{\psi}_b(\theta,x)$, and then $|\hat{\psi}_b(\theta,x)|^2$ is calculated. The center angle of the beam is found to be $\theta_c(x) = p_r(x) = \dot{y}_r(x)$, as expected. From the results in Sec. III, the mean center angle

is $\langle p_r(x)\rangle = p_0 = 0$, and the rms center angle is $\sigma_p(x) = \varepsilon(x/L)^{1/2}$. At the predictability horizon, the rms center angle is of order $\varepsilon^{2/3}$. This rms angle is small, which is consistent with the parabolic approximation. The spread in beam angle is found to be

$$\vartheta_b(x) = (2k_0 w_0)^{-1}[\eta_1^2(x) + 4k_0^2 w_0^4 \eta_2^2(x)]^{1/2}. \quad (71)$$

The high frequency limit, $k_0^{-1} \to 0$, of this expression is

$$\vartheta_b(x) \to w_0 |\eta_2(x)|, \quad (72)$$

which is the result of geometric acoustics since one recalls that $\eta_2(x) = \partial p_r(x)/\partial y_0$. From the results in Sec. III, if $x \sim x_p$ the rms value of the angle spread is exponentially large:

$$\langle \vartheta_b(x)\rangle^{1/2} \approx (2k_0 w_0 \sqrt{3})^{-1}(1+\gamma^{-2})^{1/2} \exp(\nu_0 x). \quad (73)$$

Thus the angle spread, like the beam width, exhibits a transient exponential growth at the rate of the Lyapunov exponent that saturates at long range. Again, this finite-frequency counterpart of ray chaos may be called ''wave chaos.''

VI. DISCUSSION

For more than ten years, the ocean acoustics chaos group in Miami has been searching for a finite-frequency manifestation of ray chaos that exhibits a transient exponential growth, at the rate of the Lyapunov exponent, of some measurable quantity. This effect would point toward an ocean acoustic experiment that would validate the reality of ''wave chaos.'' It appears that such an effect has been discovered—the explosive spreading of narrow-angle beams. It is possible that other finite-frequency effects can exhibit a similar transient exponential growth.

The reason that the reality of ''wave chaos'' has been doubted by most researchers is that linear wave equations, unlike the nonlinear ray equations, are believed not to exhibit exponential sensitivity to initial conditions. Indeed, the finite (nonzero) acoustic wavelength limits the complexity of a wave field, whereas in the geometric approximation there is no such limit. Note by complexity, we mean the resolvability of caustic structures. Interference and diffraction limit one's ability to resolve the exponentially growing (with range) number of caustics. Therefore, if an exponential growth of some quantity is possible at finite frequencies, it must be a transient phenomenon that saturates at long range. Such a phenomenon is what has been uncovered in this paper—the explosive spreading of beams. Numerical simulations using a PE acoustic model are needed to support the theoretical predictions made in this paper.

The basic conditions that should be satisfied in an experimental verification of the predicted effect are that the index of refraction fluctuations should be small, and they should have a length scale that is large compared to the acoustic wavelength. Then multiple small-angle forward scattering accumulates in range to cause an initially narrow beam to expand exponentially, or ''explosively.'' In ocean acoustics, such an experiment could be done in either the horizontal plane or in the vertical plane using naturally occurring mesoscale variability. An ocean acoustics experiment could also be done at higher frequencies and shorter ranges using internal wave or microscale turbulence fluctuations. Also a laboratory acoustics experiment, or even an optics experiment, might serve the purpose of verifying the theory presented in this paper.

Finally, it is noted that the novel effect presented in this paper may have implications for fundamental physics related to ''quantum chaos.'' [18] Thus the theory presented here predicts that a quantum mechanical particle, initialized as a Gaussian wave packet, that moves through a disordered medium may under certain conditions spread exponentially with increasing time at a rate determined by the Lyapunov exponent that is found in the classical limit. Observation of such an effect would reveal a manifestation of classical chaos in a linear wave mechanical system.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research, Code 321OA.

[1] D. R. Palmer, M. G. Brown, F. D. Tappert, and H. F. Bezdek, ''Classical chaos in nonseparable wave propagation problems,'' Geophys. Res. Lett. **15**, 569–572 (1988).
[2] M. G. Brown, F. D. Tappert, G. J. Goñi, and K. B. Smith, ''Chaos in underwater acoustics,'' in *Ocean Variability & Acoustic Propagation*, edited by J. Potter and A. Warn-Varnas (Kluwer Academic, Dordrecht, 1991), pp. 139–160.
[3] K. B. Smith, M. G. Brown, and F. D. Tappert, ''Ray chaos in underwater acoustics,'' J. Acoust. Soc. Am. **91**, 1939–1949 (1992).
[4] F. D. Tappert and X. Tang, ''Ray chaos and eigenrays,'' J. Acoust. Soc. Am. **99**, 185–195 (1996).
[5] M. A. Wolfson and F. D. Tappert, ''Chaos in an acoustic propagation model,'' in *The Chaos Paradigm: Developments and Applications in Engineering and Science*, edited by R. A. Katz [AIP Conf. Proc. **296**, 277–288 (1994)].
[6] M. A. Wolfson and F. D. Tappert, ''Study of horizontal multipaths and ray chaos due to ocean mesoscale structure,'' J. Acoust. Soc. Am. **107**, 154–162 (2000).
[7] K. B. Smith, M. G. Brown, and F. D. Tappert, ''Acoustic ray chaos induced by mesoscale ocean structure,'' J. Acoust. Soc. Am. **91**, 1950–1959 (1992).
[8] F. D. Tappert, D. Lee, and H. Weinberg, ''The hybrid parabolic equation—ray model,'' in Recent Progress in the Development and Application of the Parabolic Equation, edited by P. D. Scully-Power and D. Lee, NUSC Technical Document No. **7145**, 7 May 1984, pp. 6-1 to 6-14.
[9] F. D. Tappert, D. Lee, and H. Weinberg, ''High-frequency propagation modeling using HYPER,'' J. Acoust. Soc. Am. Suppl. 1 **75**, S30 (1984).
[10] F. D. Tappert, ''The parabolic approximation method,'' in *Wave Propagation and Underwater Acoustics*, Lecture Notes in Physics Vol. 70, edited by J. B. Keller and J. S. Papadakis (Springer, New York, 1977), Chap. V, pp. 224–287.
[11] L. A. Chernov, *Wave Propagation in a Random Medium* (McGraw–Hill, New York, 1960).
[12] I. C. Percival, ''Chaos in Hamiltonian systems,'' in *Dynamical Chaos*, edited by M. V. Berry, I. C. Percival, and N. O. Weiss (Princeton University Press, Princeton, NJ, 1987), pp. 131–143.
[13] B. V. Chirikov, ''Patterns in chaos,'' Chaos, Solitons Fractals **1**, 79–103 (1991).
[14] M. A. Wolfson and S. Tomsovic, ''On the stability of long-range sound propagation through a structured ocean,'' J. Acoust. Soc. Am. **109**, 2693–2703 (2001).
[15] F. D. Tappert and H. Weinberg, ''HYPER: A hybrid parabolic equation—ray model for underwater sound propagation in the vicinity of smooth caustics,'' J. Acoust. Soc. Am. Suppl. **81**, S11 (1987).
[16] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, New York, 1994).
[17] M. B. Porter and H. P. Bucker, ''Gaussian beam tracing for computing ocean acoustic fields,'' J. Acoust. Soc. Am. **82**, 1349–1359 (1987).
[18] E. J. Heller and S. Tomsovic, ''Postmodern quantum mechanics,'' Phys. Today **46**(7), 38–45 (1993).

Finite-bandwidth effects on the causal prediction of ultrasonic attenuation of the power-law form

Joel Mobley
Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6101

Kendall R. Waters and James G. Miller [a)]
Laboratory for Ultrasonics, Washington University in St. Louis, St. Louis, Missouri 63130

(Received 16 August 2002; revised 13 August 2003; accepted 2 September 2003)

Kramers–Kronig (K–K) relations exist as a consequence of causality, placing nonlocal constraints on the relationship between dispersion and absorption. The finite-bandwidth method of applying these relations is examined where the K–K integrals are restricted to the spectrum of the experimental data. These finite-bandwidth K–K relations are known to work with resonant-type data and here are applied to dispersion data consistent with a power-law attenuation coefficient (exponent from 1 to 2). Bandwidth-restricted forms of the zero and once-subtracted K–K relations are used to determine the attenuation coefficient from phase velocity. Analytically, it is shown that these transforms produce the proper power-law form of the attenuation coefficient as a stand-alone term summed with artifacts that are dependent on the integration limits. Calculations are performed to demonstrate how these finite-bandwidth artifacts affect the K–K predictions under a variety of conditions. The predictions are studied in a local context as a function of subtraction frequency, bandwidth, and power-law exponent. The K–K predictions of the power-law exponent within various decades of the spectrum are also examined. In general, the agreement between finite-bandwidth K–K predictions and exact values grows as the power-law exponent approaches 1 and with increasing bandwidth.
[DOI: 10.1121/1.1621394]

PACS numbers: 43.35.Bf, 43.35.Cg, 43.20.Hq [RLW]

Pages: 2782–2790

I. INTRODUCTION

Fundamentally rooted in causality, Kramers–Kronig (K–K) relations have proven to be powerful, practical tools with applications across many disciplines of physics. In ultrasonics, subtracted forms of the K–K relations have been used to accurately predict dispersion and attenuation in systems exhibiting resonant behavior (encapsulated microbubbles[1]) and power-law growth[2] in attenuation. The principal difficulty in applying K–K relations directly to ultrasonic data is the finite bandwidth inherent in experimentally measured spectra. By restricting the range of integration to the measurement spectrum, artifacts are introduced that can seriously impact the accuracy of K–K transformations. It is still an open question as to whether an accurate general procedure exists for applying K–K relations directly (i.e., blindly) to any type of dispersion/attenuation data, or if some information (e.g., an analytical model) about the target system beyond the bandlimited attenuation and dispersion data is required to overcome the artifacts.

Previously, we examined the finite-bandwidth K–K problem for resonant systems using experimental data from Albunex® microbubble suspensions. Within the measurement spectrum, which covered more than a decade in frequency, these data exhibited a single, well-resolved resonance with a full-width at half maximum covering about one quarter of the total bandwidth. For such a system, we found that accurate K–K inversions were possible using only the measured dispersion and attenuation data (i.e., no out-of-band extrapolations or model fitting were required). In contrast, this work focuses on the systems that exhibit monotonic dispersion and attenuation, specifically those whose attenuation coefficient can be described by a power law. The power-law form of the attenuation coefficient is consistent with the behavior of a considerable number of solids,[3] liquids,[2,3] and biological tissues[4,5] in the MHz range, and is thus of interest in ultrasonic research. An accurate model-dependent procedure for applying K–K transformations specifically to power-law systems has been demonstrated previously.[2] In that study, the data were fit to the power-law model, and the predictions are expressed as functions of the fitting parameters. These expressions for velocity and attenuation are derived by assuming the model used to fit the data can be extrapolated to the entire spectrum, from $\omega=0$ to $\omega\to\infty$. The present study is distinguished from this earlier work because here we assume that the model holds only within the bandwidth of interest, and we examine the impact this spectrally limited knowledge imposes on K–K predictions. This paper is specifically concerned with calculating the attenuation coefficient with finite-bandwidth K–K using the causally consistent form of the phase velocity as the input.

There are two aspects of the finite-bandwidth K–K problem that are specifically addressed in this work. The first part of the study examines the case where a subtraction frequency ω_0 is chosen at an interior frequency well within the bandwidth of interest (at least a factor of 2 above the lower limit). This portion of the study is a test of the K–K calculation procedure that was successfully applied to the resonant

a)Electronic mail: james.g.miller@wustl.edu

microbubble data,[1] where certain choices of ω_0 across the spectrum optimized the accuracy of the predictions. This part also has implications for composite media that have both resonant and power-law features (e.g., particles suspended in an oil). The second part of the study is concerned with predicting the power-law exponent of the attenuation coefficient from the K–K transformation of bandlimited dispersion data. In this case, both the subtraction frequency and lower limit of integration are taken to be zero. The power-law trajectory of the K–K predicted attenuation coefficient is examined in individual decades of frequency (i.e., powers of 10) below and up to the upper limit of integration.

In Sec. II, we provide some background on the transfer characteristics of media with a power-law attenuation coefficient and the causally consistent form for the phase velocity. The finite-bandwidth forms of the subtracted Kramers–Kronig relations are also introduced. In Sec. III analytical expressions for the finite-bandwidth K–K predictions of the attenuation coefficient are given. These results are the sum of two terms, the first being the proper power-law result and the second encompassing the finite-bandwidth artifact. In Sec. IV, the analytical results from Sec. III are used to calculate attenuation coefficient curves under various circumstances. The discussions of these results are woven throughout Sec. IV.

II. BACKGROUND

A. Ultrasonic attenuation and dispersion in power-law systems

The linear transport of ultrasonic waves across an isotropic medium of thickness D is accounted for in general by a Fourier-domain transfer function of the form

$$H(\omega,D)=\exp[iK(\omega)D], \tag{1}$$

where $K(\omega)=[\omega/c(\omega)]+i\alpha(\omega)$ is the complex wave number, $\alpha(\omega)$ is the attenuation coefficient, and $c(\omega)$ is the phase velocity. Experimentally, a power-law form of the attenuation coefficient

$$\alpha(\omega)=\alpha_0\omega^{1+\varepsilon} \quad \text{where} \quad 0\leqslant\varepsilon\leqslant 1, \tag{2}$$

has been found to provide an accurate fit to broadband data from a variety of solid and liquid materials within the respective measurement bandwidths. When Eq. (2) holds for the entire frequency spectrum, the phase velocity takes the causally consistent form[2,3]

$$\frac{1}{c(\omega)}-\frac{1}{c(\omega_0)}=\alpha_0\tan\left[(1+\varepsilon)\frac{\pi}{2}\right](\omega^{\varepsilon}-\omega_0^{\varepsilon}) \quad 0<\varepsilon\leqslant 1 \tag{3a}$$

and

$$=-\alpha_0\frac{2}{\pi}\ln\frac{\omega}{\omega_0} \quad \varepsilon=0. \tag{3b}$$

[Equation (3b) can be obtained by taking the $\varepsilon\to 0$ limit of Eq. (3a).] This form of the dispersion as well as the consistency of Eq. (2) and Eq. (3) have also been verified in the laboratory.[2,3] [Both $\alpha(\omega)$ and $c(\omega)$ are even functions, a property derived from the fact that $H(\omega,D)$ is the Fourier transform of a real function. Since negative frequencies are not explicitly considered, no absolute value signs are used.]

B. Kramers–Kronig relations for ultrasonic attenuation and phase velocity

Given that the medium described by $H(\omega,D)$ satisfies the physical requirements of causality and finite energy, the real and imaginary parts of $H(\omega,D)$ can be shown to form a Hilbert transform pair via Titchmarsh's theorem. The integration over negative frequencies can be mapped to the positive axis, and the resulting expressions are commonly known as Kramers–Kronig relations. The complex wave number itself derives its analytic properties from the transfer function, and under the proper conditions the complex wave number will retain the domain of analyticity required of Hilbert transform pairs. However, it will not fulfill the square integrability requirements since the attenuation coefficient must diverge as $\omega\to\infty$ to insure the square integrability of $H(\omega,D)$. In spite of this, K–K relations between $\alpha(\omega)$ and $c(\omega)$ can be derived by the method of subtractions,[6] which insures convergence of the integrals. Formed using $\gamma(\omega)=iK(\omega)$ as the basis function, the unsubtracted and once-subtracted relations are considered here. To allow for the straightforward adaptation to the restricted interval case, the relations are given in their expanded form.[1] The unsubtracted (zeroth) relation for the attenuation coefficient is

$$\alpha(\omega)=\lim_{\substack{\Omega\to\infty\\ \sigma\to 0}}\left[-\frac{1}{\pi}\int_{\sigma}^{\Omega}\frac{\dfrac{\omega'}{c(\omega')}-\dfrac{\omega}{c(\omega)}}{\omega'-\omega}d\omega'-\frac{1}{\pi}\int_{\sigma}^{\Omega}\frac{\dfrac{\omega'}{c(\omega')}+\dfrac{\omega}{c(\omega)}}{\omega'+\omega}d\omega'\right]. \tag{4}$$

Combining the two integrands into a single rational expression and counting the powers of ω', one can see that convergence requires that $\varepsilon<-1$ [where $\omega'/c(\omega')\sim\omega'^{1+\varepsilon}$]. This is, of course, inadequate for the range of power laws considered here, and higher-order subtractions must be considered. The once-subtracted relation in the expanded form is

$$\alpha(\omega)=\alpha(\omega_0)+\lim_{\substack{\Omega\to\infty\\ \sigma\to 0}}\left[-\frac{1}{\pi}\int_{\sigma}^{\Omega}\frac{\dfrac{\omega'}{c(\omega')}-\dfrac{\omega}{c(\omega)}}{\omega'-\omega}d\omega'-\frac{1}{\pi}\int_{\sigma}^{\Omega}\frac{\dfrac{\omega'}{c(\omega')}+\dfrac{\omega}{c(\omega)}}{\omega'+\omega}d\omega'+\frac{1}{\pi}\int_{\sigma}^{\Omega}\frac{\dfrac{\omega'}{c(\omega')}-\dfrac{\omega_0}{c(\omega_0)}}{\omega'-\omega_0}d\omega'+\frac{1}{\pi}\int_{\sigma}^{\Omega}\frac{\dfrac{\omega'}{c(\omega')}+\dfrac{\omega_0}{c(\omega_0)}}{\omega'+\omega_0}d\omega'\right], \tag{5}$$

where ω_0 is referred to as the subtraction frequency. By counting the powers of ω' after combining all four integrands, one can expect convergence for $\varepsilon<1$ which covers the range of power laws examined in this work. Note that for $\varepsilon=1$, there is no dispersion so this case is naturally excluded. We will refer to these K–K expressions whose integrals cover the entire positive frequency axis [i.e., $(\sigma,\Omega)\to(0,\infty)$] as "unrestricted."

The interval-restricted (i.e., finite-bandwidth) relations are given by

$$\alpha(\omega)^{(\sigma,\Omega,n=0)}=-\frac{1}{\pi}\int_\sigma^\Omega\frac{\frac{\omega'}{c(\omega')}-\frac{\omega}{c(\omega)}}{\omega'-\omega}d\omega'-\frac{1}{\pi}\int_\sigma^\Omega\frac{\frac{\omega'}{c(\omega')}+\frac{\omega}{c(\omega)}}{\omega'+\omega}d\omega', \tag{6}$$

and

$$\alpha(\omega)^{(\sigma,\Omega,n=1)}=\alpha(\omega_0)-\frac{1}{\pi}\int_\sigma^\Omega\frac{\frac{\omega'}{c(\omega')}-\frac{\omega}{c(\omega)}}{\omega'-\omega}d\omega'-\frac{1}{\pi}\int_\sigma^\Omega\frac{\frac{\omega'}{c(\omega')}+\frac{\omega}{c(\omega)}}{\omega'+\omega}d\omega'+\frac{1}{\pi}\int_\sigma^\Omega\frac{\frac{\omega'}{c(\omega')}-\frac{\omega_0}{c(\omega_0)}}{\omega'-\omega_0}d\omega'+\frac{1}{\pi}\int_\sigma^\Omega\frac{\frac{\omega'}{c(\omega')}+\frac{\omega_0}{c(\omega_0)}}{\omega'+\omega_0}d\omega', \tag{7}$$

where n is the subtraction order and the limits of integration are such that $0\leqslant\sigma<\Omega$.

III. THEORY

The results shown in this section are calculated by substituting the dispersion as given in Eq. (3a) and Eq. (3b) into the restricted interval forms, Eq. (6) and Eq. (7). The forms of the restricted K–K results depend on whether or not the evaluation frequency, ω, and the subtraction frequency, ω_0, are contained within the integration interval. The viewpoint taken in this work is that the bandwidth of integration denotes the extent of our knowledge about the complex wave number, and thus we examine only the situation where $\sigma<\omega$, $\omega_0<\Omega$. In all cases considered, $0\leqslant\varepsilon<1$. From the unsubtracted relation [Eq. (6)]

$$\alpha(\omega)^{(\sigma,\Omega,0)}=\alpha_0\omega^{1+\varepsilon}+F_0^{(\omega)}(\sigma,\Omega), \tag{8}$$

where $F_0^{(\omega)}(\sigma,\Omega)$ represents the artifactual terms that explicitly depend on the limits of integration

$$F_0^{(\omega)}(\sigma,\Omega)=-\frac{2}{\pi}\frac{\Omega-\sigma}{c(\omega)}-\frac{\alpha_0}{\pi}\tan\left[(1+\varepsilon)\frac{\pi}{2}\right]\times\left(2\sigma^{1+\varepsilon}\sum_{n=0}^{\infty}\frac{1}{2n+\varepsilon+3}\frac{\sigma^{2n+2}}{\omega^{2n+2}}-2\Omega^{1+\varepsilon}\sum_{n=0}^{\infty}\frac{1}{2n-\varepsilon-1}\frac{\omega^{2n}}{\Omega^{2n}}-2(\Omega-\sigma)\omega^{\varepsilon}-\omega^{1+\varepsilon}\ln\frac{\omega+\sigma}{\omega-\sigma}\frac{\Omega-\omega}{\Omega+\omega}\right) \tag{9a}$$

$$=-\frac{2}{\pi}\frac{\Omega-\sigma}{c(0)}-\frac{\alpha_0}{\pi}\tan\left[(1+\varepsilon)\frac{\pi}{2}\right]\times\left(2\sigma^{1+\varepsilon}\sum_{n=0}^{\infty}\frac{1}{2n+\varepsilon+3}\frac{\sigma^{2n+2}}{\omega^{2n+2}}-2\Omega^{1+\varepsilon}\sum_{n=0}^{\infty}\frac{1}{2n-\varepsilon-1}\frac{\omega^{2n}}{\Omega^{2n}}-\omega^{1+\varepsilon}\ln\frac{\omega+\sigma}{\omega-\sigma}\frac{\Omega-\omega}{\Omega+\omega}\right). \tag{9b}$$

In the limit of $\varepsilon\to0$ (logarithmic dispersion), the remainder becomes

$$\lim_{\varepsilon\to0}F_0^{(\omega)}(\sigma,\Omega)=-\frac{2}{\pi}\left(\frac{\Omega}{c(\Omega)}-\frac{\sigma}{c(\sigma)}\right)+\frac{2\alpha_0\omega}{\pi^2}\left(\ln\frac{\sigma}{\omega}\ln\frac{\omega+\sigma}{\omega-\sigma}-\ln\frac{\Omega}{\omega}\ln\frac{\Omega+\omega}{\Omega-\omega}-2\sum_{n=0}^{\infty}\frac{1}{(2n+3)^2}\frac{\sigma^{2n+3}}{\omega^{2n+3}}-2\sum_{n=0}^{\infty}\frac{1}{(2n-1)^2}\frac{\omega^{2n-1}}{\Omega^{2n-1}}\right). \tag{10}$$

In the limits $\sigma\to0$ and $\Omega\gg\omega$, this becomes

$$\alpha(\omega)^{(0,\Omega\gg\omega,0)}=\alpha_0\omega^{1+\varepsilon}-\Omega^{1+\varepsilon}\frac{2}{1+\varepsilon}\frac{\alpha_0}{\pi}\tan\left[(1+\varepsilon)\frac{\pi}{2}\right]\quad 0<\varepsilon<1 \tag{11a}$$

$$=\alpha_0\omega^{1+\varepsilon}-\Omega\frac{2}{\pi}\left(\frac{1}{c(\Omega)}+\frac{2\alpha_0}{\pi}\right)\quad\varepsilon=0, \tag{11b}$$

which diverges as $\Omega\to\infty$ as anticipated.

The relation with one subtraction yields the following:

$$\alpha(\omega)^{(\sigma,\Omega,1)}=\alpha_0\omega^{1+\varepsilon}-\alpha_0\omega_0^{1+\varepsilon}+\alpha(\omega_0)+F_1^{(\omega,\omega_0)}(\sigma,\Omega) \tag{12a}$$

$$=\alpha_0\omega^{1+\varepsilon}+F_1^{(\omega,\omega_0)}(\sigma,\Omega), \tag{12b}$$

where

$$F_1^{(\omega,\omega_0)}(\sigma,\Omega) = -\frac{\alpha_0}{\pi}\tan\left[(1+\varepsilon)\frac{\pi}{2}\right] \times\left(2\sigma^{1+\varepsilon}\sum_{n=0}^{\infty}\frac{\sigma^{2n+2}}{2n+\varepsilon+3}\left(\frac{1}{\omega^{2n+2}}-\frac{1}{\omega_0^{2n+2}}\right) -2\Omega^{1+\varepsilon}\sum_{n=1}^{\infty}\frac{1}{2n-\varepsilon-1}\frac{\omega^{2n}-\omega_0^{2n}}{\Omega^{2n}} -\omega^{1+\varepsilon}\ln\frac{\omega+\sigma}{\omega-\sigma}\frac{\Omega-\omega}{\Omega+\omega}+\omega_0^{1+\varepsilon}\ln\frac{\omega_0+\sigma}{\omega_0-\sigma}\frac{\Omega-\omega_0}{\Omega+\omega_0}\right). \tag{13}$$

In the limit $\varepsilon\to 0$, the remainder term becomes

$$\lim_{\varepsilon\to 0} F_1^{(\omega,\omega_0)}(\sigma,\Omega) = \frac{2\alpha_0}{\pi^2}\left(\omega\ln\frac{\sigma}{\omega}\ln\frac{\omega+\sigma}{\omega-\sigma}-\omega_0\ln\frac{\sigma}{\omega_0}\ln\frac{\omega_0+\sigma}{\omega_0-\sigma} -\omega\ln\frac{\Omega}{\omega}\ln\frac{\Omega+\omega}{\Omega-\omega}+\omega_0\ln\frac{\Omega}{\omega_0}\ln\frac{\Omega+\omega_0}{\Omega-\omega_0} -2\sigma\sum_{n=0}^{\infty}\frac{\sigma^{2n+2}}{(2n+3)^2}\left(\frac{1}{\omega^{2n+2}}-\frac{1}{\omega_0^{2n+2}}\right) -2\Omega\sum_{n=0}^{\infty}\frac{1}{(2n-1)^2}\frac{\omega^{2n}-\omega_0^{2n}}{\Omega^{2n}}\right). \tag{14}$$

Taking the limit of large Ω and small σ

$$\lim_{\Omega\gg\omega,\omega_0}\alpha(\omega)^{(\sigma,\Omega,1)} = \alpha_0\omega^{1+\varepsilon}+\frac{\alpha_0}{\pi}\tan\left[(1+\varepsilon)\frac{\pi}{2}\right]\frac{2}{1-\varepsilon}\frac{\omega^2-\omega_0^2}{\Omega^{1-\varepsilon}} \quad 0<\varepsilon<1 \tag{15a}$$

$$= \alpha_0\omega-\frac{4\alpha_0}{\pi^2}\frac{\omega^2-\omega_0^2}{\Omega} \quad \varepsilon=0, \tag{15b}$$

which converge to $\alpha(\omega)$ in the limit of $\Omega\to\infty$. The subtraction frequency ω_0 exerts no influence on the shape of $\alpha(\omega)^{(\sigma,\Omega,1)}$, but only serves to "anchor" $\alpha(\omega)^{(\sigma,\Omega,1)}$ to $\alpha(\omega)$ at $\omega=\omega_0$. To confirm these analytical results, the restricted bandwidth relation of Eq. (7) was numerically integrated under a variety of conditions using the form of the dispersion in Eq. (3a). In all of these comparisons, the analytical expression and numerical integrations produced essentially equivalent outcomes.

Both the zeroth- and first-order calculations produce results with the same general structure—the correct power-law form of the attenuation coefficient coexisting with additional additive terms that depend upon at least one of the limits of integration. However, in the unsubtracted result the extra terms grow with increasing Ω, while the once-subtracted result converges to the correct result as $\Omega\to\infty$.

IV. RESULTS AND DISCUSSION

A. Locally anchored performance ($\omega_0 \geqslant 2\sigma$)

The unrestricted K–K relations have demonstrated their utility for power-law systems, since they generate analytical forms for the attenuation and dispersion [Eq. (2) and Eq. (3)] that are consistent with experimental data.[2] In this model-dependent approach the K–K inversion is reduced to the simple act of determining α_0 and ε through some fitting procedure. Even though the model-fitting method has so far proven a successful approach for performing K–K transformations of power-law systems, there are two principal reasons for investigating their behavior with the more general restricted-bandwidth K–K. First, the restricted K–K method has been shown to work for resonant-type data without using any model-dependent or extrapolated parameters.[1] It is then natural to investigate the method with other types of data to assess its wider applicability. Second, it is possible to have a system that has a combination of localized resonant structures riding on a power-law background. Since many liquids exhibit power-law attenuation, suspensions of microbubbles or microspheres in such a liquid could display this composite behavior.

As shown in the previous study of the microbubble system, the accuracy of the restricted K–K prediction for an isolated resonance comes down to the choice of the subtraction frequency ω_0. Certain choices of ω_0 from within the data spectrum effectively minimize the finite-bandwidth artifacts, and the rationale behind these choices was demonstrated using an analytical model.[1] For the present power-law case, the accuracy of the restricted-bandwidth approach has more to do with the bandwidth of the data than the choice of subtraction frequency. The results shown in this section (see Fig. 1) are meant to examine the restricted K–K approach in a manner consistent with the way they are applied to the resonant data; specifically the subtraction frequency is chosen from the known bandwidth and is distinct from the band edges. Also, the limits of integration include realistic measurement bandwidths.

In Fig. 1 the restricted K–K predictions, calculated using Eq. (12), are plotted over a region around the subtraction frequency ω_0 for various limits of integration. The region of the spectrum displayed in the figure (from $0.5\omega_0$ to $1.5\omega_0$) shows only a portion of the total spectra used to calculate the four K–K curves. This spectral region displayed in Fig. 1 encompasses the widths of resonant peaks encountered in data from encapsulated microbubble[1] ($\omega_{\text{res}}/\Delta\omega^{\text{full width}}_{\text{half max}}\sim 1$) and polymer microsphere[7] ($\omega_{\text{res}}/\Delta\omega^{\text{full width}}_{\text{half max}}\sim 10$) suspensions. This serves to illustrate the behavior of the predicted curves over those frequency scales encountered in resonant data with resolved peaks. In panel (a), the $1+\varepsilon=1.1$ results are plotted. For the curve calculated using the widest bandwidth (with lower limit of integration $\sigma=0.05\omega_0$, and upper limit of integration $\Omega=50\omega_0$), the K–K prediction accounts for 89% of the variation in the exact curve. If the lower limit

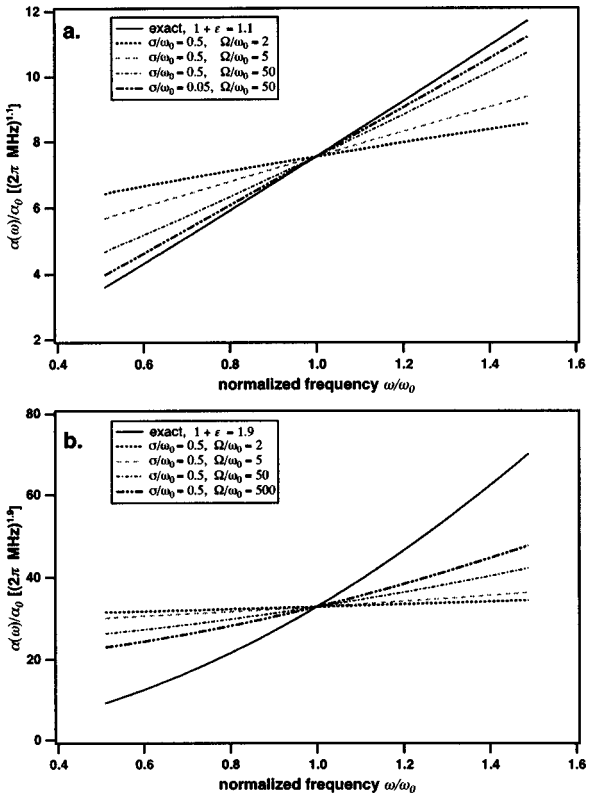


FIG. 1. (a) The comparison of $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$ for $1+\varepsilon=1.1$ around the subtraction frequency ω_0 for bandwidths indicated in the legends. (b) Comparison of $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$ for $1+\varepsilon=1.9$.

of integration is moved up to the boundary of the plotted curves ($\sigma=0.5\omega_0$) while holding the upper limit fixed, the predicted variation falls to 75% of the target value. The predicted variation decreases further as the upper limit is lowered, down to 26% of the exact value for the narrowest bandwidth case, $\sigma=0.5\omega_0$ to $\Omega=2\omega_0$. In panel (b), the $1+\varepsilon=1.9$ results are shown. Here, the predictions are relatively insensitive to changes in σ, so the only changes from curve to curve are in Ω. In this case the predicted variations in the attenuation coefficient range from 41% down to 5% of the exact value. The impact of these results on predicting peaks and trends ultimately depends on many system-dependent factors (e.g., relative peak-to-background ratio). When the peak is localized above (below) the subtraction frequency, the peak-to-background ratio will be over (under) estimated. If the resonant peaks are clearly resolved, when the peak is near ω_0 it can be reproduced without significant distortion, especially for the smaller values of ε. In fact, the restricted K–K transform is like a ''local filter'' in this case, suppressing the monotonic global trend in favor of the localized variations.

B. Predicted power-law trajectories, ω_0, $\sigma\rightarrow 0$

In this section, we apply the restricted K–K relations in a manner suited to predicting power-law exponents for the attenuation coefficient. For this purpose, the spectrum will be examined in specific decades (i.e., powers of 10) down from the upper limit of integration. In each decade, a power law is fit to the predicted curve to estimate the exponent for the curve. The subtraction frequency and lower limit of integration are taken to be equal and arbitrarily close to zero so that all terms containing σ and ω_0 as factors vanish. The once-subtracted results for $\alpha(\omega)^{(\sigma,\Omega,1)}$, expressed in Eq. (12), are calculated for four power laws, $1+\varepsilon=1.9$, 1.5, 1.1, and 1.0 (logarithmic dispersion). The objective is to identify the regions of the spectrum in which there is agreement between the predicted and exact power-law exponents. In all of the results to follow (shown in Figs. 2–5), the power-law fits to $\alpha(\omega)^{(\sigma,\Omega,1)}$ have an R^2 value $\geqslant 0.99$.

The results of the $1+\varepsilon=1.0$ case, where the dispersion is logarithmic, are shown in Fig. 2. In panel (a), $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$ are compared on a log–log plot covering three decades in frequency up to the upper limit of integration Ω. One can see that the prediction tracks the exact trajectory well for $\omega/\Omega<0.05$. In panel (b), $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$ are compared on a linear–linear plot over the range $0.001<\omega/\Omega\leqslant 0.01$, two decades down from the upper limit. A power-law fit to $\alpha(\omega)^{(\sigma,\Omega,1)}$ for just this decade yields $1+\varepsilon=0.99$. In panel (c), $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$ are compared over the top decade, $0.1<\omega/\Omega\leqslant 1$, where a power-law fit to $\alpha(\omega)^{(\sigma,\Omega,1)}$ yields $1+\varepsilon=0.77$. Thus, to match the proper power-law trajectory with restricted K–K to about 1% in this case requires knowledge of the dispersion about 1.5 orders of magnitude above the frequency scale of interest.

The five panels of Fig. 3 display the results for the $1+\varepsilon=1.1$ case. In this figure (and the ones to follow) in addition to $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$, we also plot the dominant artifactual terms from Eq. (12). For the cases examined here where ω_0 and σ are arbitrarily small, the artifact is due to two terms—referred to as the ''Ω series''

$$\frac{\alpha_0}{\pi}\tan\left[(1+\varepsilon)\frac{\pi}{2}\right]2\Omega^{1+\varepsilon}\sum_{n=1}^{\infty}\frac{1}{2n-\varepsilon-1}\frac{\omega^{2n}}{\Omega^{2n}}, \tag{16}$$

and the ''log term''

$$\frac{\alpha_0}{\pi}\tan\left[(1+\varepsilon)\frac{\pi}{2}\right]\omega^{1+\varepsilon}\ln\frac{1-\omega/\Omega}{1+\omega/\Omega}. \tag{17}$$

In panel (a), the quantities $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$ are shown on a log–log plot covering three decades of the spectrum up to Ω. Moving down the frequency scale, the two curves merge in the middle decade similar to the previous case. In panel (b), $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$ are compared on a linear plot over the third decade down from Ω. Here, the fit to $\alpha(\omega)^{(\sigma,\Omega,1)}$ yields an exponent of $1+\varepsilon=1.09$, demonstrating the high degree of agreement at this scale. In panel (c), the two quantities comprising the error are plotted over the same decade as panel (b). The two terms act in opposition to one another, with the Ω series having the greater magnitude. Note that the scale of the y axis in panel (c) is an order of magnitude smaller than that of panel (b). In panel (d), the exact and predicted K–K results are plotted for the top decade, and the deviation between the two is clear. A power-law fit over this decade produces $1+\varepsilon=0.86$. This is 78% of the target value, very similar to the previous case in the top decade. In panel (e) the error terms are shown to have similar shapes and are largely counterbalancing one another. Also

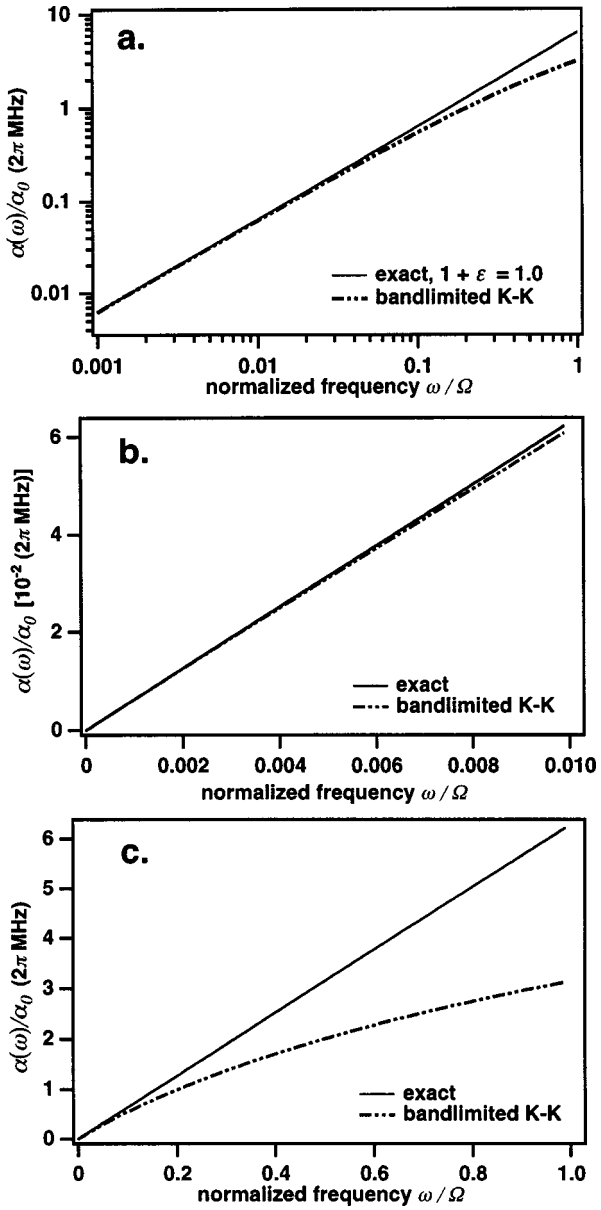


FIG. 2. (a) Log–log plot of the exact and restricted-interval once-subtracted K–K prediction for attenuation coefficient for $1+\varepsilon=1.0$, the logarithmic dispersion case, covering three decades in frequency. (b) A linear plot of the exact and predicted attenuation coefficients for the third decade, $0.001<\omega/\Omega\leqslant 0.01$. (c) A linear plot of the exact and predicted attenuation coefficient over the top decade ($\omega/\Omega<1.0$).

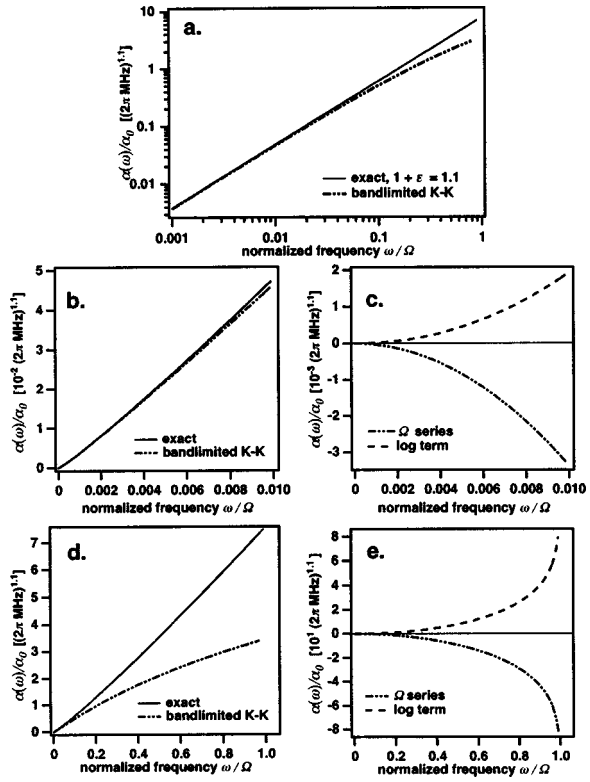


FIG. 3. (a) Log–log plot of the exact and restricted-interval once-subtracted K–K prediction for attenuation coefficient for $1+\varepsilon=1.1$. The plot covers three decades in frequency. (b) A linear plot of the exact and predicted attenuation coefficients over the third decade, $0.001<\omega/\Omega\leqslant 0.01$. (c) The two error terms, the Ω series [Eq. (16)] and the log term [Eq. (17)], in the K–K calculations for the third decade. The vertical scale is an order of magnitude narrower than that in panel (b). (d) A linear plot of the exact and predicted attenuation coefficient over the top decade ($\omega/\Omega<1.0$). (e) The two error terms in the K–K calculations for the top decade. Here, the vertical axis is 20 times broader than that of panel (d).

note that the two error terms are individually much larger than $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$, as the vertical axis is about 20 times broader than in panel (d). Both error terms individually diverge in the $\omega\rightarrow\Omega$ limit, but one can show that together these divergences cancel.

Figure 4 contains the results for the $1+\varepsilon=1.5$ case. In panel (a), the quantities $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$ are shown on a log–log plot covering three decades of the spectrum. The K–K prediction deviates noticeably from $\alpha(\omega)$ in the second decade and higher. In panel (b) $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$ are compared linearly in the $0.001<\omega/\Omega\leqslant 0.01$ decade. Here, the correspondence is very good with a power-law fit of $1+\varepsilon=1.46$. As shown in panel (c), the Ω series term is about an order of magnitude larger than the logarithmic term, but is itself about an order of magnitude smaller than $\alpha(\omega)$. In panel (d) $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$ are compared on a linear scale over the top decade. Similar to the two previous cases, the predicted exponent of $1+\varepsilon=1.13$ in this top decade is only about 75% of the target value. Panel (e) shows the two artifactual terms in $\alpha(\omega)^{(\sigma,\Omega,1)}$. The two terms have similar shapes although they differ in sign, and the Ω series clearly has the greater magnitude of the two. Note that the y axis in panel (e) is about a factor of 5 broader than that for panel (d).

The five panels of Fig. 5 display the results for the $1+\varepsilon=1.9$ case. In panel (a), the quantities $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$ are shown on a log–log plot covering six decades of the spectrum up to Ω. The K–K prediction deviates from $\alpha(\omega)$ increasingly from the smallest decade up to Ω. Panel (b) shows $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$ two decades down from the top. Here, the correspondence is not as strong as in the previous three cases, but the fit exponent of $1+\varepsilon=1.77$ for this decade accounts for 93% of the target value. For the artifact components shown in panel (c), it is clear that the Ω-series term is dominant, with no significant contribution from the log term. In panel (d) $\alpha(\omega)^{(\sigma,\Omega,1)}$ and $\alpha(\omega)$ are compared over the top decade. Here, the difference is dramatic, although the fit exponent of $1+\varepsilon=1.41$ is 74% of the target value, a similar percentage as in the other power-law cases in the top decade.

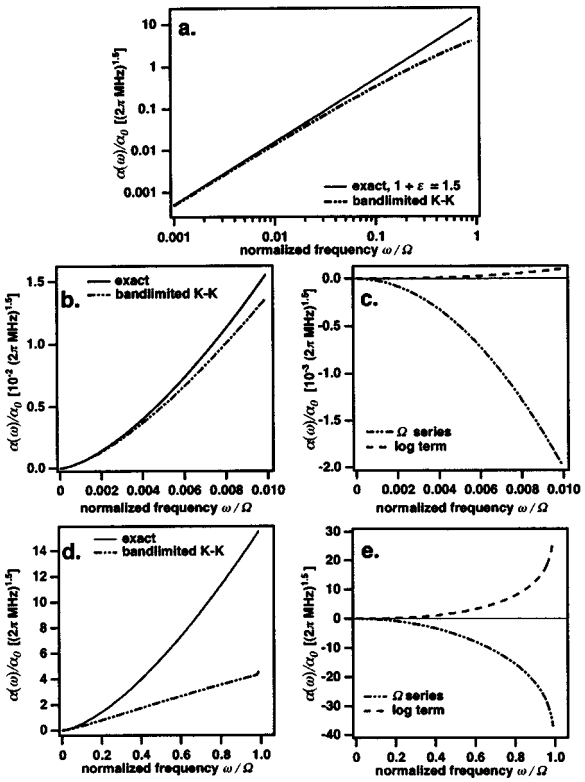


FIG. 4. (a) Log–log plot of the exact and restricted-interval once-subtracted K–K prediction for attenuation coefficient for $1+\varepsilon=1.5$. The plot covers three decades in frequency. (b) A linear plot of the exact and predicted attenuation coefficients in the third decade, $0.001<\omega/\Omega\leqslant 0.01$. (c) The two error terms, the Ω series [Eq. (16)] and the log term [Eq. (17)], in the K–K calculations for the third decade. The vertical scale is an order of magnitude smaller than that in panel (b). (d) A linear plot of the exact and predicted attenuation coefficient over the top decade ($\omega/\Omega<1.0$). (e) The two error terms in the K–K calculations over the top decade. The vertical axis is about 5 times broader than that of panel (d).

Panel (e) shows the two artifactual terms in $\alpha(\omega)^{(\sigma,\Omega,1)}$. In panel (e), the dominance of the Ω series in the artifact is clear, although the logarithmic term counterbalances it somewhat.

In each respective decade shown, the Ω series term is of a similar magnitude across the three $\varepsilon>0$ cases, while the log term provides less of a counterbalance as ε increases. To illustrate this trend, we add Eq. (16) and Eq. (17) to get the total artifact, writing the logarithm in Eq. (17) in its power series form. The first term in the total artifact series is

$$\alpha_0\omega^{1+\varepsilon}\frac{2}{\pi}\tan\left[(1+\varepsilon)\frac{\pi}{2}\right]\left(\frac{1}{1-\varepsilon}\frac{\Omega^{\varepsilon}}{\omega^{\varepsilon}}-1\right)\frac{\omega}{\Omega}. \tag{18}$$

(For $\omega<\Omega/2$, this first term approximation is accurate to better than 10%.) In the last bracket of Eq. (18), the first term is from the Ω series and the second is from the log term. The tangent can also be expanded to first order as $\tan[(1+\varepsilon)\times(\pi/2)]\approx(\varepsilon-1)(\pi/2)$ (which is accurate to better than 10% for $\varepsilon>0.66$)

$$\alpha_0\omega^{1+\varepsilon}\frac{2}{\pi}(\varepsilon-1)\frac{\pi}{2}\left(\frac{1}{1-\varepsilon}\frac{\Omega^{\varepsilon}}{\omega^{\varepsilon}}-1\right)\frac{\omega}{\Omega}$$

$$=-\alpha_0\omega^{1+\varepsilon}\left(\frac{\Omega^{\varepsilon}}{\omega^{\varepsilon}}+(\varepsilon-1)\right)\frac{\omega}{\Omega}. \tag{19}$$

As shown by the above expression, as ε increases towards 1, the log-term contribution goes to zero while the singularity in the first Ω-series term compensates for the zero in the tangent, resulting in a finite nonvanishing contribution.

The power law predicted by K–K for a given decade can be calculated using the following formula:

$$(1+\varepsilon)_k^{(\sigma,\Omega,1)}=\log_{10}\left[\frac{\alpha^{(\sigma,\Omega,1)}(10^{-k+1}\Omega,\omega_0)}{\alpha^{(\sigma,\Omega,1)}(10^{-k}\Omega,\omega_0)}\right]$$

$$=1+\varepsilon+\log_{10}\left[\frac{1+\frac{2}{\pi}\tan\left[(1+\varepsilon)\frac{\pi}{2}\right]\sum_{n=1}^{\infty}\left(\frac{10^{\varepsilon(k-1)}}{2n-\varepsilon-1}-\frac{1}{2n-1}\right)10^{(1-k)(2n-1)}}{1+\frac{2}{\pi}\tan\left[(1+\varepsilon)\frac{\pi}{2}\right]\sum_{n=1}^{\infty}\left(\frac{10^{\varepsilon k}}{2n-\varepsilon-1}-\frac{1}{2n-1}\right)10^{-k(2n-1)}}\right] \quad (0<\varepsilon<1) \tag{20a}$$

$$=1+\log_{10}\left[\frac{1-\frac{4}{\pi^2}\sum_{n=1}^{\infty}\left(\frac{1}{2n-1}\ln 10^{k-1}+\frac{1}{(2n-1)^2}\right)10^{(1-k)(2n-1)}}{1-\frac{4}{\pi^2}\sum_{n=1}^{\infty}\left(\frac{1}{2n-1}\ln 10^{k}+\frac{1}{(2n-1)^2}\right)10^{-k(2n-1)}}\right] \quad (\varepsilon=0), \tag{20b}$$

where k is a positive integer, $\omega_0\ll 10^{-k}$, and $\sigma\ll 10^{-k}\,\Omega$. Once again the logarithms in the artifact have been written in the power series form. The top decade is given by $k=1$ and the lower decades by the higher integers. In the top decade, this

formula matches the exponents numerically fit to the predicted curves to 2% and in the lower decades it matches better than 0.8%. Below the first decade (i.e., $k \geqslant 2$), the above formula is essentially determined by the first terms in the infinite series

$$(1+\varepsilon)_{k\geqslant 2}^{(\sigma,\Omega,1)} = 1+\varepsilon+\log_{10}\left[\frac{1+\frac{2}{\pi}\tan\left((1+\varepsilon)\frac{\pi}{2}\right)\left(\frac{10^{\varepsilon(k-1)}}{1-\varepsilon}-1\right)10^{(1-k)}}{1+\frac{2}{\pi}\tan\left((1+\varepsilon)\frac{\pi}{2}\right)\left(\frac{10^{\varepsilon k}}{1-\varepsilon}-1\right)10^{-k}}\right] \quad (0<\varepsilon<1), \tag{21a}$$

$$= 1+\log_{10}\left[\frac{1-\frac{4}{\pi^2}(\ln 10^{k-1}+1)10^{(1-k)}}{1-\frac{4}{\pi^2}(\ln 10^{k}+1)10^{-k}}\right] \quad (\varepsilon=0). \tag{21b}$$

As shown in this section, the restricted K–K method is clearly not practical for the prediction of power-law exponents in the $1 \leqslant 1+\varepsilon < 2$ range. Even in the most favorable case of $\varepsilon=0$, accurate predictions require the data to span a spectrum up to 50 times the lowest valid frequency in the measurement. Such wideband data are rarely available as acquiring dispersion and/or attenuation data over even a single decade in frequency is a challenging task (at least with a single set of transducers, pulsers, and amplifiers). For example, if one were interested in the attenuation behavior around 2 MHz, dispersion data would be required up to at least 100 MHz to achieve reasonable accuracy. The data would also need to extend down a decade or so below 2 MHz. A direct fit of the data to the power law model of Eq. (2) and Eq. (3) is probably the best approach as the model-independent method is so spectrally demanding.

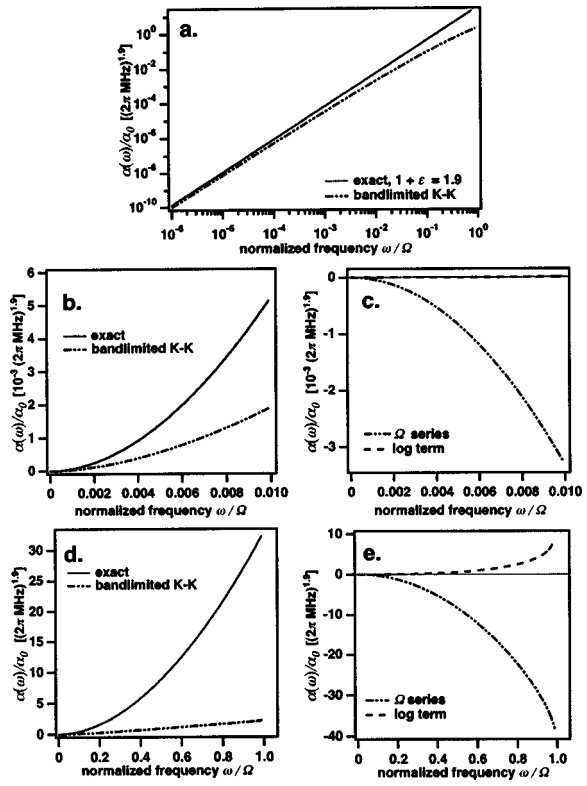


FIG. 5. (a) Log–log plot of the exact and restricted-interval once-subtracted K–K prediction for attenuation coefficient for $1+\varepsilon=1.9$. The plot covers six decades in frequency. (b) A linear plot of the exact and predicted attenuation coefficient for the third decade, $0.001<\omega/\Omega\leqslant 0.01$. (c) The two error terms, the Ω series [Eq. (16)] and the log term [Eq. (17)], in the K–K calculations for the third decade. The vertical scale is of the same order of magnitude as in panel (b). (d) A linear plot of the exact and predicted attenuation coefficients for the top decade, $\omega/\Omega<1.0$. (e) The dominant error terms in the K–K calculations for the top decade. The vertical scale is of the same order of magnitude as in panel (d).

One use of these results could be in extrapolating the range of knowledge of a system beyond the finite spectral window of our measurement system. It is possible that the power-law behavior only persists over a limited spectrum, and the system exhibits some other type of behavior elsewhere. The restricted K–K integrals could be interpreted as representing the power-law portion of the system's behavior. A successful fit of the data to the causal model within the measurement bandwidth could then indicate that the dispersion follows the model for several decades beyond the high end of the data; otherwise, the fit to the attenuation data would exhibit some deviation. The quantitative implications of this idea remain to be explored.

C. On predicting dispersion from the attenuation coefficient

As shown above, in predicting the attenuation coefficient the restricted K–K relations produce the correct power-law form for both zeroth- and first subtraction orders. Even though the zeroth-order result does not converge in the $\Omega\rightarrow\infty$ limit, as the bandwidth increases the slope of the predicted attenuation does approach the correct value. When predicting the dispersion from power-law attenuation for the various orders of restricted K–K, the results are more complex in both form and interpretation. The dispersion prediction problem differs from the attenuation case in two important respects: (1) the lowest order convergent relation is the twice-subtracted form, and (2) the lower-order nonconvergent results do not produce σ- and Ω-independent terms that correspond to the correct functional form of the dispersion. The wider implications of the dispersion results within the

context of the search for a more generally applicable K–K method are yet to be explored.

V. CONCLUSION

In this work, we have demonstrated the impact of bandwidth restriction on the causal prediction of the attenuation coefficient from dispersion for power-law systems. The calculations using both the zeroth- and first-order subtracted K–K relations generate analytical results as sums of two terms. The first term is the proper power-law form for the attenuation coefficient, while the second consists of all the artifactual factors which have explicit dependencies on the limits of integration. We have shown that the once-subtracted relation converges to $\alpha(\omega)$ as the bandwidth grows to encompass larger portions of the frequency axis while the zero-order transform diverges as $\Omega \to \infty$. As illustrated in the numerical studies, the restricted K–K predictions grow more accurate as $1+\varepsilon \to 1$. The predictions are also better when the frequency scale of interest is farther removed from the upper limit of integration. The bandwidth-restricted K–K relations reveal the causal linkage between power-law attenuation and dispersion in the analytical results. In the direct application to experimental data the link may be somewhat obscured as the finite-bandwidth artifacts can be substantial. However, with the analytical expressions available in this paper it is possible to conceive of methods for correcting these results, perhaps within the context of a general model-independent approach to finite-bandwidth K–K calculations.

ACKNOWLEDGMENTS

This work is supported in part by NIH Grant No. R37 HL40302. Dr. Mobley is also supported by an appointment to the Oak Ridge National Laboratory Postdoctoral Research Associates Program administered jointly by the Oak Ridge Institute for Science and Education and Oak Ridge National Laboratory. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Dept. of Energy under Contract No. DE-AC05-00OR22725.

[1] J. Mobley, K. R. Waters, M. S. Hughes, C. S. Hall, J. N. Marsh, G. H. Brandenburger, and J. G. Miller, "Kramers–Kronig relations applied to finite bandwidth data from suspensions of encapsulated microbubbles," J. Acoust. Soc. Am. **108**(5), 2091–2106 (2000); **112**(2), 760–761(E) (2002).

[2] K. R. Waters, M. S. Hughes, J. Mobley, G. H. Brandenburger, and J. G. Miller, "On the applicability of Kramers–Kronig relations for ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**(2), 556–563 (2000).

[3] T. L. Szabo, "Causal theories and data for acoustic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **97**, 14–24 (1995).

[4] F. A. Duck, *Physical Properties of Tissue: A Comprehensive Reference Book* (Academic, New York, 1990), p. 112.

[5] *Tissue Substitutes, Phantoms, and Computational Modeling in Medical Ultrasound, ICRP Report No. 61* (International Commission on Radiation Units and Measurements, Bethesda, MD, 1998), p. 8.

[6] H. M. Nussenveig, *Causality and Dispersion Relations* (Academic, New York, 1972), pp. 28–33.

[7] J. Mobley, K. R. Waters, C. S. Hall, J. N. Marsh, M. S. Hughes, G. H. Brandenburger, and J. G. Miller, "Measurements and predictions of the phase velocity and attenuation coefficient in suspensions of elastic microspheres," J. Acoust. Soc. Am. **106**(2), 652–659 (1999).

Open cycle traveling wave thermoacoustics: Mean temperature difference at the regenerator interface

Nathan T. Weiland[a)]
School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Ben T. Zinn
School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 18 April 2003; revised 10 August 2003; accepted 8 September 2003)

In an open cycle traveling wave thermoacoustic engine, the hot heat exchanger is replaced by a steady flow of hot gas into the regenerator to provide the thermal energy input to the engine. The steady-state operation of such a device requires that a potentially large mean temperature difference exist between the incoming gas and the solid material at the regenerator's hot side, due in part to isentropic gas oscillations in the open space adjacent to the regenerator. The magnitude of this temperature difference will have a significant effect on the efficiencies of these open cycle devices. To help assess the feasibility of such thermoacoustic engines, a numerical model is developed that predicts the dependence of the mean temperature difference upon the important design and operating parameters of the open cycle thermoacoustic engine, including the acoustic pressure, mean mass flow rate, acoustic phase angles, and conductive heat loss. Using this model, it is also shown that the temperature difference at the regenerator interface is approximately proportional to the sum of the acoustic power output and the conductive heat loss at this location. [DOI: 10.1121/1.1621859]

PACS numbers: 43.35.Ud, 44.27.+g [RR] Pages: 2791–2798

I. INTRODUCTION

Research in thermoacoustics has been limited mostly to closed cycle designs in recent years, though there is a potential for significant improvements in efficiency by using open cycle thermoacoustic devices. In such systems, a slow mean flow is superimposed on the acoustic field in order to replace one of the heat exchangers with a convective heat transfer process.[1,2] This requires, however, that the mean flow velocity be an order of magnitude smaller than the acoustic velocity in order to keep thermal convection effects from overwhelming the thermoacoustic effects in the device. In most cases, an additional thermodynamic benefit is gained in that the open cycle thermoacoustic process is very efficient in converting the thermal energy in the convected fluid into acoustic energy, or vice versa. This was the major impetus for the construction and testing of the first open cycle thermoacoustic device, Los Alamos's standing wave refrigerator,[1–3] which cools a stream of gas as it passes through the stack. Greater commercial potential exists, however, for open cycle traveling wave thermoacoustic devices, as they possess an efficiency advantage over their inherently irreversible standing wave counterparts.[4] For instance, in the natural gas liquefier being developed jointly by Los Alamos National Lab and Praxair Inc.,[5] the natural gas that is burned to power the liquefier could be routed directly through the thermoacoustic engine in an open cycle configuration, thereby eliminating the hot heat exchanger and providing a simpler and possibly more efficient means of converting fuel energy into acoustic energy.

A diagram of a basic open cycle traveling wave thermoacoustic engine is shown in Fig. 1, where the slow mean flow of hot gas approaches the regenerator from the right. Heat is removed at the cold heat exchanger on the left side of the regenerator, setting up a temperature gradient through which an acoustic traveling wave is amplified in passing from the cold end of the regenerator to the hot end. However, as will be shown in the following, the joining conditions at the interface between the regenerator and the adjacent open duct require that a substantial mean temperature difference exist between the incoming mean flow and the regenerator's solid material. Since this may have profound effects on the acoustics, thermodynamics, and efficiency of the open cycle thermoacoustic engine, the remainder of this study will focus on developing a model that can be used to predict the magnitude of this temperature difference.

The physical processes responsible for creating the temperature difference at the regenerator/open duct interface are shown schematically for various gas parcels in Fig. 2, which can be viewed as a magnification of the control volume surrounding the regenerator interface in Fig. 1. Gas parcels that start the acoustic cycle at various locations within the regenerator are shown in Figs. 2(a)–(d). For good gas–solid thermal contact within the regenerator, these parcels of gas remain at the regenerator's temperature until they enter the open duct, where they undergo isentropic temperature oscillations in concert with the pressure oscillations of the traveling acoustic wave. These oscillations cause the gas parcels to re-enter the regenerator at temperatures below that of the regenerator,[6] where they receive heat from the regenerator solid and are rapidly heated back up to the temperature of the regenerator at the interface, T_{re}. At the end of the acoustic cycle, fresh gas parcels enter the regenerator for the first time due to the steady mean flow of gas toward the regenerator, as

a)Electronic mail: gte852f@mail.gatech.edu

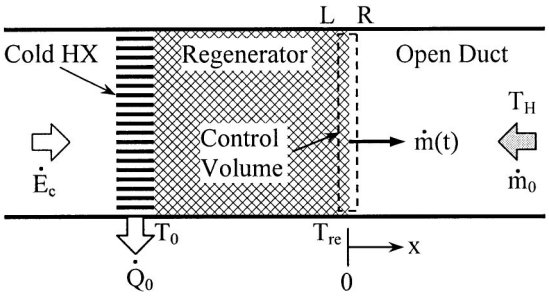


FIG. 1. A schematic of a basic open cycle traveling wave thermoacoustic engine. A steady flow of gas with mean temperature T_H flows into the regenerator, whose hot end is at the temperature T_{re}. A cold heat exchanger rejects heat to ambient temperature at the other end of the regenerator, and the resulting temperature gradient in the regenerator is used to amplify the acoustic traveling wave entering its cold end. An analysis of the control volume, which includes the interface between the regenerator and the open duct, is used to determine the difference between the temperatures T_H and T_{re}. State properties on the left and right of the control volume are labeled with subscripts "*L*" and "*R*," respectively.

depicted in Figs. 2(e) and (f). These gas parcels isentropically oscillate about their mean temperature, T_H, until they enter the regenerator, at which time they are cooled to T_{re} by contact with the regenerator's solid material. In the absence of other sources of heat input or output, a heat balance of the regenerator solid at the interface shows that the heat transferred from the solid to the cold returning gas must equal the heat transferred from the fresh gas to the regenerator solid, if the solid material is to maintain a constant mean temperature during steady-state operation. Satisfying this criterion gives rise to the mean temperature difference between the regenerator solid and the hot incoming gas, as cited earlier.

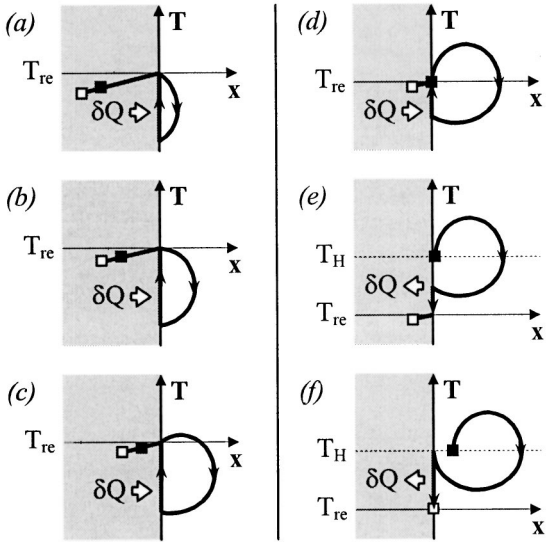


FIG. 2. A schematic of the temperature and acoustic displacement histories of various gas parcels near the regenerator/open duct interface, over the course of an acoustic cycle. The closed squares denote the starting position of each parcel, and the open squares denote the final position. The net displacement between the starting and finishing positions is due to the mean mass flux. Gas parcels in (a)–(d) start their motion inside the regenerator, undergo isentropic temperature oscillations outside the regenerator that reduce their temperatures, and are heated to the regenerator's temperature, T_{re}, upon re-entry. Gas parcels in (e) and (f) start their motions outside the regenerator, oscillate isentropically about the mean temperature, T_H, and cool down to T_{re} upon entering the regenerator at the end of the acoustic cycle.

The effect of a temperature difference within or between thermoacoustic components has been a subject of several studies in thermoacoustics. This topic was first broached by Swift,[7] who pointed out that for net solid–gas heat transfer to occur within a heat exchanger, a difference must exist between the solid temperature and the spatially and temporally averaged gas temperature within the heat exchanger. Brewster *et al.*[8] extended this concept to include the heat transfer between a thermoacoustic stack and its heat exchangers, though their analysis was limited only to standing wave thermoacoustic devices. Research that is more closely related to the problem at hand was first performed by Smith and Romm[6] for component interface losses in Stirling engines, and later on by Kittel[9] and Bauwens[10] for similar interface losses in pulse tube refrigerators. Swift[11] subsequently applied these concepts to the general field of thermoacoustics, and has developed relationships that are frequently used to estimate losses and joining conditions between components in thermoacoustic devices. However, none of these studies consider the effects of mean flow on the component interface dynamics, thus the development of a new theoretical framework is required for studying the component interface conditions in open cycle thermoacoustic devices.

It should also be noted that while this study deals primarily with open cycle, traveling wave thermoacoustic engines, the problem of determining the temperature difference at the regenerator/open duct interface is also of concern in thermoacoustic refrigerators and heat pumps of the same type, as this temperature difference critically impacts the efficiencies of all of these devices. However, this paper effectively demonstrates the procedure involved in the determination of this temperature difference by focusing solely on the traveling wave engine application.

II. MASS FLUX MODEL

The formulation used to model the open cycle thermoacoustic engine follows the formulation developed by Smith and Romm,[6] with some modifications to account for the presence of mean flow. The model considers conditions within a control volume that encloses the hot side interface of the regenerator, as shown in Fig. 1. The control volume's left and right boundaries, designated with the subscripts *L* and *R*, are located inside the solid matrix of the regenerator and in the adjoining open duct, respectively, and are close enough together that the control volume can be assumed to contain a negligible amount of mass. To simplify the analysis, it is assumed that the flow is one-dimensional, and that there is no axial conduction or mixing between the parcels of gas that move axially in and out of the control volume.

A. Mass flux

To begin, the mass flux passing through the control volume is assumed to be given by

$$\dot{m}(t) = \dot{m}_0 + \dot{m}_1 \sin(\omega t + \phi), \tag{1}$$

where $\dot{m}_0$ and $\dot{m}_1$ are the magnitudes of the mean and oscillating components of the mass flux, respectively, ω is the angular frequency of the acoustic oscillations, and ϕ is a

phase shift used to set a reference time below. Having assumed that the control volume contains a negligible amount of mass, it can be further assumed that the mass fluxes crossing the left and right sides of the control volume are equal (i.e., $\dot{m}_L = \dot{m}_R$).

It is convenient to nondimensionalize the model equations in order to eliminate the dependence on the angular frequency. Introducing a dimensionless time, $\hat{t} \equiv \omega t$, and dimensionless mass fluxes, $\hat{m} \equiv \dot{m}/\dot{m}_1$ and $\hat{m}_0 \equiv \dot{m}_0/\dot{m}_1$, Eq. (1) can be rewritten as

$$\hat{m}(\hat{t}) = \hat{m}_0 + \sin(\hat{t} + \phi). \tag{2}$$

Note that nondimensionalizing the mass fluxes by $\dot{m}_1$ instead of $\dot{m}_0$ is more practical for consideration of the limiting case in which there is no mean mass flux.

By choosing to let $\hat{m} = 0$ at $\hat{t} = 0$, the phase shift becomes

$$\phi = \sin^{-1}(-\hat{m}_0). \tag{3}$$

Here, Eq. (3) specifies that $|\hat{m}_0| \leq 1$, though for practical purposes, we require $|\hat{m}_0| \leq 0.1$ to satisfy the condition that the mean velocity be an order of magnitude smaller than the oscillating velocity in an open cycle thermoacoustic device.[2] Further, this study will only consider the processes that occur during one acoustic cycle, i.e., $0 \leq \hat{t} \leq 2\pi$, where $\hat{m} = 0$ at $\hat{t} = 0$ and $\hat{t} = 2\pi$, according to Eqs. (2) and (3).

An important parameter in this study is the time, $\hat{t}_{\text{mid}}$, at which the gas ceases to flow out of the regenerator, reverses direction, and begins to flow back into the regenerator. Noting that $\hat{m}(\hat{t}_{\text{mid}}) = 0$, that $\hat{t}_{\text{mid}}$ should be close to π in value, and that the sine function has odd symmetry about π, Eqs. (2) and (3) can be used to show that

$$\hat{t}_{\text{mid}} = \pi - 2\phi. \tag{4}$$

B. Total mass

In later sections, it will be necessary to relate the time at which a particular gas parcel leaves the regenerator to the time at which it re-enters the regenerator. This is accomplished by tracking the total mass, m, that has passed from the regenerator to the open duct during the acoustic cycle. Integrating Eq. (1) and nondimensionalizing yields

$$\hat{m}(\hat{t}) = \hat{m}_0 \hat{t} - \cos(\hat{t} + \phi) + \cos(\phi), \tag{5}$$

where the total dimensionless mass has been defined as $\hat{m} \equiv m\omega/\dot{m}_1$.

The mass flux and total mass displacement, as described by Eqs. (2) and (5), are plotted in Fig. 3, where three distinct time periods within the acoustic cycle can be identified. The first, denoted by the subscript "*a*," describes the time period $0 \leq \hat{t}_a \leq \hat{t}_{\text{mid}}$ when the gas is flowing from left to right out of the regenerator. At the time $\hat{t}_{\text{mid}}$, the gas reverses direction and starts flowing back to the left, thus the total mass that has entered the open duct since the start of the acoustic cycle, $\hat{m}$, is a maximum at this time. The second time period is defined as the time during which gas parcels that left the regenerator return to the regenerator. Since a stratified flow has been assumed, the time at which a representative gas parcel leaves the regenerator, $\hat{t}_a$, can be related to the time at which the same parcel re-enters the regenerator, $\hat{t}_b$, by equating the total mass displacement at these times as shown in Fig. 3, i.e., $\hat{m}(\hat{t}_a) = \hat{m}(\hat{t}_b)$. Using Eq. (5), $\hat{t}_b$ can thus be related to $\hat{t}_a$ by solving

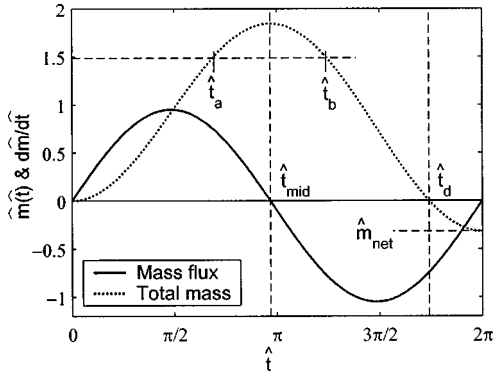


FIG. 3. Time dependence of the dimensionless mass flux and total mass displacement during an acoustic cycle in which $\hat{m}_0 = -0.05$. Three distinct time periods are identified by the dashed vertical lines. In the first time period, gas flows to the right, leaving the regenerator, until it comes to a stop at time $\hat{t}_{\text{mid}}$. In the second time period, that same gas flows back to the left, returning to the regenerator. The time at which a particular gas parcel re-enters the regenerator, $\hat{t}_b$, is linked to the time at which it leaves the regenerator, $\hat{t}_a$, by the total mass displacement. In the third time period, after all of the original gas has returned to the regenerator at time $\hat{t}_d$, fresh gas flows into the regenerator until the end of the acoustic cycle. The net mass displacement for the acoustic cycle, $\hat{m}_{\text{net}}$, is also shown.

$$\hat{m}_0 \hat{t}_a - \cos(\hat{t}_a + \phi) = \hat{m}_0 \hat{t}_b - \cos(\hat{t}_b + \phi), \tag{6}$$

where $\hat{t}_b$ falls between $\hat{t}_{\text{mid}}$ and a time $\hat{t}_d$, which marks the start of the third time period.

In the stratified flow assumption, two parcels of gas that start the acoustic cycle adjacent to each other, one inside the regenerator and one outside, can have vastly different states, thus the regenerator interface experiences a discontinuity in gas properties as the gas in the open duct flows from right to left through the control volume at time $\hat{t}_d$, the "discontinuity" time. This time corresponds to the time at which $\hat{m} = 0$ again, and can be found by solving

$$\hat{m}(\hat{t}_d) = \hat{m}_0 \hat{t}_d - \cos(\hat{t}_d + \phi) + \cos(\phi) = 0 \tag{7}$$

for $\hat{t}_d$, where $\hat{t}_{\text{mid}} < \hat{t}_d < 2\pi$. This time marks the beginning of the third time period, denoted by the subscript "*c*," $\hat{t}_d \leq \hat{t}_c \leq 2\pi$, during which fresh gas parcels enter the regenerator for the first time due to the presence of a mean flow. At the end of the acoustic cycle, a net mass of

$$\hat{m}(\hat{t} = 2\pi) \equiv \hat{m}_{\text{net}} = 2\pi\hat{m}_0 \tag{8}$$

has been transported into the regenerator as a result of the mean mass flux as depicted in Fig. 3.

III. THERMODYNAMIC MODEL

Next, the thermodynamic properties of the gas need to be determined on each side of the control volume for all times during the acoustic cycle. To facilitate this analysis, it is assumed that the pressure in the control volume does not depend on axial location, and is given by

$$\hat{p}(\hat{t}) = 1 + \hat{p}_1 \sin(\hat{t} + \phi + \theta), \tag{9}$$

where $\hat{p} \equiv p/p_0$ and $\hat{p}_1 \equiv p_1/p_0$. Here, p_0 is the mean pressure, p_1 is the amplitude of the acoustic pressure, and θ is the phase angle by which the acoustic pressure leads the acoustic mass flux, where attention is restricted to $-\pi/2 < \theta < \pi/2$ in this work, since acoustic energy flux is positive according to the sign conventions of this study.

A. Gas temperatures

The temperature and position histories of various gas parcels near the regenerator interface are shown schematically in Fig. 2 above, although it is not necessary to calculate these histories, as the only temperatures of interest in this study are those of the gas parcels as they enter and leave the control volume. To simplify matters, it is assumed that the gas properties can be related by the ideal gas equation of state, and that there is perfect thermal contact between the gas and the solid within the regenerator (i.e., $T_L = T_{\mathrm{re}}$ for all $\hat{t}$). Since accounting for the transient heating effects in the regenerator would require a finite-width control volume, thermal conductivity and specific heat properties of the gas, and details on the regenerator's internal geometry, neglecting these effects greatly simplifies the model at the expense of a slightly diminished accuracy.

During the first time period, gas exits the regenerator at temperature T_{re} and crosses the right-hand boundary of the control volume, so $T_{R,a} = T_{\mathrm{re}}$. Upon exiting the regenerator, gas parcels undergo isentropic temperature oscillations as a result of the acoustic pressure oscillations in the open duct to the right of the regenerator, as depicted in Figs. 2(a)–(d). For an isentropic process in an ideal gas, the pressure and temperature are related by

$$\frac{T}{T_{\mathrm{ref}}} = \left(\frac{p}{p_{\mathrm{ref}}} \right)^{(\gamma-1)/\gamma}, \tag{10}$$

where γ is the ratio of specific heats. The reference conditions for a particular parcel of gas correspond to the temperature and pressure of the gas when it first encounters the isentropic environment upon exiting the regenerator at time $\hat{t}_a$. The time at which this gas parcel returns to the regenerator is determined by solving Eq. (6) for $\hat{t}_b$, and the dimensionless temperature of the gas parcel at this time, $\hat{T}_{R,b}$, is then found from Eq. (10):

$$\hat{T}_{R,b} = \left[\frac{1 + \hat{p}_1 \sin(\hat{t}_b + \phi + \theta)}{1 + \hat{p}_1 \sin(\hat{t}_a + \phi + \theta)} \right]^{(\gamma-1)/\gamma}, \tag{11}$$

where the temperature has been nondimensionalized with respect to the regenerator temperature (i.e., $\hat{T} \equiv T/T_{\mathrm{re}}$).

The temperature of the incoming gas during the third time period can be similarly obtained, though the reference states for temperature and pressure must be redefined. In the open duct, the gas temperature isentropically oscillates about a mean temperature, T_H, as a result of the pressure oscillations, as depicted in Figs. 2(e) and (f). In the course of these oscillations, $T = T_H$ when $p = p_0$, so taking T_H and p_0 as the reference states in Eq. (10) for the gas parcels that begin the acoustic cycle inside the open duct, the temperature at the time the gas parcels first enter the regenerator, $\hat{T}_{R,c}$, can be expressed as

$$\hat{T}_{R,c} = \hat{T}_H [1 + \hat{p}_1 \sin(\hat{t}_c + \phi + \theta)]^{(\gamma-1)/\gamma}. \tag{12}$$

B. Thermal energy balance

Having determined the temperature of the gas crossing each of the control volume's boundaries, the heat balance within the control volume can now be examined. Assuming that kinetic energy and viscous effects are negligible, a quasi-one-dimensional energy equation can be written as

$$\frac{\partial(\rho e)}{\partial t} + \frac{\partial(\rho u h)}{\partial x} + \nabla \cdot \mathbf{q} = 0, \tag{13}$$

where ρ is the gas density, e is the internal energy, u is the gas velocity in the x direction, h is the enthalpy, and $\mathbf{q}$ is the heat flux vector. Using the ideal gas law and the state relationships $e = h - p/\rho$ and $dh = c_p dT$, the first term in Eq. (13) can be written as: $\partial(\rho e)/\partial t = (\partial p/\partial t)/(\gamma - 1)$.

Integrating over the control volume then yields

$$\frac{V}{\gamma - 1} \frac{dp}{dt} + [\dot{m} h]_R - [\dot{m} h]_L + \int \nabla \cdot \mathbf{q} \, dV = 0. \tag{14}$$

Assuming that the periphery of the device is well-insulated, the heat flux vector, $\mathbf{q}$, consists of two primary components, axial heat conduction and lateral heat transfer between the gas and the regenerator's solid material. Using the divergence theorem and Fourier's law for the conductive heat flux yields

$$\frac{V}{\gamma - 1} \frac{dp}{dt} + \left[\dot{m} h - Ak \frac{dT}{dx} \right]_R - \left[\dot{m} h - Ak \frac{dT}{dx} \right]_L = \dot{Q}, \tag{15}$$

where A is the cross-sectional area occupied by the gas, k is the thermal conductivity of the gas, and $\dot{Q}$ is generically defined as the heat transfer rate between the regenerator solid and the gas in the control volume, where heat transfer to the gas is positive. This equation can be further simplified by applying the above assumption of an infinitely thin control volume, which eliminates the first term on the left-hand side of Eq. (15). Finally, if thermal conduction effects in the gas are assumed to be negligible, Eq. (15) can be written nondimensionally as

$$\hat{Q} = \hat{m}(\hat{T}_R - 1), \tag{16}$$

where the dimensionless heat flux is defined as $\hat{Q} \equiv \dot{Q}/\dot{m}_1 c_p T_{\mathrm{re}}$.

Although thermal conduction in the gas has been neglected, conduction in the solid of the regenerator may still be important, as the thermal conductivities of solids are generally orders of magnitude larger than those for gases, and the solid material constitutes a significant volume fraction of the regenerator. To account for this conductive heat flux, a thermal energy balance of the regenerator solid is required. Assuming that heat transfer to the gas and conduction through the left side of the control volume are the only heat fluxes in or out of the solid, this energy balance can be written nondimensionally as

$$\frac{d\hat{E}_s}{d\hat{t}} = \hat{Q}_k - \hat{Q}, \tag{17}$$

where $\hat{E}_s \equiv \omega E_s / \dot{m}_1 c_p T_{re}$, in which E_s is the thermal energy of the solid in the control volume, and $\hat{Q}_k \equiv -A_s k_s (dT/dx)_{int} / \dot{m}_1 c_p T_{re}$, where A_s is the cross-sectional area of the solid, k_s is the thermal conductivity of the solid, and $(dT/dx)_{int}$ is temperature gradient of the solid at the regenerator interface.

In steady state operation, the solid material of the regenerator maintains a constant temperature, T_{re}, thus there must be no net energy change in the solid material of the regenerator over the course of an acoustic cycle, i.e.,

$$\int_0^{2\pi} \frac{d\hat{E}_s}{d\hat{t}} d\hat{t} = \int_0^{2\pi} \hat{Q}_k d\hat{t} - \int_{\hat{t}_{mid}}^{\hat{t}_d} \hat{Q}_b d\hat{t}_b - \int_{\hat{t}_d}^{2\pi} \hat{Q}_c d\hat{t}_c = 0. \tag{18}$$

In this expression, the gas–solid heat transfer integral has been split into its contributions from the second and third time periods, $\hat{Q}_b$ and $\hat{Q}_c$, respectively, while the contribution from the first time period has been neglected since no heat is transferred between the gas and the regenerator solid during this time. Substituting Eqs. (11) and (12) for $\hat{T}_{R,b}$ and $\hat{T}_{R,c}$, respectively, into Eq. (16), expressions for the heat fluxes during the second and third time periods are obtained:

$$\hat{Q}_b(\hat{t}_b) = \hat{m}(\hat{t}_b) \left(\left[\frac{\hat{p}(\hat{t}_b)}{\hat{p}(\hat{t}_a)} \right]^{(\gamma-1)/\gamma} - 1 \right), \tag{19}$$

$$\hat{Q}_c(\hat{t}_c) = \hat{m}(\hat{t}_c) (\hat{T}_H [\hat{p}(\hat{t}_c)]^{(\gamma-1)/\gamma} - 1). \tag{20}$$

In each of these equations, the mass fluxes and pressures are given (for the appropriate times) by Eqs. (2) and (9), respectively, and $\hat{t}_a$ in Eq. (19) is related to $\hat{t}_b$ by Eq. (6).

IV. THE TEMPERATURE DIFFERENCE

Substituting Eqs. (19) and (20) into the heat balance of Eq. (18) yields, after some manipulation, an equation for the ratio of the mean temperature of the incoming gas to the temperature of the regenerator solid:

$$\hat{T}_H = \frac{\int_{\hat{t}_d}^{2\pi} \hat{m} d\hat{t}_c + \int_{\hat{t}_{mid}}^{\hat{t}_d} \hat{m} \left(1 - \left[\frac{\hat{p}(\hat{t}_b)}{\hat{p}(\hat{t}_a)} \right]^{(\gamma-1)/\gamma} \right) d\hat{t}_b + 2\pi \hat{Q}_k}{\int_{\hat{t}_d}^{2\pi} \hat{m} [\hat{p}(\hat{t}_c)]^{(\gamma-1)/\gamma} d\hat{t}_c}. \tag{21}$$

The integrations in Eq. (21) are performed numerically on a PC to determine the value of $\hat{T}_H$ that satisfies the heat balance. In an actual open cycle thermoacoustic engine, the mean flow gas temperature would be a known, fixed quantity, and the steady-state heat balance of Eq. (18) would be solved to determine the temperature at the hot face of the regenerator. Regardless, Eq. (21) solves for the ratio between the two temperatures, so knowing one temperature allows for the solution of the other temperature.

A plot of the calculated time history of the temperature at the right-hand side of the control volume is shown for representative choices of independent parameters in Fig. 4(a). The right-side temperature, T_R, is equal to the regenerator temperature during the first time period, and is lower than the regenerator temperature during the second time period. This is consistent with the results of Smith and Romm,[6] who have shown that if work is generated by the engine (i.e., $-\pi/2 < \theta < \pi/2$), then the isentropic oscillations outside the regenerator cause the gas to re-enter the regenerator at lower temperatures. At the end of the second time period, $\hat{t}_d$, a discontinuity in temperature occurs as the fresh, hotter gas enters the regenerator for the first time. The plot of the gas–solid heat transfer rate versus time in Fig. 4(b) shows the added effect of the mass flux, and illustrates the heat balance described in Eq. (18), where the area under the curves during the second and third time periods must equal one another in the absence of a conductive heat flux. The various curves in Figs. 4(a) and (b) show the effects of varying mean mass flow rates on the temperature and heat flux histories, which will be discussed further in a later section.

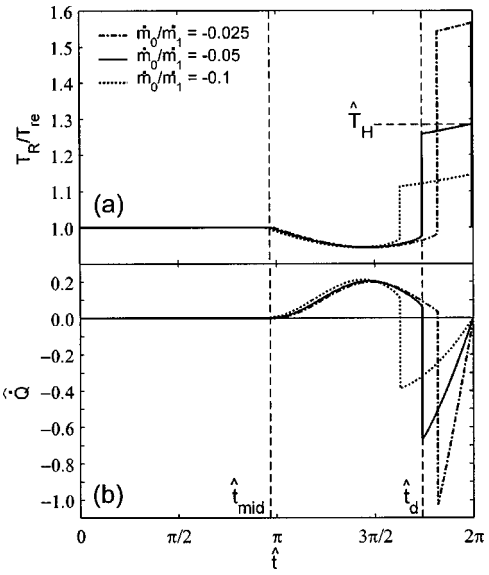


FIG. 4. Time dependence of (a) the dimensionless temperature on the right side of the control volume and (b) the dimensionless heat flux from the regenerator solid, during an acoustic cycle for the conditions: $\hat{p}_1 = 0.1$, $\gamma = 1.4$, $\theta = 0$, and $\hat{Q}_k = 0$. The mean incoming gas temperature and relevant time periods are marked for $\hat{m}_0 = -0.05$ (solid line), which correspond to the same time periods shown in Fig. 3. For comparison, the temperature and heat flux histories for $\hat{m}_0 = -0.025$ and $\hat{m}_0 = -0.1$ are also shown.

A. Temperature difference approximation

Predicting the mean temperature difference with Eq. (21) is fairly cumbersome, and is not very practical for quick design calculations. Several approximations can be made to this equation to yield a simpler expression that reasonably estimates the temperature difference, $\hat{T}_H$, and provides a clearer understanding of how it is affected by various independent parameters. First, by expanding the pressure term in the denominator of Eq. (21) in a binomial series and assuming that the acoustic pressure amplitude is much smaller than the mean pressure (i.e., $\hat{p}_1 \ll 1$), the denominator of Eq. (21) can be approximated by

$$\int_{\hat{t}_d}^{2\pi} \hat{m} [\hat{p}(\hat{t}_c)]^{(\gamma-1)/\gamma} d\hat{t}_c \approx \int_{\hat{t}_d}^{2\pi} \hat{m} d\hat{t}_c = 2\pi \hat{m}_0, \tag{22}$$

where the last integral is equal to the net mass displacement for the entire acoustic cycle (see Fig. 3). Substitution of Eq. (22) into Eq. (21) for the temperature difference yields

$$\hat{T}_H \approx 1 + \frac{1}{2\pi \hat{m}_0} \int_{\hat{t}_{\text{mid}}}^{\hat{t}_d} \hat{m} \left(1 - \left[\frac{\hat{p}(\hat{t}_b)}{\hat{p}(\hat{t}_a)} \right]^{(\gamma-1)/\gamma} \right) d\hat{t}_b + \frac{\hat{Q}_k}{\hat{m}_0}, \tag{23}$$

where Eq. (22) has also been applied to the first term in the numerator of Eq. (21).

To simplify the pressure term within the integral in Eq. (23), the pressure ratio can be expanded in a Taylor series with the use of Eq. (9), and the resulting expression can be further expanded in a binomial series to account for the pressure ratio's exponent. Extracting the θ dependence from this result and using Eq. (6) yields

$$\begin{aligned} 1 - \left[\frac{\hat{p}(\hat{t}_b)}{\hat{p}(\hat{t}_a)} \right]^{(\gamma-1)/\gamma} \approx{}& \frac{\gamma-1}{\gamma} \hat{p}_1 \hat{m}_0 (\hat{t}_a - \hat{t}_b) \sin\theta \\ &+ \frac{\gamma-1}{\gamma} \hat{p}_1 \cos\theta [\sin(\hat{t}_a + \phi) \\ &- \sin(\hat{t}_b + \phi)], \end{aligned} \tag{24}$$

where all terms that are second order and higher in the acoustic pressure amplitude have been neglected. This expression can be further simplified by assuming that the mean mass flux is much smaller than the acoustic mass flux, i.e., $\hat{m}_0 \ll 1$, which eliminates the first term on the right-hand side of Eq. (24). Substituting Eqs. (2) and (24) into Eq. (23) and eliminating terms of the order of $\hat{m}_0$ from the integral yields

$$\begin{aligned} \hat{T}_H \approx{}& 1 + \frac{\hat{Q}_k}{\hat{m}_0} + \frac{\gamma-1}{\gamma} \frac{\hat{p}_1 \cos\theta}{2\pi \hat{m}_0} \int_{\hat{t}_{\text{mid}}}^{\hat{t}_d} \sin(\hat{t}_b + \phi) \\ &\times [\sin(\hat{t}_a + \phi) - \sin(\hat{t}_b + \phi)] d\hat{t}_b, \end{aligned} \tag{25}$$

where the integral now contains only variables that are a function of $\hat{m}_0$. In the limiting case in which $\hat{m}_0$ approaches zero, it can be shown that: $\phi \to 0$, $\hat{t}_a \to 2\pi - \hat{t}_b$, $\hat{t}_{\text{mid}} \to \pi$, and $\hat{t}_d \to 2\pi$.

Using these limits to evaluate the integral in Eq. (25) yields a compact expression for the estimated temperature difference:

$$\hat{T}_H \approx 1 - \frac{\gamma-1}{\gamma} \frac{\hat{p}_1 \cos\theta}{2\hat{m}_0} + \frac{\hat{Q}_k}{\hat{m}_0}. \tag{26}$$

Note that in this estimate, $\hat{T}_H > 1$, since $\hat{m}_0$ and $\hat{Q}_k$ are always negative by the sign conventions used in this study, and $-\pi/2 \leqslant \theta \leqslant \pi/2$.

Given the number of approximations made in arriving at this result, Eq. (26) works remarkably well as an estimate for the true temperature difference given in Eq. (21), with errors of less than 5% for $\hat{p}_1 \leqslant 0.1$, $\gamma \leqslant 1.67$, $-\pi/2 \leqslant \theta \leqslant \pi/2$, $-0.1 \leqslant \hat{m}_0 \leqslant -0.0001$, and $\hat{Q}_k \leqslant 0$. Even for higher acoustic pressure amplitudes (e.g., $\hat{p}_1 \leqslant 0.3$), Eq. (26) approximates the temperature difference to within about 10% of its true value.

To take the estimation of the temperature difference a step further, note that the acoustic energy at the hot side of the regenerator can be approximated by

$$\dot{E}_{re} \approx \frac{p_1 \dot{m}_1 \cos\theta}{2\rho_{\text{re}}}. \tag{27}$$

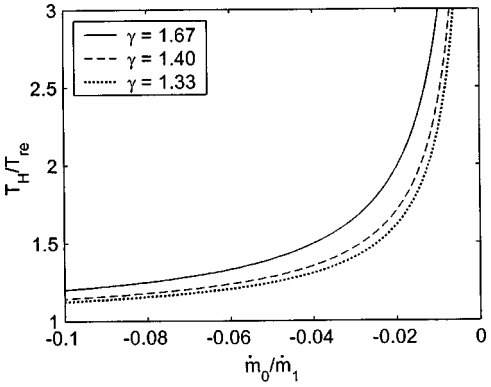


FIG. 5. The dependence of the temperature ratio of the incoming gas and the regenerator solid upon the mean to oscillating mass flux ratio and the ratio of specific heats, γ, for the conditions: $\hat{p}_1 = 0.1$, $\theta = 0$, and $\hat{Q}_k = 0$. Note that the negative mean mass flux indicates gas flow toward the regenerator.

By using Eq. (27), the ideal gas law, and the definitions of the dimensionless variables, Eq. (26) can be rearranged to produce the following interesting result:

$$\dot{m}_0 c_p (T_{\text{re}} - T_H) \approx \dot{E}_{\text{re}} + A_s k_s \left(\frac{dT}{dx} \right)_{\text{int}}. \tag{28}$$

Equation (28) indicates that a substantial temperature difference must exist across the regenerator interface if the engine is going to produce a significant amount of acoustic power. A more broad interpretation of Eq. (28) states that of the thermal energy convected to the regenerator interface at temperature T_H, some of it is convected into the regenerator at temperature T_{re}, some of it is conducted through the regenerator solid, and the remainder is used to create the acoustic energy exiting the regenerator. This acoustic energy is not generated by the processes occurring at the regenerator interface, but rather, this term represents a ''thermoacoustic heat flux'' into the regenerator. Note that in an ideal regenerator of a traveling wave thermoacoustic device, that the acoustic energy flux traveling in one direction is accompanied by a ''thermoacoustic heat flux'' of equal magnitude traveling in the opposite direction.[11] The thermal energy carried by this thermoacoustic ''heat-pumping'' effect, as it is also called, is converted into work in the form of acoustic energy within the regenerator. Therefore, in accordance with Eq. (28), a larger temperature difference at the regenerator interface results in a larger ''thermoacoustic heat flux'' into the regenerator, consequently increasing the acoustic power that is generated with this thermal energy.

B. Effects on the temperature difference

The above-presented analysis shows that the temperature difference between the gas and regenerator depends upon the following five dimensionless parameters: the ratio of mean to oscillating mass fluxes, $\hat{m}_0$, the ratio of acoustic to mean pressures, $\hat{p}_1$, the phase angle by which oscillating pressure leads oscillating mass flux, θ, the ratio of specific heats, γ, and the thermal conduction loss in the regenerator, $\hat{Q}_k$. The parameter with the largest effect on the temperature difference is $\hat{m}_0$, as shown in Fig. 5. As the magnitude of mean mass flux decreases relative to the magnitude of the

acoustic mass flux, $\hat{T}_H$ becomes fairly large. The functional relationship between the temperature difference and the mean mass flux is shown in Eq. (26), where the temperature difference is a function of the inverse of the mean mass flux.

This dependence on the mean mass flux can be explained by considering its effect on the discontinuity time, $\hat{t}_d$, given in Eq. (7). A reduction in $\hat{m}_0$ increases $\hat{t}_d$, thus decreasing the fraction of the acoustic cycle during which the fresh gas enters the regenerator. This effect is depicted in the heat flux plot of Fig. 4(b), where the time at which the discontinuity occurs for $\hat{m}_0=-0.1$ is earlier than the discontinuity time for $\hat{m}_0=-0.05$. As shown in the figure, the total heat transferred to the gas during the second time period (the area under the curve) is only slightly increased by the increased limits of integration. The primary effect of decreasing the mean mass flux and increasing the discontinuity time is to decrease the time allotted for heat transfer from the fresh gas to the regenerator solid. To compensate, the mean temperature of the incoming gas must increase so that the net heat input from the gas (the area under the curve for the third time period) equals that of the heat output from the solid, as expressed in the heat balance in Eq. (18).

As Fig. 5 illustrates, the mean temperature difference between the gas in the regenerator and the open duct is not necessarily small, and can be much larger than the other types of temperature differences noted in the thermoacoustics literature to date.[6–11] This is primarily a result of the mean flow's role as the heat source, whereas other analyses have only investigated temperature differences associated with the use of heat exchangers. Temperature differences of the magnitudes seen here can have a profound effect on the efficiency of the engine, although Eq. (28) would suggest that it may not necessarily be advantageous to minimize this temperature difference, as it is directly linked to the acoustic power output of the engine.

Figure 5 also describes the effect of the ratio of specific heats, γ, on the magnitude of the temperature difference as expressed in Eq. (21). These effects are the result of the influence of γ on the isentropic relationship between temperature and pressure outside the regenerator, as described by Eq. (10). As Eq. (26) shows, the ratio of specific heats has a smaller effect upon the temperature difference than the other independent parameters, however, it is the only one of the five independent parameters that depends upon the gas properties within the engine. As such, monatomic gases ($\gamma\approx1.67$) result in higher temperature differences than nonmonatomic gases. For nonmonatomic gases the mean gas temperature plays a small role, as gaseous combustion products ($T\approx2000\,\mathrm{K}$), for which $\gamma\approx1.33$, will yield slightly lower temperature differences according to Fig. 5, than will ambient-temperature nonmonatomic gases, for which $\gamma\approx1.4$.

The dependence of the temperature difference upon the acoustic phase lag, θ, and the acoustic pressure magnitude, $\hat{p}_1$, is shown in Fig. 6. Consistent with the cos θ dependence predicted in Eq. (26), the acoustic phase shift produces the largest temperature differences near $\theta=0$, corresponding to traveling wave acoustic phasing. This condition maximizes the difference in a gas parcel's temperature from exit to return to the regenerator, which increases $\hat{Q}_b$, the heat flux

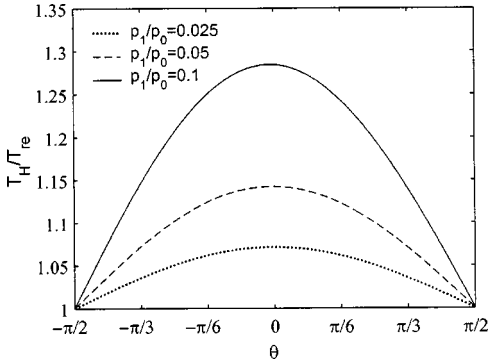


FIG. 6. The dependence of the temperature ratio of the incoming gas and the regenerator solid upon the phase angle by which oscillating pressure leads the oscillating mass flux, θ, and the acoustic to mean pressure ratio, $\hat{p}_1$, for the conditions: $\hat{m}_0=-0.05$, $\gamma=1.4$, and $\hat{Q}_k=0$.

from the regenerator to the incoming gas, and requires that $\hat{Q}_c$ and $\hat{T}_H$ increase correspondingly. The temperature difference decreases away from $\theta=0$, until no temperature difference exists for standing wave acoustic phasing at $\theta=\pm\pi/2$. Since the acoustic displacement, pressure, and temperature are all in phase in a standing wave, a gas parcel will undergo isentropic oscillations outside of the regenerator but will return to the regenerator at the same temperature and pressure as when it left.[11] The addition of mean flow introduces slight phase differences between temperature and displacement over the course of an acoustic cycle, but the resulting heat fluxes cancel one another for $\theta=\pm\pi/2$.

Figure 6 also shows the approximately linear relationship between the acoustic pressure magnitude and the temperature difference of Eq. (21), with higher acoustic pressures resulting in higher temperature differences. This relationship is predicted by Eq. (26) and can be linked to the Taylor series expansion in Eq. (24), where the temperature of the gas returning to the regenerator during the second part of the acoustic cycle, $\hat{T}_{R,b}$, is shown to be approximately proportional to $\hat{p}_1$ for small values of the acoustic pressure magnitude. Therefore, the total heat flux from the solid to the gas during the second part of the acoustic cycle is proportional to $\hat{p}_1$, which is in turn approximately proportional to the temperature difference required to maintain the heat balance. Recognizing the link between the temperature difference and the heat flux across the regenerator's interface, the results in Fig. 6 for both the pressure and phase angle are in qualitative agreement with the results of Smith and Romm.[6]

Finally, Fig. 7 shows the combined effects of the conductive loss term, $\hat{Q}_k$, and the mean mass flux on the temperature difference. While the mean mass flux effect is the same as that shown in Fig. 5, the addition of a conductive heat loss increases the required temperature difference at the regenerator's interface. For a given mean mass flux, this dependence is shown to be approximately linear, as predicted by Eq. (26) for the estimated temperature difference. This behavior is best explained in conjunction with Eq. (18), which states that the total energy change in the regenerator solid over the course of an acoustic cycle is equal to the heat transfer from the gas to the solid during the third time period, minus the heat transferred from the solid to the gas during

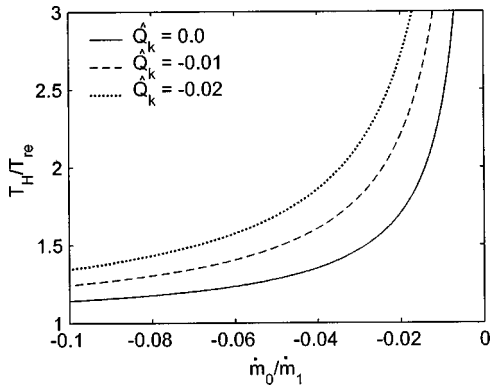


FIG. 7. The dependence of the temperature ratio of the incoming gas and the regenerator solid upon the mean to oscillating mass flux ratio and the dimensionless conductive heat loss, $\hat{Q}_k$, for the conditions: $\hat{p}_1=0.1$, $\theta=0$, and $\gamma=1.4$.

the second time period, minus the heat leaving the control volume via thermal conduction. Therefore, for increasing heat conduction losses, an increasing heat input from the third time period is required (which is proportional to the temperature difference), in order to enforce the condition that there be zero net energy change in the regenerator's solid from one acoustic cycle to the next. As such, minimizing the conductive heat leak increases the portion of the thermal energy input available for conversion to acoustic energy.

V. CONCLUSIONS

The theoretical analysis developed in this study allows the determination of the temperature difference that occurs at the regenerator interface in open cycle traveling wave thermoacoustic engines, when mean flow replaces the hot heat exchanger as a means for supplying heat to the engine. This temperature difference is shown to be a function of five independent dimensionless parameters, and is well approximated by Eq. (26) above. With slight modifications, the theoretical framework developed in this study could also be used to evaluate open cycle designs for traveling wave thermoacoustic refrigerators and heat pumps. The temperature differences predicted by these analyses could significantly affect the performance of these open cycle devices, and may positively or negatively affect their feasibility as alternatives to their closed cycle counterparts. Previous studies on thermoacoustic temperature differences typically associate them with loss mechanisms, where minimizing them generally improves the performance of the thermoacoustic device. This is not found to be the case in the open cycle thermoacoustic engine, however, since reducing the temperature difference has the direct effect of reducing the acoustic power output of the engine, according to Eq. (28).

Predicting the existence of this temperature difference is the first step in the evaluation of the feasibility of open cycle traveling wave thermoacoustic devices, and suggests several subsequent avenues of research. For instance, the losses that result from the heat transferred across the potentially large temperature difference could be analyzed and compared to the losses in the hot heat exchanger of a closed cycle thermoacoustic engine. Also, the effects of this temperature difference on the theoretical efficiencies of open cycle devices could be studied and compared to those of closed cycle devices. Finally, given the potential for increasing the efficiencies of thermoacoustic devices by employing open cycle configurations, these analyses need to be verified in the laboratory to determine the accuracy of the assumptions and predictions that are made in these studies.

ACKNOWLEDGMENT

The authors would like to thank Greg Swift for his helpful discussion of this topic.

[1] R. S. Reid, W. C. Ward, and G. W. Swift, "Cyclic thermodynamics with open flow," Phys. Rev. Lett. **80**, 4617–4620 (1998).
[2] R. S. Reid and G. W. Swift, "Experiments with a flow-through thermoacoustic refrigerator," J. Acoust. Soc. Am. **108**, 2835–2842 (2000).
[3] R. S. Reid, "Open cycle thermoacoustics," Ph.D. thesis, Georgia Institute of Technology, School of Mechanical Engineering, 1999.
[4] S. Backhaus and G. W. Swift, "A thermoacoustic-Stirling heat engine: Detailed study," J. Acoust. Soc. Am. **107**, 3148–3166 (2000).
[5] J. J. Wollan, G. W. Swift, S. Backhaus, and D. L. Gardner, "Development of a thermoacoustic natural gas liquefier," AIChE Meeting, New Orleans, LA, 11–14 March 2002. Available at: http://www.lanl.gov/thermoacoustics/Pubs/index.html.
[6] J. L. Smith and M. Romm, "Thermodynamic loss at component interfaces in Stirling cycles," Proceedings of the 27th Intersociety Energy Conversion Engineering Conference, San Diego, CA, Society of Automotive Engineers, 1992, pp. 5.529–5.532.
[7] G. W. Swift, "Analysis and performance of a large thermoacoustic engine," J. Acoust. Soc. Am. **92**, 1551–1563 (1992).
[8] J. R. Brewster, R. Raspet, and H. E. Bass, "Temperature discontinuities between elements of thermoacoustic devices," J. Acoust. Soc. Am. **102**, 3355–3360 (1997).
[9] P. Kittel, "The temperature profile within pulse tubes," Adv. Cryog. Eng. **43**, 1927–1932 (1998).
[10] L. Bauwens, "Interface loss in the small amplitude orifice pulse tube model," Adv. Cryog. Eng. **43**, 1933–1940 (1998).
[11] G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America, Melville, NY, 2002).

Vibration of prolate spheroidal shells with shear deformation and rotatory inertia: Axisymmetric case

Sabih I. Hayek[a)]
Active Vibration Control Laboratory, Department of Engineering Science and Mechanics, Penn State University, 212 Earth and Engineering Sciences Building, University Park, Pennsylvania 16802-6812

Jeffrey E. Boisvert
Naval Undersea Warfare Center, Newport, Rhode Island 02841

(Received 31 July 2002; revised 15 July 2003; accepted 11 August 2003)

This paper presents the derivation of the equations for nonaxisymmetric motion of prolate spheroidal shells of constant thickness. The equations include the effect of distributed mechanical surface forces and moments. The shell theory used in this derivation includes three displacements and two thickness shear rotations. Thus, the effects of membrane, bending, shear deformation, and rotatory inertia are included in this theory. The resulting five coupled partial differential equations are self-adjoint and positive definite. The frequency-wave-number spectrum has five branches, two acoustic and three optical branches representing flexural, extensional, torsional, and two thickness shear. For the case of axisymmetric motion, these were computed for various spheroidal shell eccentricities and thickness-to-length ratios for a large number of modes. The axisymmetric dynamic response for damped shells of various eccentricities and thicknesses under point and ring surface forces are presented. [DOI: 10.1121/1.1616925]

PACS numbers: 43.40.Ey [JHG] Pages: 2799–2811

I. INTRODUCTION

Prolate spheroidal shells (PSS) are shells of revolution that can span a wide range of shapes, from spherical to needle-like. They are essentially an ellipsoid of revolution about its major axis and may have a constant thickness or a variable thickness defined by two confocal ellipses. The ellipse is characterized by an interfocal length$=d$ and eccentricity$=1/a$, "a" being the radial coordinate of the shell's mid-surface in the prolate spheroidal coordinate system. In the limits $d\to 0$, $a\to\infty$, and $(da/2)\to R$, the ellipsoidal surface approaches a spherical surface whose radius is R.

DiMaggio and Silbiger[1] were the first to consider the free extensional axisymmetric torsional vibrations of a PSS of variable thickness. The equations of motion were based on the Lagrangian of the shell using Hamilton's variational principle. The exact solution was found in terms of the angular prolate spheroidal wave function $S_{0n}(\eta)$. Silbiger and DiMaggio[2] developed the equations for axisymmetric extensional-flexural vibration of PSS of a variable thickness. Using Rayleigh–Ritz method, they obtained the resonance frequencies and mode shapes, but the solution was limited to low eccentricity of the shell $(1/a=0.7)$. Shiraishi and DiMaggio[3] obtained the solution for the same by use of perturbation techniques. The perturbation parameter was the eccentricity $(1/a)$ and their perturbation series was extended to $O(1/a^4)$. It should be noted that the zero order term of the pertubation solution is the spherical shell resonances and mode shapes. Again these frequencies were valid for low eccentricity $(1/a=0.7)$. Nemergut and Brand[4] developed the extensional-flexural vibration of a constant-thickness shell using a numerical integration scheme. DiMaggio and Rand[5] obtained the solution of variable-thickness flexural-extensional vibration by finite difference techniques. Rand[6] obtained the exact solution for axisymmetric torsional vibration of solid prolate spheroids and prolate spheroidal shells using the exact theory of elasticity.

Burroughs and Magrab[7] were the first to include bending and shear deformation in the shell equations using a variational approach in terms of shear, normal and moment resultants. They developed five coupled equations for nonaxisymmetric motion for a shell with constant thickness. However, they obtained solutions for only the axisymmetric vibration of the nontorsional branches using Galerkin's method, using sinusoidal comparison functions. They showed that for axisymmetric vibrations, the resonance frequencies for the flexural modes, increase with (h/R), while the longitudinal modes are not influenced by the addition of bending or shear deformations. There is a new branch representing thickness-shear mode, which was shown to be inversely proportional to (h/R). Chen and Ginsberg[8] revisited the problem by focusing on the loci of the eigenvalues of three PSS for the axisymmetric vibration, but they included extensional and bending terms only and they used a Ritz series formulation in their solution, employing sinusoidal comparison functions.

The first attempt at examining the fluid loading effects on the vibration of submerged PSS was made by Hayek and DiMaggio.[9,10] The complex natural frequencies of submerged PSS using membrane strain energy only were found for shells of variable thickness using perturbation techniques on the strain energy density and the surface acoustic pressure. Rand and DiMaggio[11] examined the resonance frequencies of extensional-flexural vibration of fluid-filled PSS. The presence of a fluid inside the shell increases the number of resonance frequencies because of the three-dimensionality of

a)Author to whom correspondence should be addressed; electronic mail: sihesm@engr.psu.edu

the cavity resonances of the interior of the PSS. Yen and DiMaggio[12] obtained the solution for the forced vibration of submerged PSS due to axisymmetric time-harmonic surface forces. They used numerical integration techniques after transforming the infinite exterior acoustic medium domain to a finite domain. Bedrosian and DiMaggio[13] obtained the transient solution of an axisymmetric extensional vibration of a submerged PSS undergoing sudden application of a uniform surface source at $t=0$.

Berger[14] was the first to include bending effects in the problem of forced vibration of a submerged PSS of constant thickness. He used a finite difference approach for the shell equations and mapped the infinite region of the exterior acoustic medium into a finite region. He presented the response of the shell due to a normal transient loading. Lee and DiMaggio[15] obtained the response of a fluid-filled PSS as a model of a human head. They included bending theory in their model and obtained the response of the PSS due to harmonic surface sources. Ross and Johns[16] obtained the solution for the axisymmetric vibration of free and submerged (both sides) of a hemi-PSS and conducted experiments to support their analytic predictions. Prikhod'ko[17] obtained the solution for a slender submerged PSS, but the solution is valid for sufficiently slender shells in order to neglect the fluid loading on the shell ends. Pauwelussen[18] used FEM to solve the problem of axisymmetric response of submerged PSS due to a shock load. Chen and Ginsberg[19] obtained the acoustic radiation for PSS including bending effects due to axisymmetric mechanical surface forces.

Acoustic scattering from PSS has not been treated as well as vibration and acoustic radiation from PSS. Chertock *et al.*[20] obtained the flexural response of PSS due to an incident underwater explosion and tested their theory with experiments on a parallelepiped box-shell. Silbiger[21] studied the scattering from PSS due to an incident time-harmonic plane wave by approximating the response due to one resonant mode. Jones-Oliveira[22,23] obtained the transient scattering from an axially incident plane wave, i.e., axisymmetric response, from a PSS that includes extensional and bending effects using expansions in terms of Legendre polynomials. In addition, the first eight resonant frequencies for the flexural and extensional branches of two different PSS were determined.

II. EQUATIONS OF MOTION OF PROLATE SPHEROIDAL SHELLS

In this paper, we have employed the prolate spheroidal coordinate system found in Flammer.[24] Basically, these coordinates represent ellipsoidal or hyperboloidal surfaces [see Fig. 1(a)]. The coordinate system transformation from Cartesian coordinates are given by:

$$x=\frac{d}{2}[(1-\eta^2)(\xi^2-1)]^{1/2}\cos\phi, \quad -1\leqslant\eta\leqslant 1,$$

$$y=\frac{d}{2}[(1-\eta^2)(\xi^2-1)]^{1/2}\sin\phi, \quad 1\leqslant\xi\leqslant\infty,$$

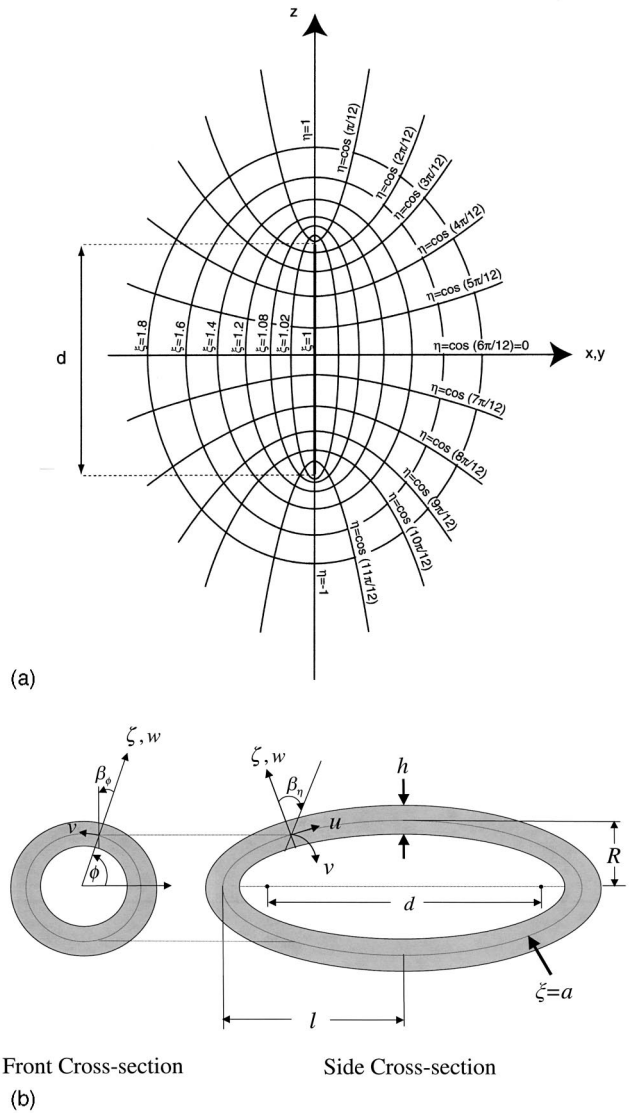


FIG. 1. (a) Prolate spheroidal coordinate system. (b) Prolate spheroidal shell geometry.

$$z=\frac{d}{2}\eta\xi, \quad 0\leqslant\phi\leqslant 2\pi,$$

where d represents the interfocal distance of the generating ellipse.

It should be noted that in the limit $\xi\to\infty$, $d\to 0$, the prolate spheroidal coordinates reduce to spherical coordinates, i.e., $d\xi/2\to r$, $\eta\to\cos\theta$, as $d\to 0$ and $\xi\to\infty$.

The theory of thin curvilinear shells depends on assumptions made in the theory of elastic media. Specifically, it is assumed that the shell is thin, and deformations are small. The theories of thin shells are developed in many books and special higher order theories are given in research reports and archival papers. Of the former, there are complete treatments of this subject in Refs. 25 and 26.

In this study, the equations of motion of the vibration of shells are derived from the Lagrangian, **L** using Hamilton's variational principle:

$$\mathbf{L}=-\mathbf{U}+\mathbf{K}+\mathbf{X}, \tag{2.1}$$

where **U** is the strain energy of the shell, **K** is the kinetic energy of the shell, and **X** is the potential for the work done by distributed external surface forces. The potential and kinetic energy densities were derived using kinematic variables defined by Kraus.[25] These were then integrated over the surface of the spheroidal shell in terms of a Taylor series in the thickness. For a consistent theory of thin shells, only terms of order h and h^3 were retained and higher order terms were neglected throughout. It should be noted that Kraus[25] applies variational methods on the strain energy density in terms of force and moment resultants. This could lead to non-self-adjoint and non-positive-definite system of partial differential equations (PDE). The use of variational methods on the Lagrangian in terms of the normal and shear strains, changes in curvatures, and thickness shear angles instead of the force and moment resultants guarantees that the resulting PDE are self-adjoint and positive-definite, resulting in an elastic impedance matrix that is symmetric.

Consider a prolate spheroidal shell of constant thickness, h, whose midsurface is defined by $\xi=a>1$ and interfocal distance, d, as shown in Fig. 1(b). The mid-surface of the shell has a half-length, $l=da/2$, and a maximum radius, $R=d\sqrt{a^2-1}/2$, so that the ratio of maximum diameter/length is $\sqrt{a^2-1}/a$, and the eccentricity is $e=1/a$.

The displacement field can be written as:

$$\begin{aligned} U(\zeta,\eta,\phi)&=u(\eta,\phi)+\zeta\beta_\eta(\eta,\phi),\\ V(\zeta,\eta,\phi)&=v(\eta,\phi)+\zeta\beta_\phi(\eta,\phi),\\ W(\eta,\phi)&=w(\eta,\phi), \end{aligned} \tag{2.2}$$

where U and V are the displacements in the tangential direction along the η and ϕ directions, and W is the displacement normal to the midsurface of the shell. Here, the independent variables u, v, and w are the displacements of a point on the midsurface of the shell and β_η and β_ϕ are the rotations of the normal to the midsurface. To nondimensionalize the independent displacement variables let

$$\bar{u}=u/l, \quad \bar{v}=v/l, \quad \bar{w}=w/l.$$

Using variational methods on the Lagrangian, Hayek and Boisvert[27] derived the five coupled self-adjoint PDE on the five dynamic variables that can be symbolically written as

$$\begin{aligned} &L_{uu}\bar{u}+L_{uv}\bar{v}+L_{uw}\bar{w}+L_{u\beta_\eta}\beta_\eta+L_{u\beta_\phi}\beta_\phi\\ &\quad=\frac{l^2}{c_p^2}\left\{\left(1+\epsilon\frac{a^4}{C^2}\right)\frac{\sqrt{AC}}{a^2}\ddot{\bar{u}}+\epsilon(1+D^2)\ddot{\beta}_\eta\right\}\\ &\qquad-\frac{1-\nu^2}{Eh}l\frac{\sqrt{AC}}{a^2}q_u, \end{aligned} \tag{2.3}$$

$$\begin{aligned} &L_{vu}\bar{u}+L_{vv}\bar{v}+L_{vw}\bar{w}+L_{v\beta_\eta}\beta_\eta+L_{v\beta_\phi}\beta_\phi\\ &\quad=\frac{l^2}{c_p^2}\left\{\left(1+\epsilon\frac{a^4}{C^2}\right)\frac{\sqrt{AC}}{a^2}\ddot{\bar{v}}+\epsilon(1+D^2)\ddot{\beta}_\varphi\right\}\\ &\qquad-\frac{1-\nu^2}{Eh}l\frac{\sqrt{AC}}{a^2}q_v, \end{aligned} \tag{2.4}$$

$$\begin{aligned} &L_{wu}\bar{u}+L_{wv}\bar{v}+L_{ww}\bar{w}+L_{w\beta_\eta}\beta_\eta+L_{w\beta_\phi}\beta_\phi\\ &\quad=\frac{l^2}{c_p^2}\frac{\sqrt{AC}}{a^2}\left(1+\epsilon\frac{a^4}{C^2}\right)\ddot{\bar{w}}-\frac{1-\nu^2}{Eh}l\frac{\sqrt{AC}}{a^2}q_w, \end{aligned} \tag{2.5}$$

$$\begin{aligned} &L_{\beta_\eta u}\bar{u}+L_{\beta_\eta v}\bar{v}+L_{\beta_\eta w}\bar{w}+L_{\beta_\eta\beta_\eta}\beta_\eta+L_{\beta_\eta\beta_\phi}\beta_\phi\\ &\quad=\frac{l^2}{c_p^2}\epsilon\left\{\frac{\sqrt{AC}}{a^2}\ddot{\beta}_\eta+(1+D^2)\ddot{\bar{u}}\right\}-\frac{1-\nu^2}{Eh}\frac{\sqrt{AC}}{a^2}m_{\beta_\eta}, \end{aligned} \tag{2.6}$$

$$\begin{aligned} &L_{\beta_\phi u}\bar{u}+L_{\beta_\phi v}\bar{v}+L_{\beta_\phi w}\bar{w}+L_{\beta_\phi\beta_\eta}\beta_\eta+L_{\beta_\phi\beta_\phi}\beta_\phi\\ &\quad=\frac{l^2}{c_p^2}\epsilon\left\{\frac{\sqrt{AC}}{a^2}\ddot{\beta}_\varphi+(1+D^2)\ddot{\bar{v}}\right\}-\frac{1-\nu^2}{Eh}\frac{\sqrt{AC}}{a^2}m_{\beta_\varphi}, \end{aligned} \tag{2.7}$$

where $c_p^2=E/\rho_s(1-\nu^2)$ is the square of the extensional plate velocity, E is Young's modulus, ρ_s is the shell density, ν denotes Poisson's ratio, $\epsilon=h^2/12l^2$ is the bending coefficient, $A=a^2-1$,

$$\begin{aligned} &B=1-\eta^2, \quad C=a^2-\eta^2, \quad D=\frac{\sqrt{a^2-1}}{\sqrt{a^2-\eta^2}},\\ &F=\frac{1-\nu}{2}, \quad \gamma=3/7, \quad \kappa=\pi^2/12, \end{aligned} \tag{2.8}$$

and $(\cdot)$ indicates partial differentiation with time. In Eqs. (2.3)–(2.7), q_u, q_v, and q_w are the distributed surface forces along u, v, and w, and m_{β_η} and m_{β_ϕ} are distributed surface moments. The expressions for the operators L are given in the Appendix.

III. GALERKIN EXPANSION FOR NONAXISYMMETRIC VIBRATION

Since there is no closed-form orthogonal set of eigenfunctions for the PSS, one must use a set of nonorthogonal basis functions that satisfy the boundary conditions. However, the modal decomposition in the circumferential coordinate is exact.

The nondimensional displacement field can thus be expanded in terms of basis functions satisfying continuity conditions at the apexes of the closed spheroidal shell, i.e.,

$$\begin{aligned} &\bar{u}(\pm1,\phi)=0, \quad \bar{v}(\pm1,\phi)=0, \quad \bar{w}(\pm1,\phi) \text{ bounded},\\ &\beta_\eta(\pm1,\phi)=0, \quad \text{and } \beta_\phi(\pm1,\phi)=0. \end{aligned}$$

For the general case of nonaxisymmetric vibration, circular functions in the circumferential coordinate ϕ and associated Legendre functions in the angular coordinate η were chosen as the basis functions, since they are the ones used for the limiting case of a spherical shell. The displacement field is expressed in terms of appropriate basis functions that satisfy the continuity conditions as follows:

$$\bar{u}=\sum_{m=0}^{\infty}\sum_{n=m+1}^{\infty} U_{mn}P_n^{m+1}(\eta)\cos m\phi,$$

$$\bar{v}=\sum_{m=0}^{\infty}\sum_{n=m+1}^{\infty} V_{mn}P_n^{m+1}(\eta)\sin m\phi,$$

$$\bar{w}=\sum_{m=0}^{\infty}\sum_{n=m}^{\infty} W_{mn}P_n^{m}(\eta)\cos m\phi, \tag{3.1}$$

$$\beta_\eta=\sum_{m=0}^{\infty}\sum_{n=m+1}^{\infty} D_{mn}P_n^{m+1}(\eta)\cos m\phi,$$

$$\beta_\phi=\sum_{m=0}^{\infty}\sum_{n=m+1}^{\infty} E_{mn}P_n^{m+1}(\eta)\sin m\phi.$$

A harmonic time dependence $e^{i\omega t}$ will be assumed throughout this paper. One must truncate the series in (3.1) so that N nontrivial basis functions are used in each series. Substituting the expansion in (3.1) into Eqs. (2.3)–(2.7), one would obtain five coupled ordinary differential equations for the mth modal components. Each equation is an infinite series in terms of the basis functions. Multiplying the first and fourth equations by $P_k^{m+1}(\eta)\cos m\phi$, the second and fifth equations by $P_k^{m+1}(\eta)\sin m\phi$, and the third by $P_k^m(\eta)\cos m\phi$, $k=m,m+1,...$, and integrating over η from -1 to $+1$, and integrating over ϕ from 0 to 2π, one obtains $5N\times 5N$ algebraic equations on the $5N$ displacement coefficients. These equations take the form

$$\sum_{n=m+1}^{N+m+1}\{p_1(m,k,n)U_{mn}+q_1(m,k,n)V_{mn}+r_1(m,k,n)W_{mn}+s_1(m,k,n)D_{mn}+t_1(m,k,n)E_{mn}+\Omega^2 xp_1(m,k,n)U_{mn}$$

$$+\Omega^2 xs_1(m,k,n)D_{mn}\}=-\frac{(1-\nu^2)l}{2\pi Eha^2}\epsilon_m\int_{-1}^{+1}\int_0^{2\pi} q_u(\eta,\varphi)\sqrt{AC}P_k^{m+1}(\eta)\cos(m\varphi)d\varphi\,d\eta, \tag{3.2}$$

$$\sum_{n=m+1}^{N+m+1}\{p_2(m,k,n)U_{mn}+q_2(m,k,n)V_{mn}+r_2(m,k,n)W_{mn}+s_2(m,k,n)D_{mn}+t_2(m,k,n)E_{mn}+\Omega^2 xq_2(m,k,n)V_{mn}$$

$$+\Omega^2 xt_2(m,k,n)E_{mn}\}=-\frac{(1-\nu^2)l}{2\pi Eha^2}\epsilon_m\int_{-1}^{+1}\int_0^{2\pi} q_v(\eta,\varphi)\sqrt{AC}P_k^{m+1}(\eta)\sin(m\varphi)d\varphi\,d\eta, \tag{3.3}$$

$$\sum_{n=m}^{N+m}\{p_3(m,k,n)U_{mn}+q_3(m,k,n)V_{mn}+r_3(m,k,n)W_{mn}+s_3(m,k,n)D_{mn}+t_3(m,k,n)E_{mn}+\Omega^2 xr_3(m,k,n)W_{mn}\}$$

$$=-\frac{(1-\nu^2)l}{2\pi Eha^2}\epsilon_m\int_{-1}^{+1}\int_0^{2\pi} q_w(\eta,\varphi)\sqrt{AC}P_k^{m}(\eta)\cos(m\varphi)d\varphi\,d\eta, \tag{3.4}$$

$$\sum_{n=m+1}^{N+m+1}\{p_4(m,k,n)U_{mn}+q_4(m,k,n)V_{mn}+r_4(m,k,n)W_{mn}+s_4(m,k,n)D_{mn}+t_4(m,k,n)E_{mn}+\Omega^2 xp_4(m,k,n)U_{mn}$$

$$+\Omega^2 xs_4(m,k,n)D_{mn}\}=-\frac{(1-\nu^2)l}{2\pi Eha^2}\epsilon_m\int_{-1}^{+1}\int_0^{2\pi} m_{\beta_\eta}(\eta,\varphi)\sqrt{AC}P_k^{m+1}(\eta)\cos(m\varphi)d\varphi\,d\eta, \tag{3.5}$$

$$\sum_{n=m+1}^{N+m+1}\{p_5(m,k,n)U_{mn}+q_5(m,k,n)V_{mn}+r_5(m,k,n)W_{mn}+s_5(m,k,n)D_{mn}+t_5(m,k,n)E_{mn}+\Omega^2 xq_5(m,k,n)V_{mn}$$

$$+\Omega^2 xt_5(m,k,n)E_{mn}\}=-\frac{(1-\nu^2)l}{2\pi Eha^2}\epsilon_m\int_{-1}^{+1}\int_0^{2\pi} m_{\beta_\varphi}(\eta,\varphi)\sqrt{AC}P_k^{m+1}(\eta)\sin(m\varphi)d\varphi\,d\eta. \tag{3.6}$$

The nondimensional frequency parameter is defined by $\Omega=\omega l/c_p$. The coefficients p, q, r, s, t are the elastic components of these equations, and the coefficients xp, xq, xr, xs, xt are the inertial terms.[28]

IV. NUMERICAL RESULTS—AXISYMMETRIC VIBRATION

For axisymmetric motion of PSS, $\partial/\partial\phi=0$, then the five coupled PDE's (2.3)–(2.7), (3.2)–(3.6) and in the Appendix

TABLE I. Comparison of the first eight frequencies of the five branches for a spherical shell.

	Nontorsional								Torsional			
	Flexural			Extensional			Non-tor. th. sh.		Torsional		Tors. th. sh.	
n	a	b	c	a	b	c	a	b	a	b	a	b
0	None	None	None	1.612	1.611	1.612	18.707	18.746	None	None	18.656	18.746
1	10^{-3}	4×10^{-7}	0.0	1.970	1.969	1.976	18.846	18.815	10^{-3}	0.00	18.694	18.764
2	0.708	0.708	0.708	2.711	2.711	2.726	19.050	18.953	1.181	1.181	18.751	18.802
3	0.871	0.870	0.873	3.620	3.619	3.641	19.318	19.157	1.867	1.867	18.826	18.858
4	1.003	1.003	1.014	4.577	4.576	4.605	19.643	19.423	2.505	2.505	18.920	18.933
5	1.173	1.173	1.202	5.551	5.551	5.587	20.023	19.748	3.124	3.124	19.031	19.026
6	1.400	1.400	1.462	6.534	6.533	6.576	20.453	20.127	3.734	3.734	19.161	19.137
7	1.682	1.682	1.798	7.520	7.519	7.569	20.927	20.555	4.339	4.338	19.308	19.266
8	2.013	2.014	2.208	8.508	8.507	8.565	21.443	21.028	4.940	4.939	19.472	19.412

[a]Present study, spheroidal shell $a=100$, $l/R=1.000\,05$, $h/l=0.1$.
[b]Results from Wilkinson, Ref. 29 for spherical shell, $h/R=0.1$.
[c]Results from Hayek, Ref. 30 for spherical shell, $h/R=0.1$, without shear deformation.

decouple into three coupled equations on u, w, and β_η and two coupled equations on v and β_ϕ. It can be seen that the expressions in (A2), (A5) in the first equation, (A12) and (A15) in the third equation and (A17) and (A20) in the fourth equation vanish, rendering the first, third and fourth equations dependent on u, w, and β_η exclusively.

Furthermore, for axisymmetric motion, expressions in (A6), (A8), and (A9) in the second equation, (A21), (A23), and (A24) in the fifth equation vanish. This means that the second and fifth equations are exclusively dependent on v and β_ϕ. This will cause a decoupling of the $5N\times5N$ algebraic equations into a $3N\times3N$ nontorsional system on U_{0n}, W_{0n}, and D_{0n}, and a $2N\times2N$ torsional system on V_{0n} and E_{0n}.

The resonance frequencies are computed for various values of "a" and various thickness to half-length ratios (h/l). These were obtained using a root search method on the $3N\times3N$ determinant for nontorsional vibration, and on the $2N\times2N$ determinant for torsional vibration. The numerical results are plotted in the form of frequency spectra, Ω vs n. To determine convergence of the resonance frequencies, the roots were calculated using $N=20$, 40, 80, 160, and 320 basis function sets. For this study, adequate convergence was found using $N=80$ basis functions, with a frequency resolution of 0.001.

A. Comparison with previous studies

To assess the accuracy of the numerical results of this study, one compares first the results of this study of a spheroidal shell with $a=100$ (i.e., length/diameter$=1.000\,05$) to the results for a spherical shell of Wilkinson[29] and Hayek.[30] Wilkinson's work included membrane, bending and thickness shear based on Naghdi's theory of thin shells based on Newtonian formulation of the stress resultants. However, the rotatory inertias for the thickness shear included terms h^3 and h^5, inconsistent with the rest of his terms, which were limited to h and h^3. He also used a shear modulus correction factor $\kappa=5/6$ instead of the more accurate Mindlin correction of $\kappa=\pi^2/12$. Hayek's[30] work includes only membrane and bending terms and uses Lagrangian approach on the strain energy density in terms of kinematical variables. Comparison of the numerical results of this study with Wilkinson's and Hayek's exact values for $h/l=0.1$ and $\nu=0.3$ is shown in Table I. Examination of the first eight nondimensional frequencies for the flexural, extensional, and torsional branches with Wilkinson's are accurate to within 0.05%. Comparisons with Hayek's results show differences from 0% to 4% for the first six resonances of the flexural branch. This branch is the most influenced by bending, rotatory inertia, and thickness shear. Comparison of the first eight resonances of the exten-

TABLE II. Comparison of the first eight frequencies of the flexural, extensional, and nontorsional thickness shear branches for a spheroidal shell, $a=1.4142$, $h/l=0.035\,36$.

	Flexural		Extensional		Non-tors. th. sh.	
n	a	b	a	b	a	b
0	None	None	1.979	2.828	52.612	53.5
1	0.001	0.000	2.647	3.125	52.684	53.6
2	1.043	1.032	3.423	3.380	52.786	53.9
3	1.289	1.287	4.358	4.313	52.920	54.0
4	1.386	1.386	5.410	5.346	53.086	54.3
5	1.473	1.485	6.518	6.435	53.282	54.6
6	1.567	1.598	7.650	7.552	53.509	55.0
7	1.672	1.725	8.793	8.683	53.767	55.3

[a]Present study.
[b]Burroughs and Magrab, Ref. 7, $\Omega=a\Omega_{(\text{Ref. 7})}$, $h/l=(h/a)_{(\text{Ref. 7})}/a$.

TABLE III. Comparison of the first eight frequencies of the flexural, extensional, and nontorsional thickness shear branches for a spheroidal shell, $a=1.4142$, $h/l=0.003\,536$.

	Flexural		Extensional		Non-tors. th. sh.	
n	a	b	a	b	a	b
0	None	None	1.979	2.86	525.625	529
1	0.001	0.000	2.648	3.01	525.632	529
2	1.043	1.032	3.425	3.38	525.643	529
3	1.278	1.273	4.362	4.31	525.656	529
4	1.336	1.343	5.415	5.35	525.673	529
5	1.358	1.358	6.524	6.45	525.692	529
6	1.376	1.414	7.657	7.55	525.716	529
7	1.395	1.457	8.801	8.68	525.742	529

[a]Present study.
[b]Burroughs and Magrab, Ref. 7, $\Omega=a\Omega_{(\text{Ref. 7})}$, $h/l=(h/a)_{(\text{Ref. 7})}/a$.

sional branch shows differences of 0% to 0.7% indicating that the extensional branch is not sensitive to thickness shear deformation. In fact, Hayek[30] pointed out that it also is not sensitive to the bending terms for the first 20 modes. Comparisons of the nontorsional and torsional thickness shear frequencies show differences of 0.01–1.9 %. This difference is due to the addition of the thickness shear rotatory inertia terms of order h^5 in Wilkinson's equations.

Numerical results of this study are compared to those of Burroughs and Magrab[7] for the shape parameter $a=1.4142$ for $h/l=0.035\,36$ in Table II and for $h/l=0.003\,536$ in Table III, for $\nu=0.3$. It should be noted that $\Omega=a\Omega_{(\text{Ref. 7})}$ and $h/l=(h/a)_{(\text{Ref. 7})}/a$. Numerical results in Ref. 7 are given for nontorsional branches using $N=15$ comparison functions that are based on sinusoidal functions. Comparison of the resonances of the flexural branches shows excellent agreement. Except for poor agreement in the first two modes of the extensional branch ($n=0$ and 1), the agreement with the remaining six resonances is excellent. Comparison of the resonances of the thickness shear branch show differences of up to 3% due to the smaller number of comparison functions used (15 vs 80 in this study).

In order to track down the differences for first two modes of the extensional branch, comparison is made of the results of this study with two other studies by Chen and Ginsberg[8] and Jones-Oliviera[23] in Table IV for $a=1.4142$, $h/l=0.027\,45$, and $\nu=0.3$. These studies involved a theory of shells that included membrane and bending terms only. One can see that the frequencies of the extensional branch have excellent agreement for the first six modes, as is the case for the extensional branch for spherical shells in Table I for Wilkinson's[29] and Hayek's[30] results.

Further comparison of our results with those of Jones-Oliviera[22] and Chen and Ginsberg[8] is shown in Table V for a long shell with $a=1.005$, $h/l=0.000\,6744$, and $\nu=0.3$. There is an excellent agreement with Chen and Ginsberg in the low frequency range for the first six modes, due to the higher number of comparison functions used in their work ($N=30$) vs those of Jones–Oliviera ($N=8$).

B. Frequency spectra

For each spheroidal shell, represented by the geometric spheroidal shape parameter "a" and the thickness to half-length ratio h/l, the frequency spectrum has five branches; three nontorsional branches being the flexural, the extensional, and the nontorsional thickness shear, and two torsional branches being the torsional and the torsional thickness shear. The flexural and torsional branches are acoustic branches, in that the starting point of the branch is $\Omega=0$ and $n=0$. For the flexural branch however, the $n=0$ mode does not exist, hence the first resonance is $\Omega=0$ at $n=1$. The extensional and thickness shear branches are optical branches in that the $n=0$ mode has a nonzero frequency,

TABLE IV. Comparison of the first nine frequencies of the flexural and extensional branches for a spheroidal shell, $a=1.4142$, $h/l=0.027\,45$.

	Flexural			Extensional		
n	a	b	c	a	b	c
0	None	None	None	1.979	1.979	1.979
1	0.001	0.001	–	2.648	2.646	2.648
2	1.043	1.042	1.043	3.424	3.423	3.425
3	1.285	1.285	1.285	4.360	4.359	4.362
4	1.372	1.374	1.373	5.412	5.413	5.416
5	1.445	1.468	1.447	6.520	6.524	6.525
6	1.522	1.550	1.526	7.652	7.660	7.658
7	1.606	1.736	1.612	8.796	9.093	8.803
8	1.697	1.835	1.705	9.946	10.313	9.954

[a]Present study with shear deformation and rotatory inertia ($N=80$).
[b]Jones-Oliviera, Ref. 23, without shear deformation, $\Omega=\Omega_{(\text{Ref. 23})}/\sqrt{13}$ ($N=8$).
[c]Chen and Ginsberg, Ref. 8, without shear deformation, $\Omega=a/\sqrt{a^2-1}\sqrt{\lambda_{(\text{Ref. 8})}}$ ($N=30$).

TABLE V. Comparison of the first six frequencies of the flexural branch for a very long spheroidal shell, $a=1.005$ and $h/l=0.000\,6744$.

Mode number	a	b	c
1	0.002	0.170	0.757
2	1.783	1.798	2.143
3	3.275	3.288	3.600
4	4.725	4.734	4.979
5	6.124	6.134	6.507
6	7.428	7.435	7.894

[a]Present study with shear deformation ($N=80$).
[b]Chen and Ginsberg, Ref. 8, without shear deformation, $\Omega=a/\sqrt{a^2-1}\sqrt{\lambda_{(\text{Ref. 8})}}$ ($N=30$).
[c]Jones-Oliviera, Ref. 22, without shear deformation, $\Omega=\Omega_{(\text{Ref. 22})}/\sqrt{13}$ ($N=8$).

referred to as the cutoff frequency. The extensional branch has a cutoff frequency that is independent of thickness. The two thickness-shear branches have nearly equal cutoff frequencies that are inversely proportional to the thickness. The slope of the branches of the frequency spectrum in the high frequency/high wave-number region represents the group velocity c_g. Since the frequency spectra are plotted in terms of the nondimensional frequency Ω vs mode number n, the group velocity depends on the specific geometry of the shell. The chord length of the shell along its major axis is given by $2l[E(1/a)]$ where $E(1/a)$ is the complete Elliptic function. For high mode numbers n, the structural wavelength is given by $\lambda_n \approx 4l[E(1/a)]/n$ and the wave number k_n is $2\pi/\lambda_n$. The group velocity in the high frequency range $c_g=\partial\omega/\partial k \approx \omega/k$, so that $c_g/c_p \approx 2E(1/a)\Omega/\pi n$, which depends on the geometry of the shell. Thus, the slope in the frequency spectra Ω vs n can be written in terms of the group velocity, i.e., $\Omega/n=[1.57/E(1/a)]c_g/c_p$.

For $\nu=0.3$, one can normalize the group velocity c_g to the shear velocity c_s, i.e., $c_s/c_p=0.59$, where $c_s^2=G/\rho_s$, with G denoting the shear modulus. Thus, the various branches have high-frequency slopes $\Omega/n=[0.93/E(1/a)]c_g/c_s$. The flexural branch's group velocity approaches the Rayleigh wave velocity, which, for Poisson's ratio of $\nu=0.3$ is $c_R=0.923c_s$, as shown by Mindlin.[31] Thus, the flexural branch has a group velocity $c_g=0.923c_s$. The torsional and torsional thickness-shear branches have a group velocity that approaches the shear velocity, so that their group velocity is $c_g=c_s$. The extensional and non-torsional thickness shear group velocity approach the plate velocity $c_g=c_p$.

Frequency spectra were computed for $\nu=0.3$, $a=10$, 3.162, 1.4142, 1.3, 1.2, 1.1, 1.05, and 1.01 with $h/l=0.1$, 0.05, and 0.005. Numerical results were computed over a frequency range $\Omega=0$ to 45.0 with a step size of 0.001. The shell with $a=10$ is nearly spherical ($l/R=1.005$) and the shell with $a=1.01$ has a slender shape with $l/R=7.1$. Of these, numerical results are presented for $a=10$, 1.4142, 1.05, and 1.01.

The frequency spectrum for a nearly spherical shell with $a=10$ and $h/l=0.1$, is shown in Fig. 2. These results are very close to the results of Wilkinson's[29] for a spherical shell with $h/R=0.1$. It can be seen that $c_g \approx c_s$ for the torsional and the torsional thickness shear branches, and $c_g \approx c_p$ for the extensional and the nontorsional thickness shear branches. The flexural branch has a lower $c_g \approx c_R$ as predicted by the formulation above. The cutoff frequencies for the torsional thickness shear branches are nearly identical and that for the extensional branch depends on the thickness and shape parameters h/l and "a," respectively.

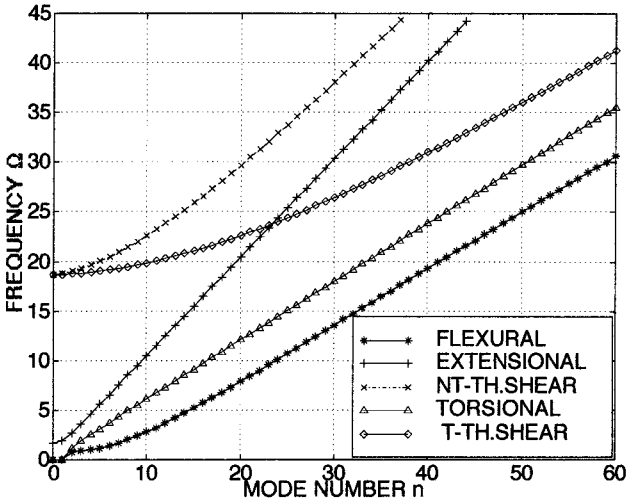


FIG. 2. Frequency spectrum for nearly spherical shell, $a=10$ and $h/l=0.1$.

The frequency spectra for a spheroidal shell with shape parameter $a=1.414$ and $h/l=0.1$ and 0.005 are shown in Figs. 3 and 4, respectively. This shape is commonly used as a basis for numerical comparisons by many authors. Since the shape is nearly spherical ($l/R=2$ vs $l/R=1$ for a spherical shell), one would expect the results to be similar to a spherical shell, see Figs. 2 and 3. One can see that the group velocities match the predicted ones for both figures. Since the thickness shear modes have a cutoff frequency around $\Omega=371$, these are not shown in Fig. 4. The torsional and extensional branches are almost the same for both thicknesses, since they are insensitive to the thickness. However, the flexural branch for the thin shell, $h/l=0.005$ is dominated by the membrane component of the energy, so that it is

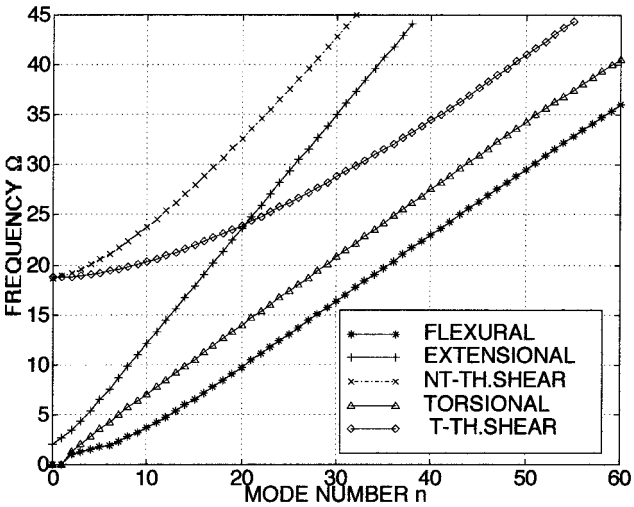


FIG. 3. Frequency spectrum for a spheroidal shell, $a=1.4142$, $h/l=0.1$.

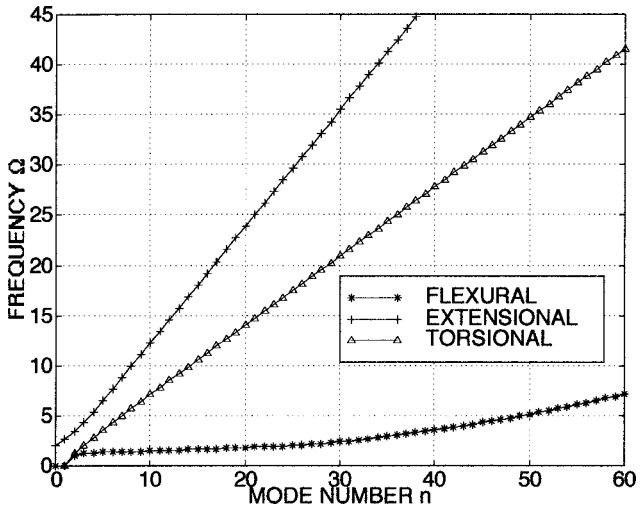


FIG. 4. Frequency spectrum for a spheroidal shell, $a=1.4142$, $h/l=0.005$.

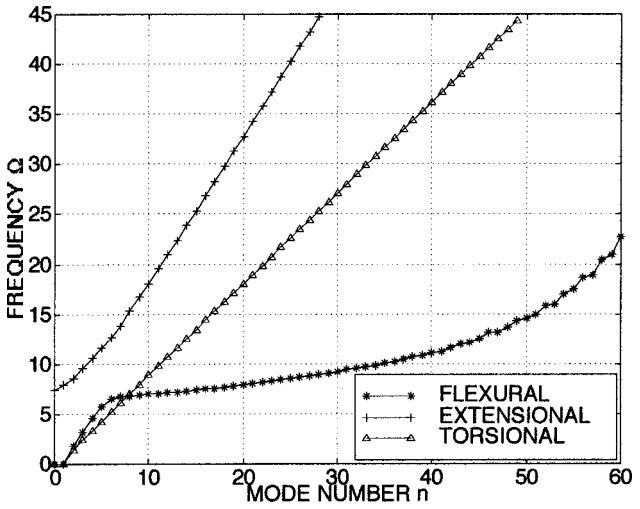


FIG. 6. Frequency spectrum for a long spheroidal shell, $a=1.01$, $h/l=0.005$.

almost flat for $n=2$ to $n=30$, then the bending component takes over, which is sensitive to the thickness. This is seen to exist for $h/l=0.1$ only for $n=2$ to 6.

The frequency spectra for $a=1.01$ and $h/l=0.1$ and 0.005 are shown in Figs. 5 and 6, respectively. For the thicker shell, $h/l=0.1$, the flexural and extensional branches deviate from the spherical ($a=10$) or nearly spherical ($a=1.414$) results. The cutoff frequency of the extensional branch has increased, and the extensional branch is dispersive for modes up to $n=20$. The flexural branch is also dispersive up to $n=40$. Since the ($a=1.01$) shell is an elongated spheroid, it starts to approach the shape of a cylinder. The flexural branch starts to resemble the axisymmetric flexural branch of a cylindrical shell, as shown in Vogelsong.[32] The flexural branch for the thin shell $h/l=0.005$ again exhibits the membrane effect for $n=6$ to 40. Both shells have approximately the same torsional branches and extensional cutoff frequencies. However, the extensional branches for both shells are different for an elongated shell.

To investigate the influence of the shape (or elongation factor) "a" on the flexural branch, results for $a=1.01$, 1.05, 1.414, and 10 are shown in Fig. 7 for $h/l=0.1$ and in Fig. 8 for $h/l=0.005$. It can be seen that the frequencies increase monotonically with decreasing "a," but all of them have a group velocity that approach the predicted values earlier. In addition, for the thin shell with $h/l=0.005$ in Fig. 8, all the shells have predominant membrane effects in the flexural branch.

To investigate the influence of the thickness on the flexural branch, plots are shown for $h/l=0.1$, 0.05 and 0.005 for $a=1.414$ in Fig. 9 and for $a=1.01$ in Fig. 10. In general, the frequencies increase monotonically with the thickness parameter, but with a significant dominance of the membrane component for as many as 30 modes.

The resonance density is defined as the number of modes per unit frequency and is denoted by $n(\omega)=\partial n/\partial\omega$, where n is the mode number. The normalized resonance density, $\partial n/\partial\Omega$, is shown in Fig. 11 for $a=10$, 1.414, and 1.01, and $h/l=0.005$, 0.05, and 0.01. To reduce the clutter of overlapping data, some data above a certain frequency were removed. It can be seen that the resonance density peaks at the extensional cutoff frequency for each shape parameter "a,"

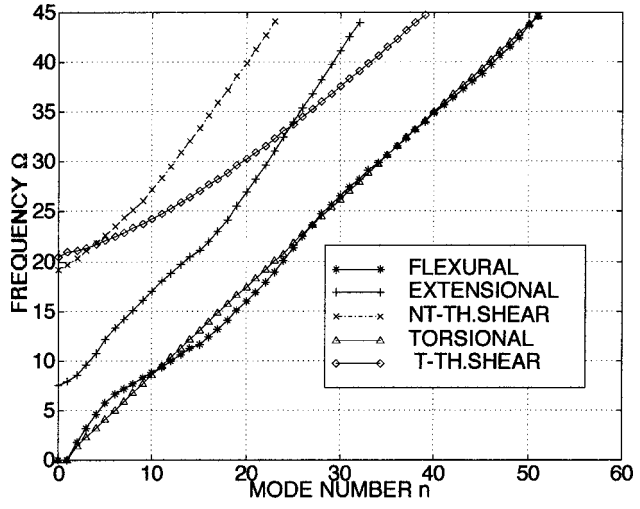


FIG. 5. Frequency spectrum for a long spheroidal shell, $a=1.01$, $h/l=0.1$.

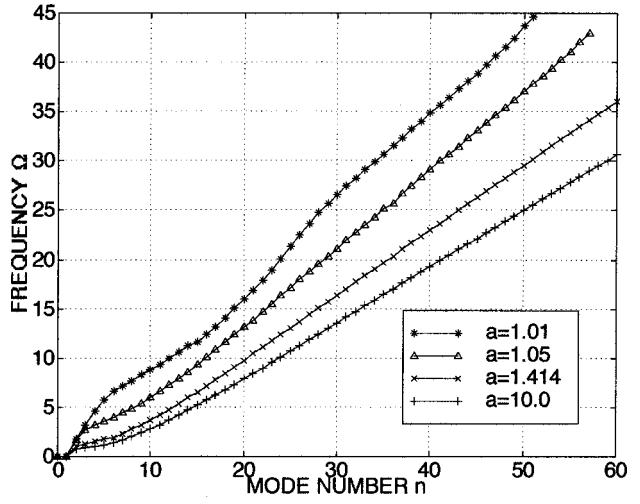


FIG. 7. Flexural branch spectrum versus shape factor a, $h/l=0.1$.

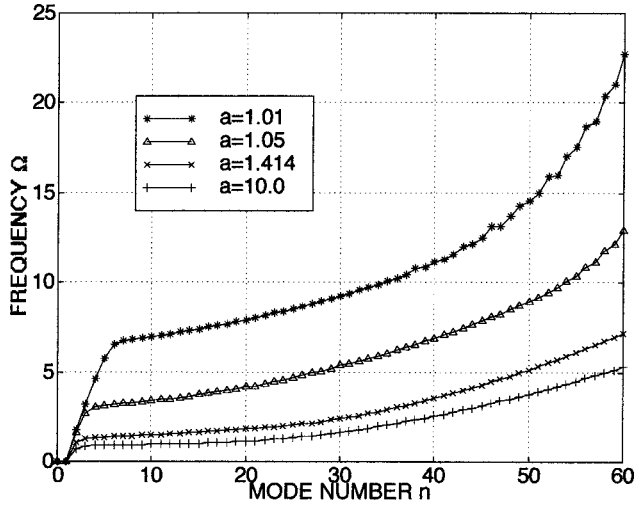


FIG. 8. Flexural branch spectrum versus shape factor a, $h/l=0.005$.

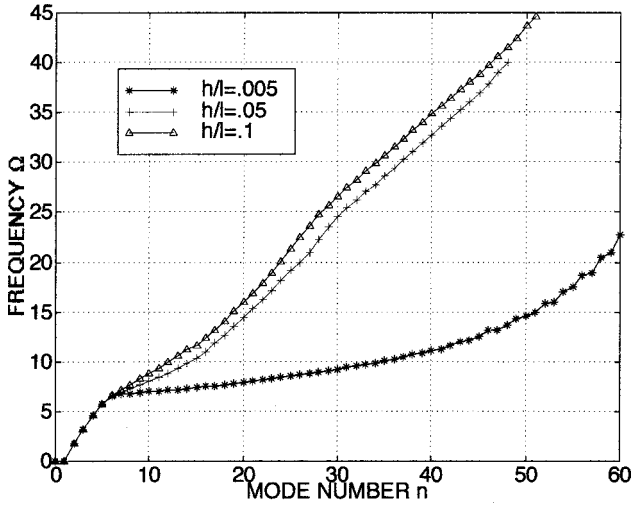


FIG. 10. Flexural branch spectrum versus thickness ratio h/l, $a=1.01$.

with the peak inversely proportional to the thickness. Thus, for $a=10$, the peak occurs at $\Omega=1.62$, for $a=1.414$ it occurs at $\Omega=1.99$, and for $a=1.01$ at $\Omega=7.41$. For higher frequencies, the resonance density becomes a constant and is proportional to $1/c_R$.

C. Dynamic response

The dynamic response of spheroidal shells under axisymmetric surface forces can be obtained from the solution of the nonhomogeneous Eqs. (2.3)–(2.7). For the response due to an axisymmetric uniform normal ring load at $\eta=\eta_0$, let

$$q_u=q_v=m_{\beta\eta}=m_{\beta\phi}=0$$

and

$$q_w=\frac{\delta(\eta-\eta_0)a^2}{2\pi l^2\sqrt{a^2-1}\sqrt{a^2-\eta_0^2}},$$

where δ is the Dirac delta function. When $\eta_0=1.0$, the ring load becomes a unit point force applied at the apex of the spheroidal shell.

Substituting the normal force of q_w in Eq. (2.5), and the comparison functions (3.1), results in $5N\times 5N$ algebraic equations of the type

$$[\mathbf{Z}][\mathbf{Y}]=[\mathbf{f}],$$

where $\mathbf{Z}$ represents the $5N\times 5N$ dynamic matrix of the shell, $\mathbf{Y}$ is the $5N$ displacement amplitude vector and $\mathbf{f}$ denotes the forcing vector, where the components of $\mathbf{f}$ are given by

$$f_n^0=\frac{(1-\nu^2)P_k(\eta_0)}{2\pi Ehl},$$

$$k=0,\quad n=1,\quad k=1,2,3...,\quad n=5k-1$$

$$=0\quad \text{otherwise.}$$

Computations were made for driving ring/point mobility and transfer mobilities, with a constant structural damping introduced by using a complex Young's modulus, $E^*=E(1+i\overline{\eta})$, with the damping coefficients being $\overline{\eta}=0.01$, for $\nu=0.3$. Computations were made for $a=10$ and 1.01, with either a point load at the apex $\eta_0=1$ or uniformly distributed ring load located at η_0. Three points were picked for the

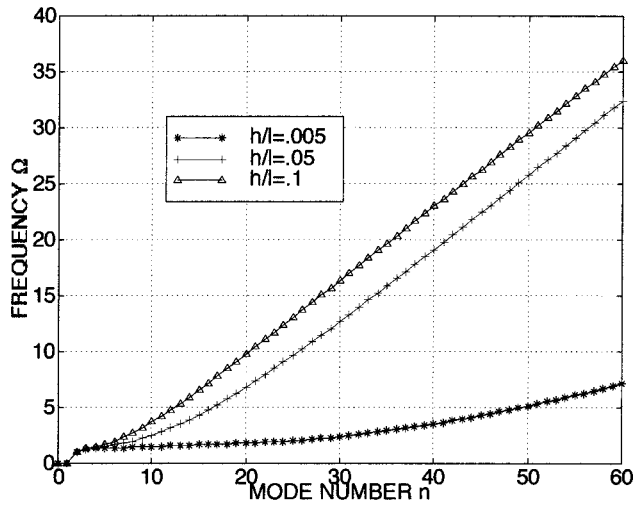


FIG. 9. Flexural branch spectrum versus thickness ratio h/l, $a=1.4142$.

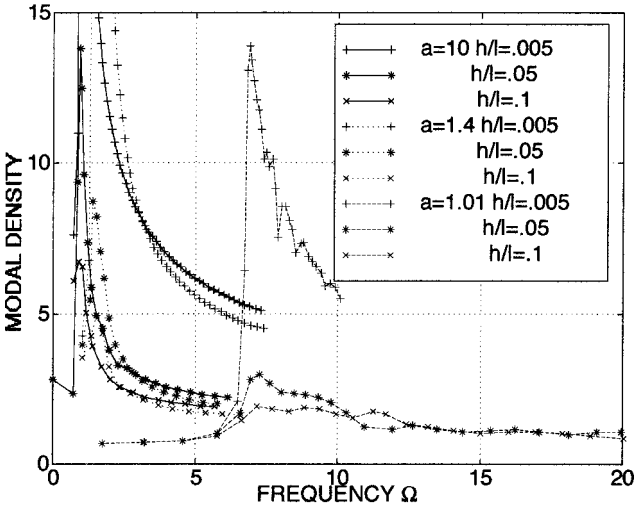


FIG. 11. Resonance density versus frequency for flexural branch.

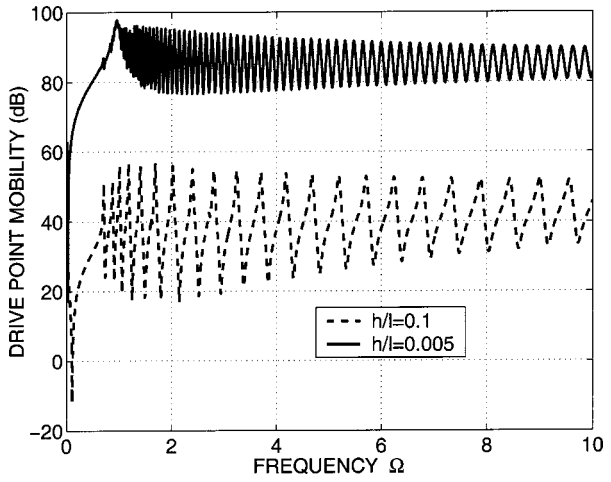


FIG. 12. Driving point mobility versus frequency for a point force at $\eta_0 = 1$ for a damped, nearly spherical shell, $a = 10$.

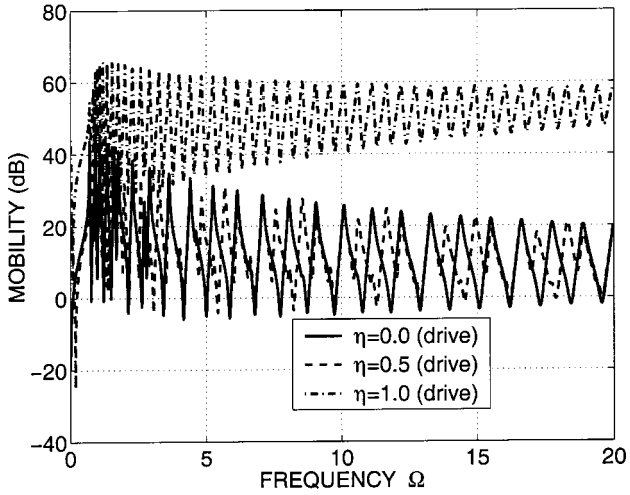


FIG. 14. Driving ring and point mobilities versus frequency for drive points $\eta_0 = 0$, 0.5, and 1.0 for a damped nearly spherical shell, $a = 10$, $h/l = 0.05$.

driving point or transfer mobilities at η or $\eta_0 = 0$, 0.5, and 1. In order to obtain normalized results, the normal load is normalized by $El^2/(1-\nu^2)$ and the normalized shell velocity is given by $\Omega\overline{w}$, so that the mobility is plotted in dB.

The driving point mobility for two shells, with thicknesses $h/l = 0.1$ and 0.005 for a nearly spherical shell, $a = 10.0$, is shown in Fig. 12 and for a long spheroidal shell, $a = 1.01$, in Fig. 13. In both cases, it can be seen that the thin shell ($h/l = 0.005$) has a higher mean-line mobility than the thicker shell ($h/l = 0.1$). This is consistent in that the thinner shell is less elastically stiff. The very high frequency asymptote of the mean line of the driving point mobility should approach that of an infinite elastic plate of the same thickness. However in the mid-to-high frequency range, the mean line at the drive point for a shell is highly influenced by the geometric mean of the radii of curvatures. For a shell to behave like a plate of the same thickness, the ratio of flexural wavelength to the local radius of curvature at the drive point should be $\ll 1$. For example, using a value of 0.01 for the ratio, the frequency must exceed $\Omega > 350$ for the ($a = 10$) shell, and $\Omega > 1700$ for the ($a = 1.01$) shell, well beyond the computed range in this study. An equivalent thickness plate has a normalized (constant) mobility given by $\sqrt{3}/(4(h/l)^2)$, which gives a value of 33 dB for $h/l = 0.1$, and 85 dB for $h/l = 0.005$. To compare with the plate mobility, the mobility in the low frequency range in Figs. 12 and 13 show that the shell mean lines are 77–85 dB for $h/l = 0.005$, and 42–46 dB for $h/l = 0.1$.

To examine the influence of the location of the drive point on the driving point mobility, two drive rings at $\eta_0 = 0$ and 0.5 and a driving point at $\eta_0 = 1.0$ are shown in Fig. 14 for a nearly spherical shell ($a = 10$) and in Fig. 15 for a long spheroidal shell ($a = 1.01$), with $h/l = 0.05$ in both cases. In general, the mean lines of the driving ring mobility for $\eta_0 = 0$ and 0.5 are close, as the radii of curvatures at these points are approximately the same. The driving point mobility at $\eta_0 = 1.0$ is higher than the two ring mobilities, partly due to the fact that the ring load is nearly a self-equilibrating surface load while the point load is not, so that the net total mechanical force of a point force on the shell is much higher than an equivalent uniformly distributed ring

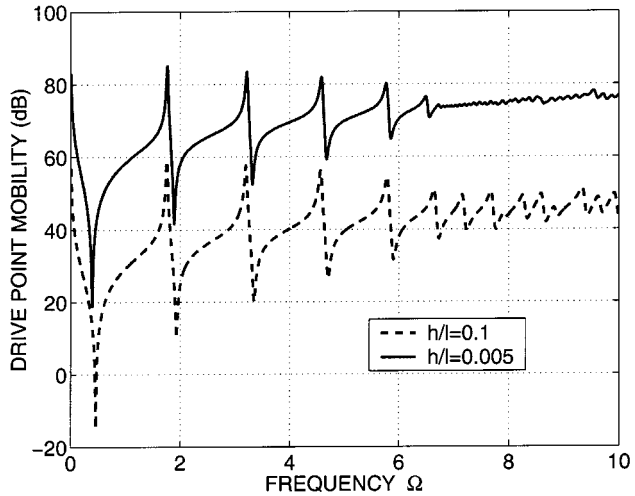


FIG. 13. Driving point mobility versus frequency for a point force at $\eta_0 = 1$ for a damped long spheroidal shell, $a = 1.01$.

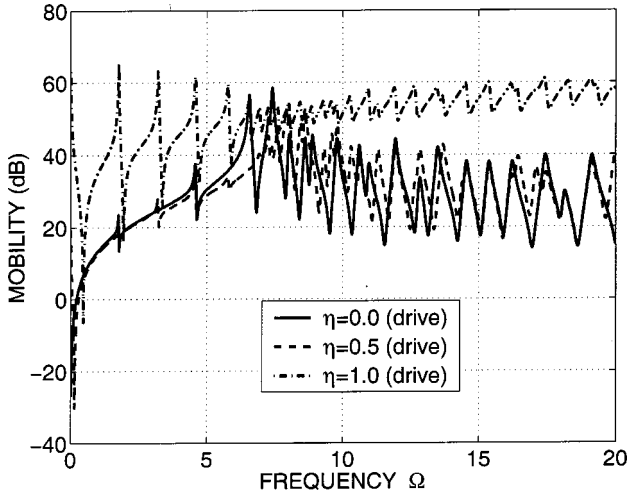


FIG. 15. Driving ring and point mobilities versus frequency for drive points $\eta_0 = 0$, 0.5, and 1.0 for a damped long spheroidal shell, $a = 1.01$, $h/l = 0.05$.

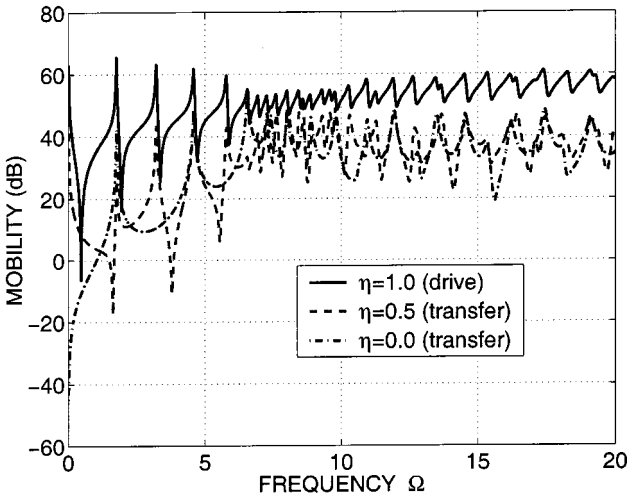


FIG. 16. Drive and transfer mobilities versus frequency for a drive point at $\eta_0=1$ for a damped long spheroidal shell, $a=1.01$, $h/l=0.05$.

load. It is noted that the mean line of the mobility for the long spheroidal shell ($a=1.01$) is lower than the nearly spherical shell ($a=10.0$) in the frequency range below $\Omega=10$. However, plotting the mobility for both shell parameters to a range of $\Omega=40$ (not shown in this paper) shows that the two mean-lines approach the same limit, the limit being influenced by the thickness of the shell ($h/l=0.05$) in the high frequency range.

To investigate the influence of the drive point on the transfer mobility, the transfer mobility at $\eta=0$ and 0.5 as well as the driving point mobility are shown in Fig. 16 for a point load at $\eta_0=1.0$, for a long spheroidal shell with $a=1.01$ and $h/l=0.05$. In the higher frequency range of Fig. 16, the mean-line response at $\eta=0$ (equator) is approximately 5 dB below the corresponding response at $\eta=0.5$ (midpoint between the equator and the apex); and they are about 20–25 dB below the mean line of the driving point mobility.

V. SUMMARY AND CONCLUSIONS

The self-adjoint PDE for the vibration of a prolate spheroidal shell with extensional, bending, thickness shear and rotatory inertias were derived. These were used to compute the frequency spectra for various shell geometries in terms of the shape factor "a" and thickness-to-half-length ratios. The frequency spectrum consists of five branches: flexural, torsional, extensional, torsional thickness shear, and nontorsional thickness shear. Each branch of the frequency spectrum was shown to have the proper high frequency asymptote, i.e., the group velocity. The resonance density of the flexural branches was computed for all the shells being discussed. The effect of shell thickness h/l and shape parameter "a" on the flexural branch was presented in the form of comparative plots.

The dynamic response of the spheroidal shells due to ring and point forces were presented for various shell thicknesses h/l and shape parameters "a." The driving and transfer ring/point mobilities were computed and presented for a structural damping of 0.01.

In general, the frequencies of the flexural branch are lower than those of the torsional branch, both being acoustic branches. The flexural branch has a group velocity close to the Rayleigh velocity, while the torsional branch has one close to the shear velocity. The extensional branch is an optical branch with a cutoff frequency almost independent of the thickness of the shell, but highly dependent on the shape factor "a." It has a group velocity of that of the plate velocity. The nontorsional and torsional thickness shear branches are optical branches with almost equal cutoff frequencies. The former has a group velocity of plate and the latter that of the shear velocity. The cutoff frequency is inversely proportional to the thickness of the shell.

In general, the flexural branch's frequencies increase monotonically with thickness and with the shape parameter "a" denoting longer spheroidal shells. The resonance density of the flexural branches show that for every shape parameter "a," the resonance density has a peak at the cutoff frequency of that shell geometry, and is inversely proportional to the thickness of the shell.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research Code 321 and the ONR/ASEE Summer Faculty Program.

APPENDIX: PDE OPERATORS

$$L_{uu}\bar{u}=DB\frac{\partial^2\bar{u}}{\partial\eta^2}-\eta D(1+D^2)\frac{\partial\bar{u}}{\partial\eta}+\frac{E}{DB}\frac{\partial^2\bar{u}}{\partial\phi^2}+\left[-\eta^2\frac{D}{B}\nu a^2\frac{D}{C}-F\kappa a^2\frac{D^3}{C}\right]\bar{u}+\epsilon\left\{-a^4\frac{BD}{C^3}\left[B\frac{\partial^2\bar{u}}{\partial\eta^2}+\eta(3-7D^2)\frac{\partial\bar{u}}{\partial\eta}\right]+F\frac{a^4D^3}{A^3}\frac{\partial^2\bar{u}}{\partial\phi^2}-\left[a^4\eta^2\frac{D}{AC^2}-F\gamma\kappa a^6\frac{D^3B}{C^4}\right]\bar{u}\right\},\tag{A1}$$

$$L_{uv}\bar{v}=(1+F)\frac{\eta}{B}\frac{\partial\bar{v}}{\partial\phi}+(\nu+F)\frac{\partial^2\bar{v}}{\partial\eta\partial\phi}+\epsilon\left\{\frac{a^4\eta}{AC^2}(1+F)\frac{\partial\bar{v}}{\partial\phi}\right\},\tag{A2}$$

$$L_{uw}\bar{w}=\frac{aD\sqrt{B}}{\sqrt{C}}[(1+F\kappa)D^2+\nu]\frac{\partial\bar{w}}{\partial\eta}+a\frac{\eta\sqrt{B}}{C^{5/2}}\times(4A+B)\bar{w}+\epsilon\left\{a^5A\frac{\sqrt{B}}{C^{5/2}}\left[-\frac{B}{C^2}(1+F\gamma\kappa)\times\frac{\partial\bar{w}}{\partial\eta}+\eta\left[\frac{1}{A^2}-\frac{6}{C^2}+9\frac{A}{C^3}\right]\bar{w}\right]\right\},\tag{A3}$$

$$L_{u\beta_\eta}\beta_\eta = F\kappa D^2\beta_\eta + \epsilon\left\{ a^2\frac{B^2}{C^2}\frac{\partial^2\beta_\eta}{\partial\eta^2} - 4a^2\eta\frac{AB}{C^3}\frac{\partial\beta_\eta}{\partial\eta} - F\frac{a^2}{AC}\frac{\partial^2\beta_\eta}{\partial\phi^2} + \left[\frac{a^2\eta^2}{C^2} - F\gamma\kappa\frac{a^4AB}{C^4}\right]\beta_\eta\right\}, \tag{A4}$$

$$L_{u\beta_\phi}\beta_\phi = \epsilon\left\{ -(1+F)a^2\eta\frac{D}{AC}\frac{\partial\beta_\phi}{\partial\phi}\right\}, \tag{A5}$$

$$L_{uu}\overline{u} = (\nu+F)\frac{\partial^2\overline{u}}{\partial\eta\partial\phi} - \frac{\eta}{B}(1+F)\frac{\partial\overline{u}}{\partial\phi} + \epsilon\left\{ -(1+F)\frac{a^4\eta}{AC^2}\frac{\partial\overline{u}}{\partial\phi}\right\}, \tag{A6}$$

$$L_{uv}\overline{v} = FDB\frac{\partial^2\overline{v}}{\partial\eta^2} - FD(1+D^2)\eta\frac{\partial\overline{v}}{\partial\eta} + \frac{1}{DB}\frac{\partial^2\overline{v}}{\partial\phi^2} + F\left[a^2\frac{D}{C} - \eta^2\frac{D}{B} - \kappa a^2\frac{D}{A}\right]\overline{v} + \epsilon\left\{ -Fa^4\frac{BD}{C^3}\left[B\frac{\partial^2\overline{v}}{\partial\eta^2} + \eta(3-7D^2)\frac{\partial\overline{v}}{\partial\eta}\right] + a^4\frac{D^3}{A^3}\frac{\partial^2\overline{v}}{\partial\phi^2} - \frac{Fa^4D}{C^2A}\left[\eta^2 + \gamma\kappa\frac{a^2B}{A}\right]\overline{v}\right\}, \tag{A7}$$

$$L_{uw}\overline{w} = \frac{a}{\sqrt{AB}}[(1+F\kappa)+\nu D^2]\frac{\partial\overline{w}}{\partial\phi} + \epsilon\left\{ (1+F\gamma\kappa)\frac{a^5\sqrt{B}}{A^{3/2}C^2}\frac{\partial\overline{w}}{\partial\phi}\right\}, \tag{A8}$$

$$L_{v\beta_\eta}\beta_\eta = \epsilon\left\{ (1+F)\eta a^2\frac{D^3}{A^2}\frac{\partial\beta_\eta}{\partial\phi}\right\}, \tag{A9}$$

$$L_{v\beta_\phi}\beta_\phi = F\kappa\beta_\phi + \epsilon\left\{ -4F\frac{a^2AB}{C^3}\eta\frac{\partial\beta_\phi}{\partial\eta} - \frac{a^2}{AC}\frac{\partial^2\beta_\phi}{\partial\phi^2} + Fa^2\frac{B^2}{C^2}\frac{\partial^2\beta_\phi}{\partial\eta^2} + F\left(\frac{\eta^2a^2}{C^2} + \gamma\kappa\frac{a^4B}{AC^2}\right)\beta_\phi\right\}, \tag{A10}$$

$$L_{wu}\overline{u} = -a\frac{\sqrt{B}}{C^{3/2}}[A(1+F\kappa)+\nu C]\frac{\partial\overline{u}}{\partial\eta} + \frac{a\eta}{\sqrt{BC}}\left[1+(\nu+F\kappa)D^2 - 3F\kappa\frac{AB}{C^2}\right]\overline{u} + \epsilon\left\{ a^5\frac{AB^{3/2}}{C^{9/2}}(1+F\kappa\gamma)\frac{\partial\overline{u}}{\partial\eta} + \eta\frac{a^5\sqrt{B}}{AC^{5/2}}[1+3F\kappa\gamma D^4(2-3D^2)]\overline{u}\right\}, \tag{A11}$$

$$L_{wv}\overline{v} = -\frac{a}{\sqrt{AB}}(1+F\kappa+\nu D^2)\frac{\partial\overline{v}}{\partial\phi} + \epsilon\left\{ -(1+F\kappa\gamma)\frac{a^5\sqrt{B}}{A^{3/2}C^2}\frac{\partial\overline{v}}{\partial\phi}\right\}, \tag{A12}$$

$$L_{ww}\overline{w} = F\kappa D\left[B\frac{\partial^2\overline{w}}{\partial\eta^2} - (1+D^2)\eta\frac{\partial\overline{w}}{\partial\eta} + \frac{1}{D^2B}\frac{\partial^2\overline{w}}{\partial\phi^2}\right] - \frac{a^2D}{A}[1+D^4+2\nu D^2]\overline{w} + \epsilon\left\{ -F\kappa\gamma\frac{a^4DB}{C^3}\left[B\frac{\partial^2\overline{w}}{\partial\eta^2} + (3-7D^2)\eta\frac{\partial\overline{w}}{\partial\eta}\right] + F\kappa\gamma\frac{a^4D^3}{A^3}\frac{\partial^2\overline{w}}{\partial\phi^2} + a^6\frac{D^3B}{C}\left[\frac{1}{C^3} - \frac{1}{A^3}\right]\overline{w}\right\}, \tag{A13}$$

$$L_{w\beta_\eta}\beta_\eta = F\kappa\frac{\sqrt{A}}{a\sqrt{B}}\left[B\frac{\partial\beta_\eta}{\partial\eta} - \eta\beta_\eta\right] + \epsilon\left\{ -a^3\frac{\sqrt{BA}}{C^3}\left[B(F\kappa\gamma+1)\frac{\partial\beta_\eta}{\partial\eta}\right] - \frac{a^3\sqrt{B}}{C^2\sqrt{A}}\eta[1+3F\kappa\gamma D^2(1-2D^2)\beta_\eta]\right\}, \tag{A14}$$

$$L_{w\beta_\phi}\beta_\phi = F\kappa\frac{\sqrt{C}}{a\sqrt{B}}\frac{\partial\beta_\phi}{\partial\phi} + \epsilon\left\{ [1+F\kappa\gamma]\frac{a^3\sqrt{B}}{AC^{3/2}}\frac{\partial\beta_\phi}{\partial\phi}\right\}, \tag{A15}$$

$$L_{\beta_\eta u}\overline{u} = F\kappa D^2\overline{u} + \epsilon\left\{ a^2\frac{B^2}{C^2}\frac{\partial^2\overline{u}}{\partial\eta^2} - 4\frac{a^2AB}{C^3}\eta\frac{\partial\overline{u}}{\partial\eta} - F\frac{a^2}{AC}\frac{\partial^2\overline{u}}{\partial\phi^2} + \left[\frac{a^2\eta^2}{C^2} - F\kappa\gamma\frac{a^4AB}{C^4}\right]\overline{u}\right\}, \tag{A16}$$

$$L_{\beta_\eta v}\overline{v} = \epsilon\left\{ -(1+F)\frac{a^2\eta}{C^{3/2}\sqrt{A}}\frac{\partial\overline{v}}{\partial\phi}\right\}, \tag{A17}$$

$$L_{\beta_\eta w}\overline{w} = -F\kappa\frac{\sqrt{AB}}{a}\frac{\partial\overline{w}}{\partial\eta} + \epsilon\left\{ \frac{a^3\sqrt{B}}{C^3\sqrt{A}}\left[(1+F\kappa\gamma)AB\frac{\partial\overline{w}}{\partial\eta} + \eta C(-1+3D^2(1-2D^2))\overline{w}\right]\right\}, \tag{A18}$$

$$L_{\beta_\eta\beta_\eta\beta_\eta}\beta_\eta = -F\kappa\frac{DC}{a^2}\beta_\eta + \epsilon\left\{D\left[B\frac{\partial^2\beta_\eta}{\partial\eta^2} - \eta(1+D^2)\frac{\partial\beta_\eta}{\partial\eta}\right] + \frac{F}{DB}\frac{\partial^2\beta_\eta}{\partial\phi^2} + D\left(-\frac{\eta^2}{B} - \nu\frac{a^2}{C} + F\kappa\gamma\frac{a^2B}{C^2}\right)\beta_\eta\right\}, \quad \text{(A19)}$$

$$L_{\beta_\eta\beta_\phi\beta_\phi}\beta_\phi = \epsilon\left\{(\nu+F)\frac{\partial^2\beta_\phi}{\partial\eta\partial\phi} + (1+F)\frac{\eta}{B}\frac{\partial\beta_\phi}{\partial\phi}\right\}, \quad \text{(A20)}$$

$$L_{\beta_\phi u}\bar{u} = \epsilon\left\{(1+F)\frac{a^2\eta D}{AC}\frac{\partial\bar{u}}{\partial\phi}\right\}, \quad \text{(A21)}$$

$$L_{\beta_\phi v}\bar{v} = F\kappa\bar{v} + \epsilon\left\{\frac{Fa^2B}{C^2}\left[B\frac{\partial^2\bar{v}}{\partial\eta^2} - 4\frac{A}{C}\eta\frac{\partial\bar{v}}{\partial\eta}\right] - \frac{a^2}{AC}\frac{\partial^2\bar{v}}{\partial\phi^2} + \frac{F}{C^2}\left[a^2\eta^2 + \kappa\gamma\frac{a^4B}{A}\right]\bar{v}\right\}, \quad \text{(A22)}$$

$$L_{\beta_\phi w}\bar{w} = -F\kappa\frac{\sqrt{C}}{a\sqrt{B}}\frac{\partial\bar{w}}{\partial\phi} + \epsilon\left\{-(1+F\kappa\gamma)\frac{a^3\sqrt{B}}{AC^{3/2}}\frac{\partial\bar{w}}{\partial\phi}\right\}, \quad \text{(A23)}$$

$$L_{\beta_\phi\beta_\eta}\beta_\eta = \epsilon\left\{(\nu+F)\frac{\partial^2\beta_\eta}{\partial\eta\partial\phi} - (1+F)\frac{\eta}{B}\frac{\partial\beta_\eta}{\partial\phi}\right\}, \quad \text{(A24)}$$

$$L_{\beta_\phi\beta_\phi\beta_\phi}\beta_\phi = -F\kappa\frac{\sqrt{AC}}{a^2}\beta_\phi + \epsilon\left\{FDB\frac{\partial^2\beta_\phi}{\partial\eta^2} - F\eta D(1+D^2)\frac{\partial\beta_\phi}{\partial\eta} + \frac{1}{DB}\frac{\partial^2\beta_\phi}{\partial\phi^2} + Fa^2D\left[\frac{1}{C} - \frac{\eta^2}{a^2B} - \kappa\gamma\frac{B}{AC}\right]\beta_\phi\right\}. \quad \text{(A25)}$$

[1] F. L. DiMaggio and A. Silbiger, "Free extensional torsional vibrations of a prolate spheroidal shell," J. Acoust. Soc. Am. **33**, 56–58 (1961).
[2] A. Silbiger and F. L. DiMaggio, "Extensional axisymmetric second class vibrations of a prolate spheroidal shell," Columbia University Contract No. Nonr 266 (67) Tech. Rep. No. 3, 1961.
[3] N. Shiraishi and F. L. DiMaggio, "Perturbations solution for the axisymmetric vibrations of prolate spheroidal shells," J. Acoust. Soc. Am. **34**, 1725–1731 (1962).
[4] P. J. Nemergut and R. S. Brand, "Axisymmetric vibrations of prolate spheroidal shells," J. Acoust. Soc. Am. **38**, 262–265 (1965).
[5] F. L. DiMaggio and R. Rand, "Axisymmetric vibrations of prolate spheroidal shells," J. Acoust. Soc. Am. **40**, 179–186 (1966).
[6] R. H. Rand, "Torsional vibrations of elastic prolate spheroids," J. Acoust. Soc. Am. **44**, 749–751 (1968).
[7] C. B. Burroughs and E. B. Magrab, "Natural frequencies of prolate spheroidal shells of constant thickness," J. Sound Vib. **57**, 571–581 (1978).
[8] P. T. Chen and J. H. Ginsberg, "Modal properties and eigenvalue veering phenomena in the axisymmetric vibration of spheroidal shells," J. Acoust. Soc. Am. **92**, 1499–1508 (1992).
[9] S. I. Hayek and F. L. DiMaggio, "Axisymmetric vibrations of submerged spheroidal shells," Ph.D. thesis, Columbia University, New York, 1964 and Columbia University Tech. Report No. 4 CU-1-64-ONR(67)-CE, 1964.
[10] S. I. Hayek and F. L. DiMaggio, "Complex natural frequencies of vibrating submerged spheroidal shells," Int. J. Solids Struct. **6**, 333–351 (1970).
[11] R. Rand and F. L. DiMaggio, "Vibrations of fluid-filled spherical and spheroidal shells," J. Acoust. Soc. Am. **42**, 1278–1286 (1967).
[12] T. Yen and F. L. DiMaggio, "Forced vibrations of submerged spheroidal shells," J. Acoust. Soc. Am. **41**, 618–626 (1967).
[13] B. Bedrosian and F. L. DiMaggio, "Transient response of submerged spheroidal shells," Int. J. Solids Struct. **8**, 111–129 (1972).
[14] B. S. Berger, "The dynamic response of a prolate spheroidal shell submerged in an acoustic medium," J. Appl. Mech. **41**, 925–929 (1974).
[15] T. Lee and F. L. DiMaggio, "Effect of bending on the axisymmetric vibrations of a spheroidal model of the head," J. Biomech. **9**, 803–812 (1976).
[16] C. T. F. Ross and T. Johns, "Vibrations of submerged hemi-ellipsoidal domes," J. Sound Vib. **91**, 363–373 (1983).
[17] V. Y. Prikhod'ko, "Forced vibrations of submerged spheroidal shells," Sov. Phys. Acoust. **34**, 80–83 (1988).
[18] J. P. Pauwelussen, "Validation of the underwater shock analysis program USA, Part II: Numerical computations," TNO-IWECO, Inst. Mechanical Construction, Delft, The Netherlands, 1987.
[19] P. T. Chen and J. H. Ginsberg, "Variational formulation of acoustic radiation from submerged spheroidal shells," J. Acoust. Soc. Am. **94**, 221–233 (1993).
[20] G. Chertock, A. E. Hirsh, and T. F. Ogilvie, "The flexural response of a submerged or floating body to an underwater explosion. Part 1—Theory and Part II—Experiments with floating box models," Department of the Navy, The David W. Taylor Model Basin, Reports Nos. C-517 and C-519, 1952 and 1953, respectively.
[21] A. Silbiger, "Scattering of sound by an elastic prolate spheroid," J. Acoust. Soc. Am. **35**, 564–570 (1963).
[22] J. B. Jones-Oliveira, "Fluid–solid interaction of a prolate spheroidal shell structure loaded by an acoustic shock wave," Ph.D. dissertation in Mathematical Physics, Massachusetts Institute of Technology, Cambridge, MA, 1990.
[23] J. B. Jones-Oliveira, "Transient analytic and numerical results for the fluid–solid interaction of prolate spheroidal shells," J. Acoust. Soc. Am. **99**, 392–407 (1996).
[24] C. Flammer, *Spheroidal Wave Functions* (Stanford University Press, Stanford, CA, 1957).
[25] H. Kraus, *Thin Elastic Shells* (Wiley, New York, 1967).
[26] E. Ventsel and T. Krauthammer, *Thin Plates and Shells* (Marcel Dekker, New York, 2000).
[27] S. I. Hayek and J. E. Boisvert, "Equations of motion for nonaxisymmetric vibrations of prolate spheroidal shells," NUWC-NPT Report No. 11,214, 18 February 2000.
[28] For the sake of brevity, the expressions for the coefficients are not included. They may be obtained upon request from the authors.
[29] J. P. Wilkinson, "Natural frequencies of closed spherical shells," J. Acoust. Soc. Am. **38**, 367–368 (1965).
[30] S. I. Hayek, "Vibration of a spherical shell in an acoustic medium," J. Acoust. Soc. Am. **40**, 342–348 (1966).
[31] R. D. Mindlin, "Influence of rotatory inertia and shear on flexural motions of isotropic, elastic plates," J. Appl. Mech. **18**, 31–38 (1951).
[32] R. Vogelsong, "Nearfield acoustic radiation from a point-excited cylindrical shell," Ph.D. thesis, Penn State University, University Park, PA, 1987.

Structural acoustic silencers—Design and experiment[a)]

Sripriya Ramamoorthy and Karl Grosh[b)]
Department of Mechanical Engineering, University of Michigan, 2200 EECS Building, Ann Arbor, Michigan 48105

Tony G. Nawar
Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan 48105

(Received 14 November 2002; revised 11 July 2003; accepted 11 August 2003)

The effectiveness of introducing flexible structural layers into air conveying ducts for controlling noise is investigated through theoretical and experimental means, focusing at low frequencies where conventional passive silencing technology is least effective. Previous theoretical work has shown that using flexible rather than rigid walls has the potential to achieve high transmission losses. The physical mechanisms responsible for structural acoustic silencing, including the relation between transmission loss peaks and structural resonance corresponding to different transverse structural modes, are presented. Sensitivity of the performance to acoustic and structural boundary conditions is discussed. To eliminate radiated noise from these walls (breakout noise), a rigid walled cavity is introduced under the flexible plate. The challenge is to find means to reject plane waves in the two-duct system. Designs that overcome these issues and achieve appreciable transmission loss are investigated. Results based on three-dimensional finite element simulations are compared with experimental results. [DOI: 10.1121/1.1616926]

PACS numbers: 43.50.Gf, 43.40.Fz, 43.20.Mv [DKW] Pages: 2812–2824

I. INTRODUCTION

Noise control can be achieved by passive or active means. In this paper, experimental and theoretical methods will be used to investigate the use of a passive device, a structural acoustic silencer, to obtain low-frequency transmission loss. Passive noise control includes reactive devices (e.g., quarter-wavelength resonators, Helmholtz resonators, side branches) and dissipative devices (e.g., porous materials like foam) (Munjal, 1987; Pierce, 1989). Current passive approaches do not achieve broadband transmission loss at low frequencies. Active control provides another effective means to obtain low-frequency noise control: active sources are introduced and controlled by noise control algorithms (see, e.g., Nelson and Elliott, 1992). However, for the same noise control performance, it is preferable to use a passive noise control device as it provides a lower cost alternative that is, in most cases, easier to install and maintain.

A structural acoustic silencer consists of a flexible layer inserted into nominally rigid piping. Transmission loss is achieved by two mechanisms—reflection of energy due to an impedance mismatch, and coupling of the incoming acoustic fluctuations to structural vibrations thereby allowing for the extraction of energy through losses in the structure. Applications for structural acoustic silencers include systems like automotive exhaust systems, HVAC systems, and industrial gas lines. The advantages of structural acoustic silencers over porous materials like foam include reduction in static pressure loss, reduced cost of maintenance, and a decrease in the possibility of introducing particulate matter into the flow (i.e., no fibrous material in contact with circulating air).

Huang (1999; Huang *et al.*, 2000) performed theoretical and experimental study of structural acoustic noise control devices. Just as in the current study, a noise control device is placed in the path of flow. A two-dimensional theoretical exploration of the design of a flexible membrane showed the promise of this approach in air (Huang, 1999), and experimental results (Huang *et al.*, 2000) supported some of the predictions related to phase speed of wave in the plate.

The same idea was also independently proposed by Grosh and Dodson (1999) as an analogy to the functioning of the cochlea, and an extension to the idea followed (Ramamoorthy and Grosh, 2000)—a variable impedance plate as a means to reduce reflection while still maintaining transmission loss characteristics. An application of structural acoustic silencers to hydraulic systems using three-dimensional finite element analysis was given by Ramamoorthy *et al.* (2002). Increasing the plate orthotropy (relatively stiffening the plate in the transverse direction) was shown to be beneficial in improving the transmission loss. Tailoring the plate thickness and/or width has the potential to reduce reflections with the added advantage of higher tolerable static pressure load, along with little or no compromise on transmission loss. Typical transmission loss for hydraulic systems was predicted to be around 40 dB.

The effect of external fluid was not considered in any of the previous papers on single-duct structural acoustic silencers. The issue of breakout noise is significant in a single-duct structural acoustic silencer (SAS) model where the flexible plate is open to external air. Breakout noise can be avoided by backing the plate with a cavity (Ramamoorthy *et al.*, 2001; Huang, 2002; Choy and Huang, 2002). However, the

[a)]Part of this material was presented at the 142nd meeting of the Acoustical Society of America, December 2001.
[b)]Electronic mail: grosh@engin.umich.edu

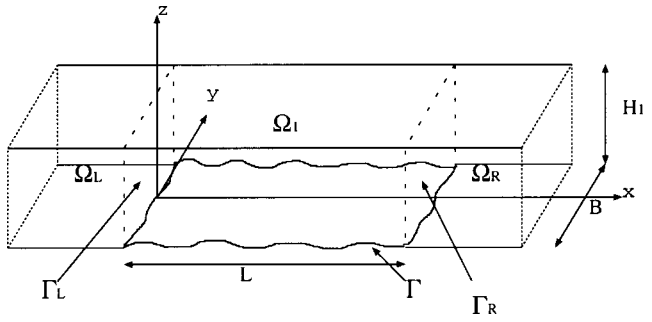


FIG. 1. Single-duct SAS showing the computational domains and boundaries. Ω_1 is the three-dimensional computational domain, Γ is the fluid–structure interface, Γ_L is the left DtN boundary, Γ_R is the right DtN boundary, Ω_L is the infinite region to the left of Γ_L, and Ω_R is the infinite region to the right of Γ_R.

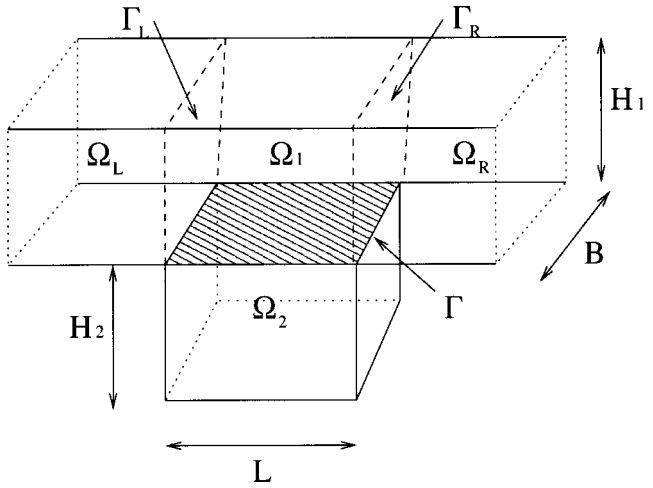


FIG. 2. Two-duct SAS showing the computational domains and boundaries. $\Omega=\Omega_1+\Omega_2$ is the three-dimensional computational domain, Ω_2 is the region of fluid in the backing cavity, Γ is the fluid–structure interface, Γ_L is the left DtN boundary, Γ_R is the right DtN boundary, Ω_L is the infinite region to the left of Γ_L, and Ω_R is the infinite region to the right of Γ_R. The coordinate system used for two-duct SAS is same as for the single-duct SAS shown in Fig. 1.

filtering characteristics of the device change as a result of backing cavity impedance.

The single-duct SAS analysis in this paper brings out an interesting relation between the TL and plate dispersion. To the authors' knowledge, this has not been presented directly in any previous publication. Cummings (2001) brings out the dispersion of coupled structural-acoustic modes in detail, but there is a gap in knowledge to relate TL to these coupled modes. The current paper fills in this gap and brings out the relation of TL to the coupled structural-acoustic modes for the first time.

II. MODELING

In this paper, two designs for structural acoustic silencers are investigated: the single-duct structural acoustic silencer, and the two-duct structural acoustic silencer. The geometry and assumptions involved in these two models are given below.

A. Model geometry

The main difference between the two models is that in single-duct SAS, the plate is loaded by pressure only on one side because external fluid loading is not considered, whereas in a two-duct SAS, a rectangular cavity backs the plate on the other side, thus avoiding breakout noise. Therefore, in two-duct SAS, the plate is loaded by pressure on both sides. The geometry of the two models is discussed in this subsection.

1. Geometry for single-duct SAS

The model (Fig. 1) consists of an infinitely long waveguide of rectangular cross-section, $B\times H_1$, with all walls rigid except a portion of one. A part of the wall at $z=0$ consists of a flexible plate or membrane of length L and width B. The fluid domain is the entire interior of the duct, which is broken into three domains: Ω_L to the left of the noise control (or computational) domain Ω_1, and Ω_R to the right. The domain of finite element analysis is Ω_1 with three of its six boundaries being the fluid structure interface (Γ), and the left and right Dirichlet to Neumann (DtN) boundaries (Γ_L, Γ_R respectively). A DtN boundary condition is an impedance termination condition that relates the fluid pressure to its normal derivative using simplifications resulting from analytical expressions for wave propagation in simple infinite domains [thus reducing the FEA mesh to finite domain; see Keller and Givoli (1989)], in this case, for the semi-infinite rigid-walled ducts Ω_L and Ω_R. Here, the DtN condition is simplified by the fact that only plane waves propagate in the rigid-walled portions of the duct for the geometry and frequencies of interest.

Note that this model has been studied previously [for recent work, see, e.g., Ramamoorthy *et al.* (2001, 2002)] [two-dimensional version in Huang (1999)].

2. Geometry for two-duct SAS

The geometry for two-duct SAS consists of a cavity with five rigid walls backing the flexible plate region of the single-duct SAS (see Fig. 2). The length and width of the cavity are the same as that of the flexible plate, and the cavity height is H_2. There are two fluid–structure interfaces—one each on the top and bottom surfaces of the plate. This design eliminates the issue of breakout noise, thereby making it a stand alone device.

B. Governing equations and assumptions

The region of numerical simulation is the three-dimensional domain $\Omega=\Omega_1$ shown in Fig. 1 for single-duct SAS and $\Omega=\Omega_1+\Omega_2$ in Fig. 2 for two-duct SAS. The fluid is assumed to be compressible and inviscid; fluid viscosity is accounted for heuristically in the plate damping factor. The flexible structure is assumed to be linearly elastic. The input excitation is an incoming right-traveling acoustic plane wave impinging on Γ_L. For the frequency range (50–2000 Hz) and geometries used in this study, only plane waves propagate in the regions $x<0$ and $x>L$. Time harmonic vibrations of the coupled system are analyzed. A time dependence of $e^{-i\omega t}$ is assumed.

Given these approximations, the coupled fluid–structure governing equations are given by

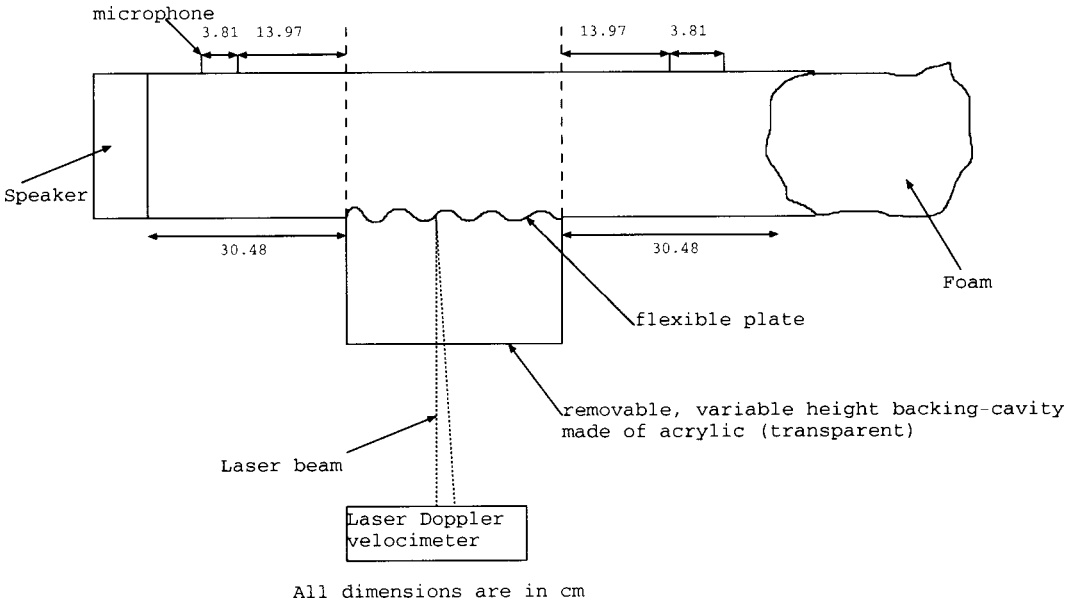


FIG. 3. Experimental setup; dimensions in the figure are not to scale; duct cross section is 2.54×2.54 cm^2.

$$\mathcal{L}_s\mathbf{u}=\hat{\mathbf{n}}(p_1-p_2)|_\Gamma \text{ on } \Gamma \quad \text{Structural equations,} \tag{1}$$

$$\nabla^2 p_1+k^2 p_1=0 \text{ in } \Omega_1 \quad \text{Fluid equations-1,} \tag{2}$$

$$\nabla^2 p_2+k^2 p_2=0 \text{ in } \Omega_2 \quad \text{Fluid equations-2,} \tag{3}$$

$$\left.\frac{\partial p_1}{\partial n}\right|_{x=0}=M_L p_1-2ikp_{1,in} \text{ on } \Gamma_L$$

$$\text{Left DtN boundary,} \tag{4}$$

$$\left.\frac{\partial p_1}{\partial n}\right|_{x=L}=M_R p_1 \text{ on } \Gamma_R \quad \text{Right DtN boundary,} \tag{5}$$

$$\frac{\partial p_1}{\partial z}=\frac{\partial p_2}{\partial z}=\rho_f\omega^2 u \text{ on } \Gamma \quad \text{Euler's relation,} \tag{6}$$

where $\mathbf{u}=u\hat{e}_z$ is the flexible structure displacement ($\hat{e}_z$ is the unit normal along the z direction), p_1 is the fluid pressure in the waveguide, p_2 is the fluid pressure in the backing cavity ($p_2=0$ for single-duct SAS), $\hat{\mathbf{n}}(p_1-p_2)|_\Gamma$ represents the fluid forcing of the flexible structure, $\hat{\mathbf{n}}$ is the unit outward normal to any bounding surface (on Γ, $\hat{\mathbf{n}}$ points along the negative z direction), ρ_f is the fluid density, $\mathcal{L}_s$ is the plate or membrane operator which acts on the displacement vector, ∇^2 is the Laplacian, $k=\omega/c$ is the acoustic wavenumber, M_L and M_R are the DtN operators that convert the infinite domain problem to an equivalent problem on a finite computational domain, and $p_{1,in}$ is the amplitude of the plane wave input from $x=-\infty$. Equation (6) is Euler's equation that relates the normal component of the acceleration to fluid pressure for an inviscid fluid. The equality of normal component of fluid velocity and velocity of the structure is prescribed on both sides of the interface Γ for the two-duct configuration.

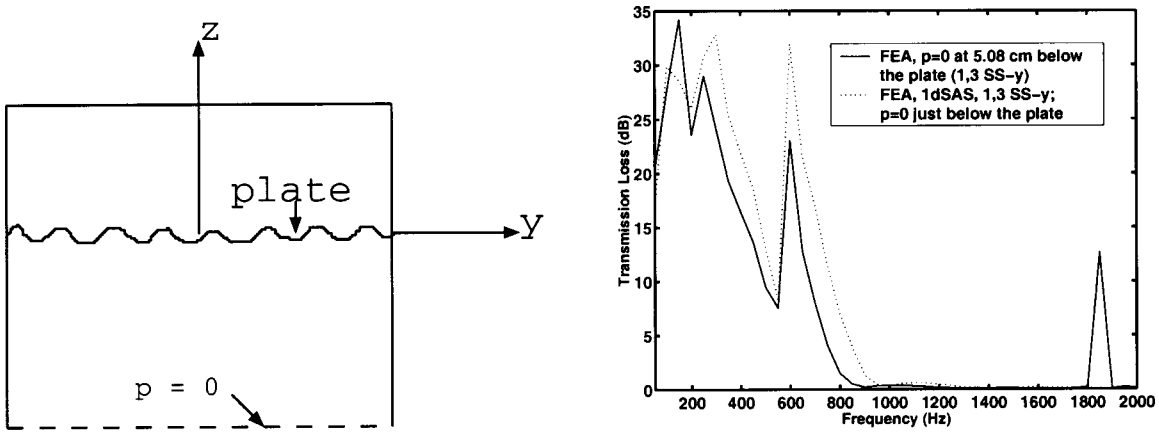
Note that the only modification to the formulation presented in Ramamoorthy *et al.* (2002) is the pressure in the backing cavity. The reader is referred to Ramamoorthy *et al.* (2002) for details on operator $\mathcal{L}_s$, 2.5-d formulation for variational equations, Dirichlet-to-Neumann (DtN) boundary condition and the finite element formulation (2.5-d).

III. EXPERIMENTAL SETUP

The 2.54×2.54-cm^2 test duct consists of three sections each 30.48 cm long. The upstream part, which is made of 1.78-cm-thick acrylic (hard walls), is connected to a 10.16-cm full-range loudspeaker unit. The silencer part consists of a 29.21-cm-long plate and a removable variable-height backing cavity. The downstream part, made of rigid walls, is terminated using foam. The experimental setup is shown in Fig. 3.

The plate made of brass shim is clamped on (glued to) rigid metallic supports. Although tension is not applied to the plate intentionally, it was observed that clamping the plate resulted in tension. The two-microphone transfer function method is implemented to separate incident and reflected waves which are used for the calculation of transmission loss across the silencer section (Chung and Blaser, 1980b, a; Boden and Abom, 1986). One pair of Larson–Davis 0.635 cm ($\frac{1}{4}$-in.) microphones (model 2520, with flat frequency response between 20 Hz and 20 kHz range) is flush-mounted with the duct walls in both the upstream and downstream sections. The separation between each pair of microphones is 3.81 cm in order to avoid singularities in the calculation of the reflection coefficient up to a frequency of 4.5 kHz. Also, the microphones, both upstream and downstream, are placed nearly 14 cm from the silencer region in order to avoid near field and to ensure plane waves at the measurement locations for the frequencies of interest. Both frequency sweep (range: 50 Hz to 2 kHz) and broadband white noise (input) experiments were carried out and the transmission loss results of both techniques agreed as expected.

The microphones are supported by two Larson-Davis preamplifiers (model 2200C) connected to a National Instruments data aquisition board (model PCI-6110E, with 12-bit analog-to-digital converters) and controlled by LabVIEW



(a) Cross section of the duct

(b) Transmission Loss

FIG. 4. Effect of external fluid loading. The plate is 0.002 54 cm thick, 2.54 cm wide, 29.21 cm long, and is simply supported along its transverse boundaries. The main duct height and width are $H_1 = B = 2.54$ cm. Outside the main duct, $p = 0$ is set at $z = -5.08$ cm, i.e., at 5.08 cm below the plate. The other four walls of the backing "cavity" at $x = 0$, $x = L$, $y = \pm B/2$ are rigid. The TL result for the simply-supported plate lateral boundary condition for single-duct SAS (without external fluid) is indicated for reference in Fig. 4(b). In the legend, "1,3 ss-y" indicates that first and third simply-supported transverse plate modes were used in FEA. This notation will also be used in the other figures in this paper, where appropriate.

software. The excitation signal to the loudspeaker is generated by a B&K signal generator (model 1050) for the frequency sweep and by LabVIEW for the white noise. A B&K power amplifier (model 2706) provided all the necessary audio amplification. Although non-plane-waves are excited in the silencer region, the 2.54×2.54-cm^2 duct cross section ensures plane wave propagation in the vicinity of the microphones. Higher-order acoustic modes have a cut-in frequency of 6.75 kHz. A Polytec scanning laser vibrometer (controller model OFV-3001, sensor head 303, scanning controller OFV-042) equipped with a velocity decoder and controlled by LabVIEW is used to scan the surface of the plate through the clear acrylic. The velocity response is referenced to the input pressure in the upstream duct (i.e., first microphone located at $x = -17.78$ cm).

For this experiment, it is impossible to precisely match the theoretical model to the experimental configuration and *vice versa*. In the next two sections, error in the predictions that arise from unmodeled effects (e.g., neglecting the external fluid pressure) and uncertainties in the experiment (e.g., plate boundary conditions) are discussed and analyzed. These sections highlight the sensitivity of the results to these variations.

IV. EFFECT OF EXTERNAL FLUID LOADING IN SINGLE-DUCT SAS

At some frequencies (for instance, above 200 Hz for a 0.002 54-cm-thick brass plate in single-duct SAS with $H_1 = 2.54$ cm), the pressure on the outer surface of the plate in single-duct SAS is experimentally measured to be comparable to the pressure on its inner surface. Thus the approximation that external pressure is zero is not strictly valid. However, the effect of external fluid can be simulated approximately by setting zero pressure at $z = -5.08$ cm (similar to the walls that are present in the experimental configuration) from the bottom surface of the plate, other walls of the "cavity" being rigid. This is equivalent to analyzing a two-duct SAS with pressure release lower wall for the backing cavity with $H_2 = 5.08$ cm [Fig. 4(a)]. The TL predicted using FEA for the 0.002 54-cm-thick brass plate in two-duct SAS with $H_2 = 5.08$ cm with pressure release lower wall for $H_1 = B = 2.54$ cm is shown in Fig. 4(b) for simply-supported lateral plate boundary condition.

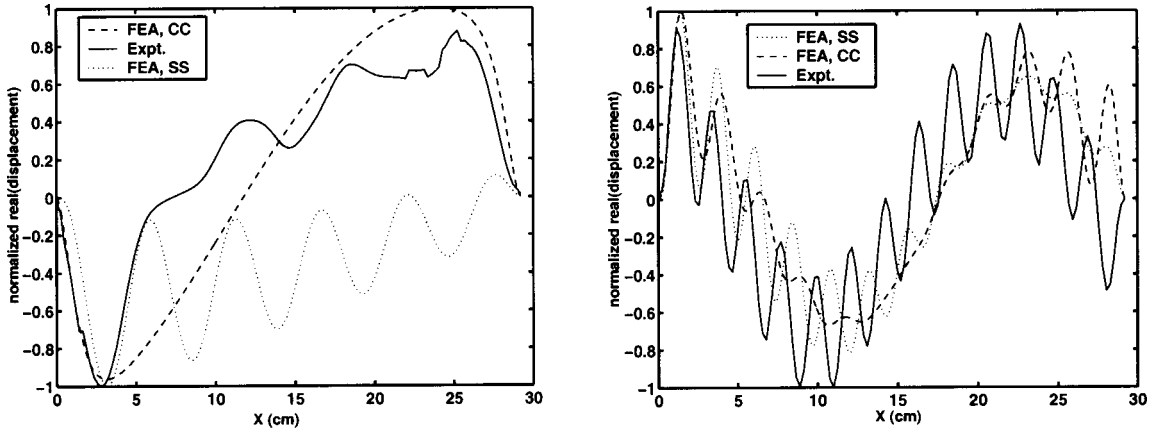
The effect of external pressure is to reduce the resonance frequency and the critical frequency, and to reduce the transmission loss at a given frequency. The reduction in TL is itself a function of frequency because the pressure radiated by the plate is frequency dependent. A similar conclusion can also be drawn from results for clamped transverse plate boundary conditions and the same fluid boundaries as used in Fig. 4(a). Note that the reduction in the resonance frequency and the critical frequency due to external fluid loading are expected because the external fluid effectively mass-loads the plate. A heavier plate results in lower TL, which explains the reduction in TL level due to external fluid loading.

V. EFFECTS OF VARIABILITY ON EXPERIMENTAL RESULTS

Variability in experimental data is introduced by lateral plate boundary conditions, acoustic boundary conditions, asymmetry, tension, and manufacturing tolerance for plate thickness. The extent to which the listed parameters affect the SAS performance will be described in this section using theoretical simulations.

Two plates each made of brass shim, of width 2.54 cm, length 29.21 cm, and thicknesses 0.002 54 and 0.0102 cm, are analyzed. Two heights for the waveguide, $H_1 = 2.54$ cm and $H_1 = 1.27$ cm, are considered. Duct width is $B = 2.54$ cm, the same as plate width. A hysteritic damping factor of 0.05 is used for modeling the plate. These material and geometric data will also be used in the results presented in the next section.

In Figs. 5(a) and (b), the FEA predictions for plate displacement profile using two different boundary conditions



(a) $f = 500$ Hz (b) $f = 1500$ Hz

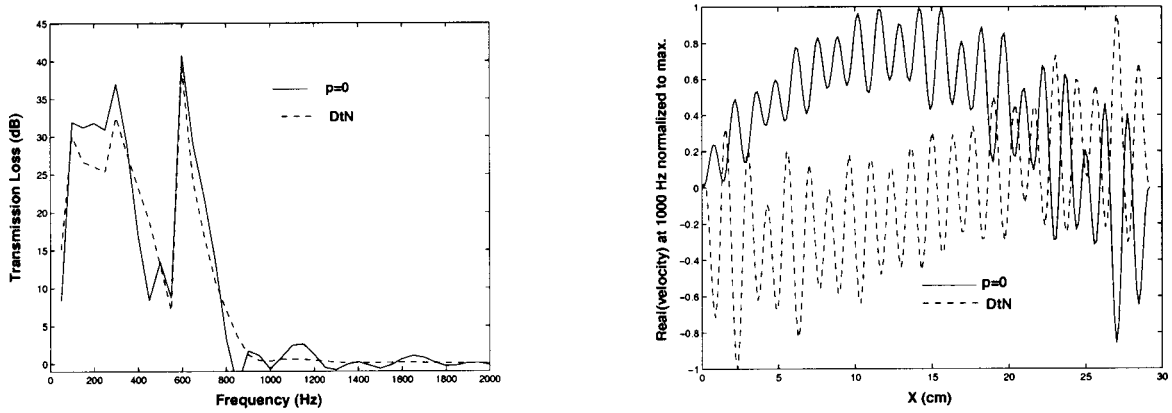
FIG. 5. Plate displacement (normalized to maximum) for single-duct SAS. The brass plate thickness is 0.0102 cm, width is 2.54 cm, and length is L=29.21 cm. The height of the waveguide is $H_1 = 2.54$ cm. The figures show plate displacement at 500 and 1500 Hz. In the legend, "CC" implies clamped and "SS" implies simply-supported lateral plate boundary condition.

are compared to experimental results for the 0.0102-cm-thick brass plate in single-duct SAS with $H_1 = B = 2.54$ cm. The predictions of the plate response are seen to be sensitive to lateral boundary conditions. In Fig. 5(a), the experimental plate response is the sum of a low-wavenumber subsonic wave and a higher-wavenumber subsonic wave. The amplitudes in this figure are normalized to the maximum displacement in each curve. At 500 Hz, the predicted peak displacement response for the simply-supported boundary condition is nearly 2.25 times larger than the corresponding value for the clamped boundary condition in Fig. 5(a), indicating that the clamped plate is stiffer at these frequencies—this will be seen to have an impact on TL. In all the curves shown in Figs. 5(a) and (b), except the low-frequency clamped case, the plate displacement response consists of a sum of the two wave components. These findings highlight the sensitivity of the plate response to boundary conditions.

In the experiments, foam is used for anechoic termination at the right end of the system. Since the absorption by foam is effective only at high frequencies, it is of interest to study the effect of acoustic boundary condition on the plate velocity and transmission loss results.

From FEA, it is observed that the effect of the acoustic boundary condition is to introduce phase shift in the plate velocity (in both subsonic and supersonic modes) and to allot different (relative) energies to each of the allowed modes. This is seen from the real part of the plate displacement [Fig. 6(b)] for the 0.002 54-cm-thick brass plate in single-duct SAS with $H_1 = B = 2.54$ cm. Figure 6(a), which compares TL for DtN and pressure release right boundary at $x = 2L$ for unit velocity piston input, shows that the transmission loss is also affected by boundary conditions, although not dramatically.

Plate tension, which is unavoidable in the experiments, is more significant in determining the response of thinner plates. The following equation relates the effects of tension and bending (focusing on low frequencies, thus modeling for the first resonance mode in the plate):



(a) Transmission Loss (b) Real part of plate velocity normalised to its maximum magnitude

FIG. 6. Effect of acoustic boundary condition. The transmission loss and the real part of plate velocity normalized to its maximum magnitude at 1000 Hz for single-duct SAS with $H_1 = 2.54$ cm, 0.002 54-cm thick brass plate for two different acoustic boundary conditions at the right-end, both for unit piston input, are shown.

$$\frac{K_T}{K_B} = \frac{T((\pi/L)^2 + (\pi/B)^2)}{[Eh^3/12(1-\nu^2)]((\pi/L)^2 + (\pi/B)^2)^2}, \tag{7}$$

where K_T and K_B are the stiffnesses due to tension and bending, respectively. This equation, when applied to a structure made of brass, whose width is 2.54 cm, for an applied uniform isotropic tension per unit width of 39 000 dynes/cm (corresponds to load due to mass of 100 g), shows that the effect of tension is higher than bending for brass plates (rather membranes) of thicknesses less than 0.0066 cm. For a 0.0066-cm-thick brass plate of width 2.54 cm, the effect of tension and bending are equal for an applied tension per unit width of 39 000 dynes/cm. This critical plate thickness is higher for tension per unit width greater than 39 000 dynes/cm and lower for an applied tension per unit width less than 39 000 dynes/cm. For the thicknesses considered in this paper, therefore, the effect of tension is high for the 0.002 54-cm-thick brass plate ($K_T/K_B \approx 15.75$ for $T = 39\,000$ dynes/cm), whereas it is relatively insignificant for the 0.0102-cm-thick brass plate ($K_T/K_B \approx 0.25$ for $T = 39\,000$ dynes/cm).

Although the SAS system is designed to be symmetric in the lateral direction, asymmetry is unavoidable in experiments (e.g., due to fixing the lateral plate ends). Therefore, antisymmetric lateral structural modes cut in the experimental setup, whereas the modeled system is symmetric. This is another source of difference between FEA predictions and experimental results, as well as variability of experimental data from two different setups.

The manufacturing tolerance for plate thickness is 10% for the 0.002 54-cm-thick brass plate (supplied by McMasterCarr). FEA simulations show that the TL resulting from a 10% increase in plate thickness (to 0.002 79 cm) has a very similar character to that of the 0.002 54-cm-thick brass plate [e.g., see the TL for the 0.002 54-cm-thick brass plate in a single duct in Fig. 4(b)]. However, an increase in the plate thickness by 10% increases the lowest frequency for which TL occurs by 50 Hz in this case.

VI. TRANSMISSION LOSS: FEA VERSUS EXPERIMENT

A. Single-duct structural acoustic silencer (single-duct SAS)

In this section, the transmission loss predicted by FEA will be compared with experimental results for single-duct SAS (for model, see Fig. 1). Transmission loss predictions are obtained by postprocessing the results of the FEA (for details, see, Ramamoorthy *et al.*, 2002).

The FEA and experimental results for transmission loss performance of single-duct SAS with $H_1 = B = 2.54$ cm and brass plate of thickness $h = 0.0102$ cm is shown in Fig. 7. The FEA results are shown for two different lateral plate boundary conditions. It is observed from Fig. 7 that the TL bandwidth is between the resonance frequency of the first plate lateral mode and the first critical frequency. The critical frequency is defined by Huang (1999) as the frequency at which the supersonic wave cuts-in in the plate.

To clarify the TL bandwidth further, consider the effect of duct height on TL examined in Fig. 8 for a brass plate of thickness 0.0254 cm. For this choice of material and geometric properties, when duct height increases the critical frequency decreases. Since the plate properties are the same in each case, the uncoupled plate resonance frequency remains the same. Hence, the starting frequency of the TL band remains the same, while the end frequency of the TL band decreases with increase in H_1. One has to be careful in interpreting the increase or decrease of critical frequency with increase in the duct height. From Huang (1999), one can notice that there are two opposing factors that determine the critical frequency as a function of duct height. As the duct height H_1 is increased, the ratio of the mass of the plate with respect to the mass of the fluid (M) decreases, thereby increasing the nondimensionalized critical frequency, but the dimensional factor ($c/2H_1$, c is the speed of sound in the fluid) decreases. Since the critical frequency is a product of these two parameters, the net effect of increasing the duct height can be a decrease in critical frequency (if the mass

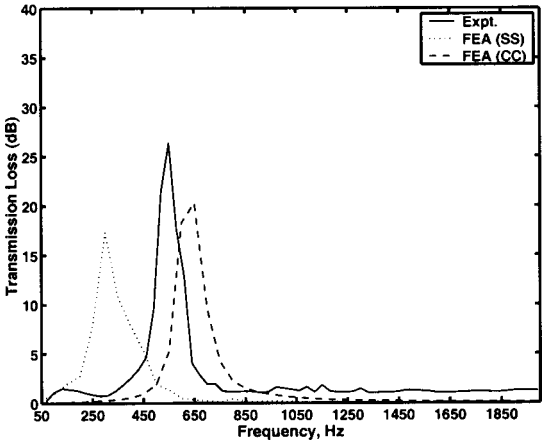


FIG. 7. Transmission loss for single-duct SAS. The brass plate width is 2.54 cm, length is 29.21 cm, and thickness is 0.0102 cm. The height of the waveguide is $H_1 = 2.54$ cm. In the legend, "CC" implies clamped and "SS" implies simply-supported lateral plate boundary condition. Only the first lateral plate mode is used in FEA result.

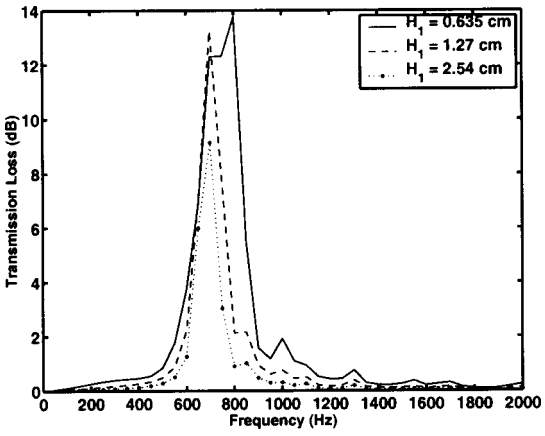
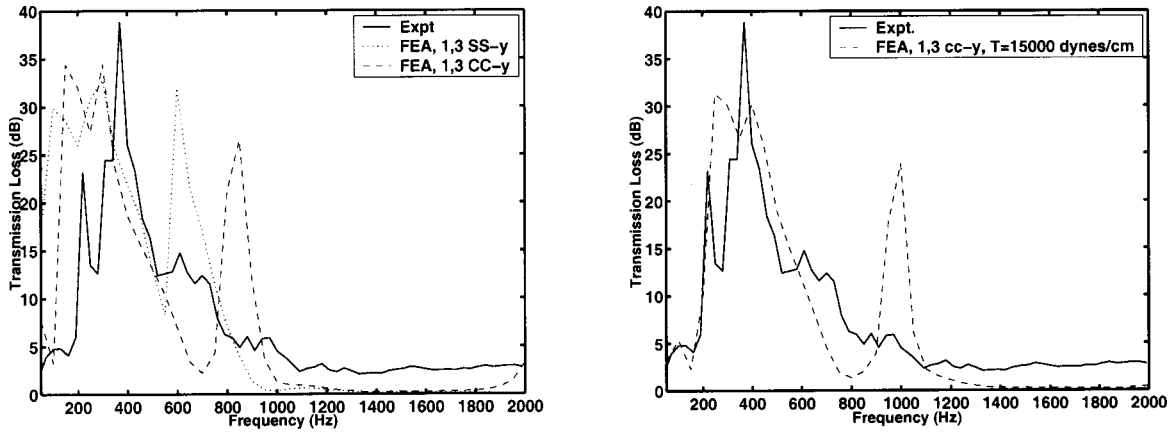


FIG. 8. Transmission loss for single-duct SAS using FEA. The brass plate width is 2.54 cm, length is 29.21 cm, and thickness is 0.0254 cm. Three different values for height of the waveguide are shown in the legend. This figure shows the TL frequency band to be from the resonance frequency of first lateral plate mode until the critical frequency. Only the first lateral plate mode is used in FEA result (with simply-supported boundary condition).



(a) $T = 0$ (b) $T = 15000\,dynes/cm$

FIG. 9. TL for single-duct SAS for a brass plate of thickness 0.002 54 cm, width 2.54 cm, and length 29.21 cm; $H_1 = B = 2.54$ cm. FEA simulations for zero tension and for applied uniform isotropic tension per unit width of $T = 15\,000$ dynes/cm with clamped transverse boundary conditions for the plate are compared with experimental result where the unintentionally applied tension is experimentally estimated to be roughly 15 000 dynes/cm assuming uniformity.

ratio M is high) or an increase in critical frequency (if M is relatively low).

TL is nonzero when the coupled mode is structure type, i.e., when the coupled mode carries most of its power flow in the structure [see Cummings (2001) for a discussion on acoustic and structure type modes]. When the lateral free structural mode cuts in, there is subsonic wave in the plate, which is evanescent in the fluid—the fluid can only propagate plane and supersonic waves. There is reflection due to modal mismatch when the plate wavenumber is subsonic and the coupled mode is structure type. However, because the fluid can propagate supersonic waves, the TL is nearly zero once the supersonic wave dominates (that is, when the coupled mode is acoustic type). Therefore, there is nonzero TL from the cut-in of the first structural mode until the first critical frequency. Once again, as the higher structural modes cut in and the coupled mode is again structure type, there is nonzero TL from the cut-in of this higher structural mode until the supersonic mode dominates again. This also explains why the TL frequency band is different for different lateral plate boundary conditions because the resonances of the free lateral structural modes depends on the lateral plate boundary condition (as seen in, e.g., Fig. 7).

The results for 0.0102-cm-thick brass plate mentioned above suggest that a thinner plate would perform better in terms of broadband low frequency TL because the resonance frequency of the first free lateral structural mode is lower and critical frequency higher [see Huang (1999) for critical frequency as a function mass of the plate] for a thinner plate, other parameters being fixed.

Figure 9(a) compares FEA (for plate in pure bending) and experimental results for TL for 0.002 54-cm-thick, 2.54-cm-wide and 29.21-cm-long brass plate in single-duct SAS with $H_1 = B = 2.54$ cm. FEA simulations and experimental results differ in frequency location of TL peak as well as TL level. This is mainly due to unintentionally applied tension in the plate in experiments. A rough estimate of the average value of tension in the plate determined experimentally using the first resonance frequency of the 0.002 54-cm-thick brass plate is nearly 15 000 dynes/cm (corresponds to load due to mass of 40 g) for clamped plate boundary condition. Predicted TL results for the 0.002 54-cm-thick brass plate for an applied uniform isotropic tension per unit width of 15 000 dynes/cm, and assuming clamped lateral plate boundary conditions simulated using FEA, are shown in Fig. 9(b).

From Fig. 9(b), it is seen that the start of the TL band in frequency in the experimental data is matched by the FEA simulation that accounts for tension in the plate. This is because the resonance frequency of the first lateral plate mode is matched by accounting for tension in the plate. Further differences in TL between the FEA simulation (with $T = 15\,000$ dynes/cm^2) and the experiment are probably due to nonuniformity of tension in the experiment, presence of second (antisymmetric) lateral plate mode (this mode introduces additional TL peak between those due to the first and third lateral plate modes), differences in structural as well as acoustic boundary conditions, and external fluid loading.

Comparing the TL attained by the 0.002 54-cm-thick brass plate [Fig. 9(a)] versus that by the 0.0102-cm-thick brass plate (Fig. 7), it is clear that the TL bandwidth and level are lower for the 0.0102-cm-thick brass plate compared to the 0.002 54-cm-thick brass plate. This is expected because the resonance frequency of the first plate cross mode is higher and the critical frequency lower for a thicker plate. One can intuitively expect that since the impedance jump experienced by the incoming wave is lower in case of thicker plate, the TL level is lower too.

The transmission loss plot for 0.002 54-cm-thick brass plate with $H_1 = 1.27$ cm in Fig. 10 shows increase in TL bandwidth with decrease in duct height for a given plate thickness due to increase in critical frequency. Also observe the existence of two distinct peaks in the FEA results for TL in Figs. 9(a) and 10. The second of the two peaks in the FEA results is due to the cut-in of the third lateral plate mode.

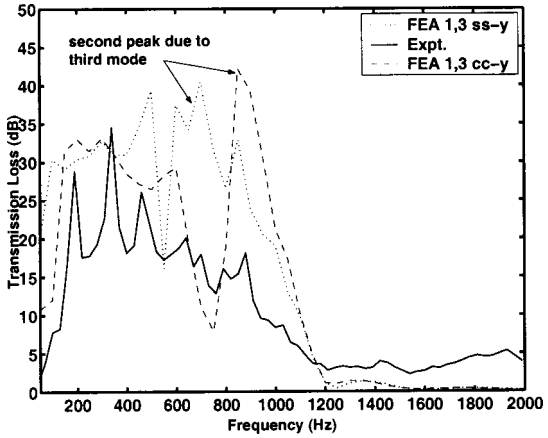


FIG. 10. Transmission loss for single-duct SAS. The brass plate thickness is 0.002 54 cm, width is 2.54 cm, and length is 29.21 cm. The height of the duct is $H_1 = 1.27$ cm. The second of the two peaks seen in FEA results is due to the cut-in of the third lateral plate mode.

B. Two-duct structural acoustic silencer (two-duct SAS)

In this section, the TL results for two-duct SAS model using FEA and experiment are presented. The dimensions of the main duct and material and geometry for the plate are same as for single-duct SAS. Two heights of the backing cavity are considered: $H_2 = 2.54$ and 25.4 cm.

The FEA and experimental results for the 0.002 54-cm-thick, 2.54-cm-wide, 29.21-cm-long brass plate in two-duct SAS with $H_1 = 2.54$ cm and $H_2 = 2.54$ and 25.4 cm are shown in Figs. 11(a) and (b), respectively. The results for the 0.0102-cm-thick brass plate in two-duct SAS [other data same as for Figs. 11(a) and (b)] are shown in Figs. 12(a) and (b). The performance of the corresponding expansion chambers (results of FEA) are indicated for reference. FEA results are shown for simply-supported as well as for clamped lateral plate boundary conditions. The improvement in TL due to two-duct SAS compared to the expansion chamber is better for cavity with $H_2 = 2.54$ cm compared to $H_2 = 25.4$ cm cavity. Even for the cavity with $H_2 = 25.4$ cm, TL at low frequencies are improved, though at the expense of losing the TL peak at a higher frequency. For example, in Fig. 11(b), the TL peak at 1500 Hz seen in the expansion chamber result is lost in the result for the corresponding two-duct SAS, but there is an improvement in TL (nearly 10 dB) around 200–600 Hz.

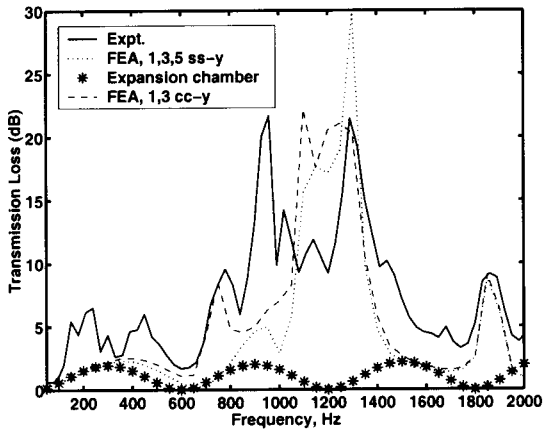
Comparing the performance of the 0.002 54-cm-thick brass plate and 0.0102-cm-thick brass plate, it is noted that the 0.002 54-cm-thick brass plate [Fig. 11(a)] has higher TL level and broader frequency band compared to the 0.0102-cm-thick brass plate [Fig. 12(a)]. Another interesting fact is that the 0.002 54-cm-thick brass plate has its TL bandwidth shifted to higher frequencies compared to that of the 0.0102-cm-thick brass plate. This trend is opposite to corresponding results for single-duct SAS.

Including a cavity under the plate increases the resonance frequencies of the system because the cavity results in an increase in stiffness. The cavity stiffness is proportional to $\rho c^2/V$ where V is the volume of the cavity. As cavity height H_2 increases, the cavity stiffness decreases. Therefore the increase in resonance frequency due to cavity stiffness reduces with H_2, as observed in the FEA simulations. Note that this is different from mass loading the plate which reduces the resonance frequency of the plate compared to its *in vacuo* value. If there is no cavity, fluid results in mass loading the plate (due to radiation impedance on the vibrating plate).

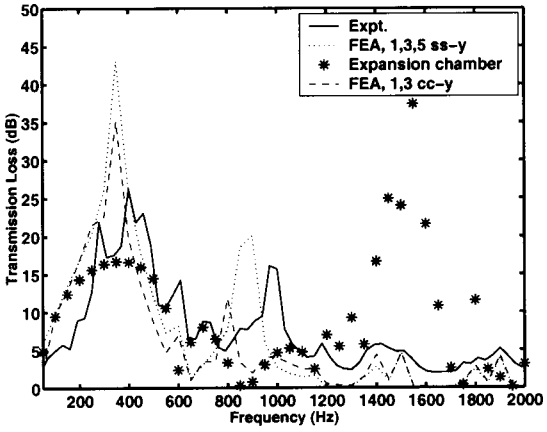
To quantitavely estimate the increase in fundamental resonance frequency of the system due to inclusion of lower cavity, consider the following approximate formula given by Dowell and Voss (1963) (under the notation used in this paper):

$$\omega_c^2 = \omega^2 + K \frac{\rho}{\rho_s} \frac{c^2}{H_2 t}, \tag{8}$$

where K is a constant determined by boundary conditions on the plate. For a plate clamped on all four sides $K = 0.44$, while for a simply-supported plate $K = 0.66$. On the right-hand side of Eq. (8), the first term is the *in vacuo* plate fundamental resonance frequency and the second term is due

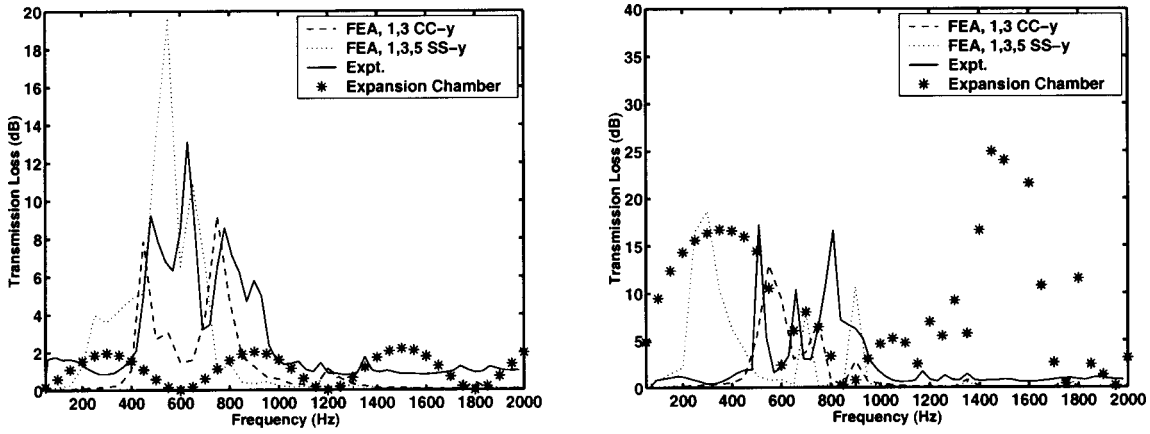


(a) $H_1 = 2.54\,cm$, $H_2 = 2.54\,cm$



(b) $H_1 = 2.54\,cm$, $H_2 = 25.4\,cm$

FIG. 11. Transmission loss for two-duct SAS. The brass plate width is 2.54 cm, length is 29.21 cm, thickness is 0.002 54 cm, cavity height $H_2 = 2.54$ cm, 25.4 cm, and duct height $H_1 = 2.54$ cm. The expansion chamber results shown in the figure are based on FEA.



(a) $H_1 = 2.54\,cm$, $H_2 = 2.54\,cm$ (b) $H_1 = 2.54\,cm$, $H_2 = 25.4\,cm$

FIG. 12. Transmission loss for two-duct SAS. The brass plate width is 2.54 cm, length is 29.21 cm, thickness is 0.0102 cm, cavity height $H_2 = 2.54$ cm, 25.4 cm, and duct height $H_1 = 2.54$ cm. The expansion chamber results shown in the figure are based on FEA.

to fluid loading on the plate. Note the inverse dependence of the second term on plate thickness, t, and cavity height, H_2.

Pretlove (1965) concludes that there are two distinct types of plate-cavity systems—one where the action of the closed cavity on the plate is negligible, characterized by acoustic stiffnesses considerably less than panel stiffnesses, and the other where there is considerable modification of the plate mode shapes and natural frequencies, characterized by acoustic stiffnesses of the same order or greater than the plate stiffness. For the current system, the 0.002 54-cm-thick brass plate over a $H_2 = 2.54$ cm cavity [cf. Fig. 11(a)] falls under the second category of plate cavity systems described by Pretlove, whereas the 0.0102-cm-thick brass plate over a $H_2 = 2.54$ cm cavity [cf. Fig. 12(a)] corresponds to the first category.

The TL for the 0.0102-cm-thick brass plate with $H_2 = 2.54$ cm cavity is more sensitive to boundary condition compared to TL for the 0.002 54-cm-thick brass plate in the same model, because the first term, which is more sensitive to boundary condition, is the dominant term in this case compared to the second term due to fluid loading. This fact is also reflected in the experimental results and their comparison with FEA predictions in Figs. 12(a) and (b).

Figure 13 shows the TL result for two-duct SAS with 0.002 54-cm-thick brass plate, $H_1 = 1.27$ cm, $H_2 = 25.4$ cm, $B = 2.54$ cm. The TL is above 15 dB until 550 Hz. Once again, the improvement in TL (of around 10 dB) over the corresponding expansion chamber (expansion chamber results shown are results of three-dimensional FEA) is seen at low frequencies, at the expense of losing the TL peak at 1500 Hz that is observed in the result for the expansion chamber.

Note that it is not possible to discuss the TL as related to plate dispersion alone for the two-duct structural acoustic silencer because the important characteristic for TL is low impedance of the plate–cavity system as experienced by the wave in the main duct. The impedance of the plate–cavity system at low frequencies is approximately equal to the difference between the plate impedance and the cavity impedance. The shifted resonances described above using the approximate formula [Eq. (8)] are the frequencies at which the net impedance of the plate–cavity system is minimum. Therefore the locations of the TL peak using either single- or two-duct SAS are determined by the frequencies of minimum impedance for the silencer system as seen by the acoustic wave in the main duct.

VII. PLATE WAVENUMBERS IN SINGLE-DUCT SAS: FEA VERSUS EXPERIMENT

The predicted and experimentally determined wavenumbers in a plate of thickness 0.002 54 cm, length 29.21 cm, and width 2.54 cm in a single-duct SAS with $H_1 = B = 2.54$ cm are presented in Figs. 14(a) and (b), respectively. The spectrum is estimated by taking a spatial FFT of the predicted and measured velocity response. In Figs. 14(a) and (b), the amplitude of the spectrum normalized to the maximum value is plotted. All components below 0.5 of the maximum are not plotted, as the additional clutter makes the plots hard to interpret.

The predicted plate dispersion [Fig. 14(a)] is very similar to the coupled-mode dispersion presented in Fig. 3 in Cummings (2001). The first coupled mode is acoustic type at

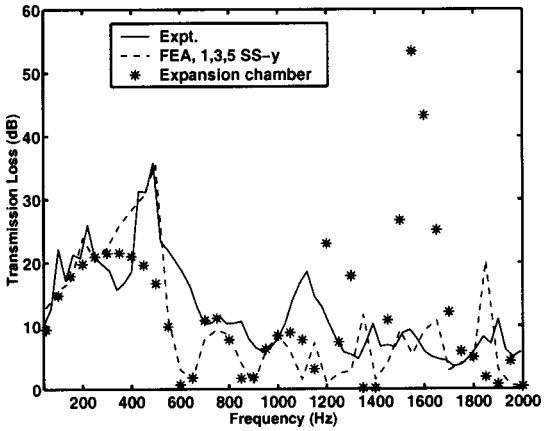


FIG. 13. Transmission loss for two-duct SAS. The brass plate thickness is 0.002 54 cm, width is 2.54 cm, and length is 29.21 cm. The height of the waveguide is $H_1 = 1.27$ cm and height of backing cavity is $H_2 = 25.4$ cm. The expansion chamber results shown in the figures are based on FEA.

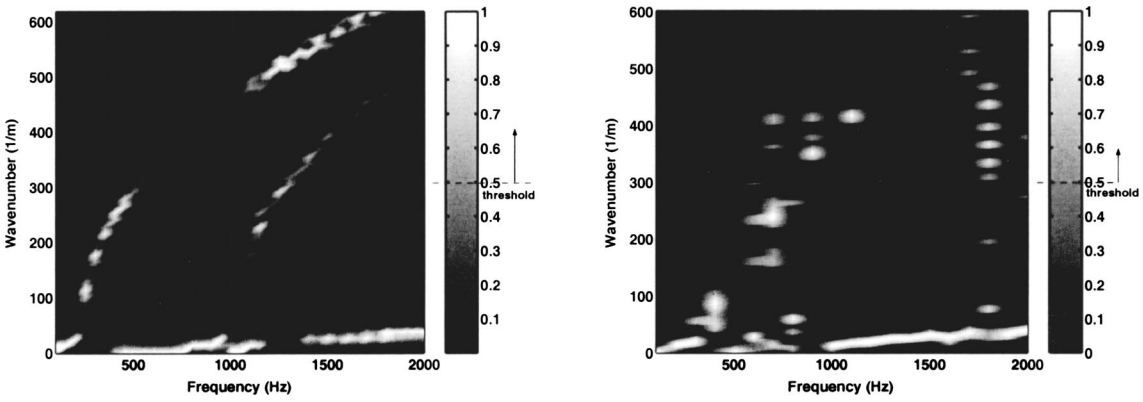


FIG. 14. Dispersion in a brass plate of thickness 0.002 54 cm and width 2.54 cm in single-duct SAS with $H_1=2.54$ cm, $L=29.21$ cm. FEA results are for clamped lateral plate boundary condition using first and third lateral plate modes. Uniform isotropic applied tension per unit width of $T=15\,000$ dynes/cm (determined experimentally as rough estimate) is used in this simulation. If no tension is applied, the theoretically predicted wavenumber-frequency plots remain nearly the same with a slight shift to the left of the corresponding plots shown for nonzero tension. The shift is due to reduced resonance frequency of lateral plate mode for zero tension. In the above figures, a threshold of 0.5 is used for the energy.

low frequencies, that is, most of the power flow is in the air. When the first free lateral structural mode cuts in [around 250 Hz in Fig. 14(a)], the mode becomes structure type, indicating that most of the power flow is in the plate. Note that the first coupled mode is low-wavenumber subsonic (very close to being a plane wave) for frequencies below 250 Hz, changing to high-wavenumber subsonic after the cut-in of the first free lateral structural mode. This is also seen (though slightly distorted) in the experimental result.

Around 400 Hz, the second coupled mode is seen in Fig. 14(a) with a supersonic wavenumber (acoustic type), changing to high-wavenumber subsonic after the cut-in of the second symmetric free lateral structural mode at 950 Hz. The third coupled mode is seen in Fig. 14(a) starting at 1000 Hz with a supersonic wavenumber. When the third symmetric free lateral structural mode cuts in, the coupled mode changes to structural type with a subsonic wavenumber, and so on.

Note that when the coupled mode is structural-type, experimental data [Fig. 14(b)] are complicated probably due to the nonuniformity in the unintentionally applied tension, lateral boundary conditions, and the cut-in of the antisymmetric free lateral structural modes. However, when the coupled mode is acoustic type, the experimental data in Fig. 14(b) closely follows the predictions in Fig. 14(a).

The implication of such a dispersion phenomena on the TL performance of the single-duct SAS is that the TL is very low in those regions of frequencies where the coupled mode is acoustic type. When the mode is structural type, a high-wavenumber subsonic mode exists in the plate and therefore there is high TL (due to modal mismatch, as mentioned earlier). This is seen clearly by comparing Figs. 14(a) and (b) with Fig. 9(b).

VIII. ANALYTICAL STUDY OF STRUCTURAL MODE COUPLING IN SINGLE-DUCT SAS

The interaction of structural cross modes is investigated in this section using Green's function to derive analytical expression for the modal radiation impedance (which leads to scattered pressure) to explain the observations related to cross mode interaction in single-duct SAS.

Consider the Green's function satisfying rigid duct boundary conditions including DtN boundaries at $x=0$ and $x=L$, given by [$e^{i\omega t}$ time harmonic dependence is assumed in this derivation in order to be compatible with reference to Huang (1999) for some simplifications in the final expressions]

$$g(\mathbf{x}-\mathbf{x}_0)=\sum_{m,n}\frac{2\pi i}{-k_x\beta_{mn}(y_0,z_0)}e^{-ik_x|x-x_0|}\phi_{mn}(y,y_0,z,z_0), \tag{9}$$

where $\phi_{mn}(y,y_0,z,z_0)=\cos(m\pi/B)(y-y_0))\cos(n\pi/H_1)(z-z_0))$, $\beta_{mn}(y_0,z_0)=BH_1/(2-\delta_m)(2-\delta_n)$ and $k_x=\sqrt{(\omega/c)^2-k_{mn}^2}$. Here, $k_{mn}^2=(m\pi/B)^2+(n\pi/H_1)^2$.

Assuming modal sum for the plate displacement (in this derivation, the y-coordinate spans from 0 to B in order to simplify expressions),

$$u(x,y)=\sum_{r,s}W_{rs}\sin\left(\frac{r\pi x}{L}\right)\sin\left(\frac{s\pi y}{B}\right). \tag{10}$$

The scattered pressure due to the plate vibration is given by

$$p_{sc}(x,y,z)=-\frac{\rho\omega^2}{4\pi}\int_{x_0=0}^{x_0=L}\int_{y_0=0}^{y_0=B}g(x,y,x_0,y_0,z,z_0=0)\times u(x_0,y_0)dx_0\,dy_0, \tag{11}$$

which, at $z=0$, simplifies to ($m=0$ only)

$$p_{sc}(x,y,z)=\frac{\rho\omega^2}{4\pi}\sum_n\left(\frac{2\pi i}{k_x}\right)\left(\frac{2-\delta_n}{BH_1}\right)\sum_{r,s}\frac{2B}{s\pi}W_{rs}\times\int_{x_0=0}^{x_0=L}\sin\left(\frac{r\pi x_0}{L}\right)e^{-ik_x|x-x_0|}\,dx_0. \tag{12}$$

The governing equation, when the input is a plane wave of magnitude p_{in}, is given by

$$\mathcal{L}_s u = -(p_{in} + p_{sc})|_{z=0}, \tag{13}$$

where $\mathcal{L}_s$ is the plate operator introduced in Eq. (1). Choosing the number of x, y modes for the plate displacement and the number of y, z modes for the fluid pressure, the above governing equation reduces to the matrix equation

$$Z_{cd} W_{cd} + \sum_{r,s} I_{cdrs} W_{rs} = -f_{cd}, \tag{14}$$

where Z is the plate impedance given by (for isotropic plate in bending)

$$Z_{cd} = \frac{BL}{4}\left(D\left(\left(\frac{c\pi}{L}\right)^4 + 2\left(\frac{c\pi}{L}\right)^2\left(\frac{d\pi}{B}\right)^2 + \left(\frac{d\pi}{B}\right)^4 \right) - \rho_s h \omega^2 \right), \tag{15}$$

where D is the bending rigidity of the plate, f is the forcing on the plate due to the plane wave input of magnitude p_{in} given by

$$f_{cd} = \int_{x=0}^{x=L} \int_{y=0}^{y=B} p_{in} \sin\left(\frac{c\pi x}{L}\right) \sin\left(\frac{d\pi y}{B}\right) dx\, dy, \tag{16}$$

and I is the modal radiation impedance given by

$$I_{cdrs} = 2i\rho\omega^2 \frac{BL^2}{\pi^2 H_1} \frac{1}{ds} \sum_n \frac{(2-\delta_n)}{k_x} J_{nrc}, \tag{17}$$

where

$$J_{nrc} = \int_{x=0}^{x=L} \frac{1}{L} \sin\left(\frac{c\pi x}{L}\right) \int_{x_0=0}^{x_0=L} \frac{1}{L} \sin\left(\frac{r\pi x_0}{L}\right) \times e^{-ik_x|x-x_0|}\, dx_0\, dx. \tag{18}$$

Simplification for this expression is given by Huang (1999) in $I_2(m,n,j)$ after replacing m by n, n by r, and j by c in that expression. In the above analysis, c, d and r, s represent two sets of indices for x, y modes for the plate displacement, only zeroth pressure mode is considered in the y direction, and the summation over n is to account for the finite number of pressure modes in the z direction.

From Eq. (17), note that the modal radiation impedance is inversely proportional to the duct height H_1. Hence, the effective scattered pressure due to the third lateral plate mode loading the first lateral plate mode is higher for the case of $H_1 = 1.27$ cm, indicating better interaction between the third and the first lateral plate modes compared to that for $H_1 = 2.54$ cm. The idea is also supported by peak plate displacement corresponding to the first mode as a function of frequency with and without the inclusion of third mode for the two duct heights considered. For $H_1 = 1.27$ cm, the inclusion of third mode increases the peak plate displacement corresponding to the first mode, whereas for $H_1 = 2.54$ cm, the peak plate displacement of the first mode remains almost the same with and without inclusion of the third mode.

IX. SIGNIFICANCE OF THREE-DIMENSIONAL ANALYSIS FOR SAS DESIGN

Figure 15 compares FEA results for TL for two- and three-dimensional models for single duct structural acoustic silencer with $H_1 = B = 2.54$ cm for plate thickness 0.0102 cm, width 2.54 cm, and length 29.21 cm. The two-dimensional analysis [e.g., analysis by Huang (1999)] predicts TL until the critical frequency, but does not bring out the fact that TL is also negligible below the resonance frequency for first plate cross mode. Also, the TL peaks due to higher lateral plate modes are not brought out by the two-dimensional analysis.

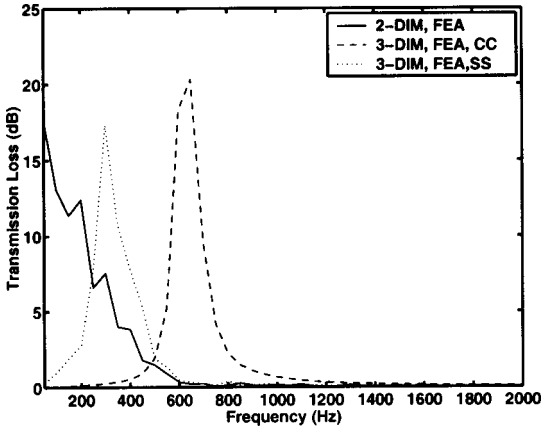


FIG. 15. Transmission loss for two- versus three-dimensional single-duct SAS. The brass plate width is 2.54 cm, length is 29.21 cm, and thickness is 0.0102 cm. Two- and three-dimensional FEA results are compared in this figure. For the three-dimensional results, only the first lateral plate mode is used in FEA.

To see how the two-duct system is affected by two-dimensional analysis, FEA simulations for the two-dimensional model are compared with corresponding results in three dimensions in Fig. 16 for the 0.0254-cm-thick, 2.54-cm-wide, and 29.21-cm-long brass plate in two-duct SAS with $H_1 = B = H_2 = 2.54$ cm. The TL peak in the two-dimensional result is at much lower frequency compared to TL peak in the three-dimensional result, because the *in vacuo* lateral mode resonance frequency of the plate is effectively zero in two-dimensional analysis, whereas it is finite in three-dimensional analysis. The TL level and frequency band can be in significant error in two-dimensional analysis.

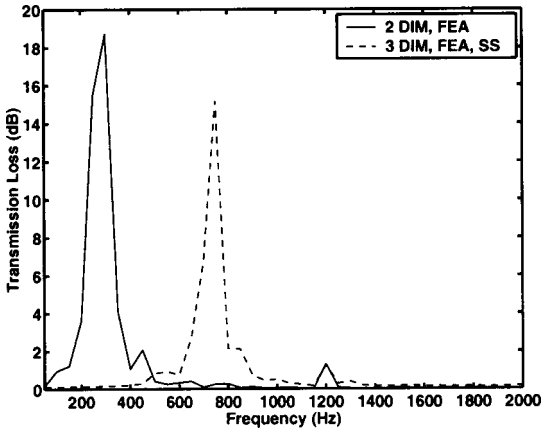


FIG. 16. Transmission loss for two- versus three-dimensional two-duct SAS. The brass plate thickness is 0.0254 cm, width is 2.54 cm, and length is $L = 29.21$ cm. The height of the waveguide and the backing cavity are both 2.54 cm. The FEA results for two- and three-dimensional models are compared to show the necessity of three-dimensional analysis in determining the response of the two-duct structural acoustic silencer. For the three-dimensional results, only the first lateral plate mode is used in FEA.

Therefore, the design of structural acoustic silencer demands three-dimensional analysis unless the cross section is wide and the plate is unsealed.

X. ADDITIONAL CONSIDERATIONS FOR PRACTICAL APPLICATION OF SAS

The examples presented in this paper are for proof of concept for SAS design. In a practical application of structural acoustic silencers, one has to take into account the following additional considerations. The TL level scales with plate length. It is appropriate to use as long a device as the application might permit. A wider duct results in greater TL at lower frequencies because the plate resonance frequency decreases with plate width. Using a material with lower value of E/ρ_s (E is the Young's modulus and ρ_s is the density of the plate) also leads to lower resonance frequencies. For example, a rubber or elastic polymer has two orders of magnitude lower value of E/ρ_s compared to brass. However, if the application involves high temperature gases, such materials may not be suitable for the purpose.

Introducing asymmetry purposely will result in additional TL peaks in a single-duct SAS due to the antisymmetric lateral plate modes. This is beneficial because the frequency band for TL is broader with the additional peaks included.

Finally, one can include as many SAS devices as desired and permitted in an application. These devices can include different plates, even a combination of single and two-duct SAS, depending on the frequencies at which TL is desired.

XI. CONCLUSIONS AND FUTURE WORK

The FEA predictions compare quite well to experimental observations for single-duct (single-duct SAS) and two-duct (two-duct SAS) models for the two plate thicknesses considered. Transmission loss for single-duct SAS is between the resonance frequency for the first plate cross mode and the critical frequency for the system, with TL at narrow frequency bands above the critical frequency from cut-in of the higher transverse mode in the plate until the supersonic mode dominates again. The TL level and bandwidth are higher for thinner plate. The plate cross modes play a significant role in determining the system performance, including transmission loss, which can be directly associated with the corresponding plot of wavenumbers in the plate. The coupling between the cross modes is higher for single-duct SAS with lower duct height and for thinner plates. The system is sensitive to both acoustic as well as plate boundary conditions. The plate boundary in experiments is a blend of simply-supported and clamped boundary conditions. Tension, which is applied unintentionally in the experiments, is important in determining the response of the thin plate considered in this study, whereas it can be neglected in the analysis of the thicker plate. The external fluid does play a role in the response of a single-duct SAS. Neglecting the external fluid in modeling the single-duct SAS leads to slightly higher critical frequency and frequency-selective higher TL.

A two-duct SAS can be designed to perform better than an expansion chamber of the same geometry. Increasing the height of the lower cavity reduces the frequency at which the TL peaks in two-duct SAS due to the reduced shift in plate resonance frequency attributed to cavity fluid loading. One such model which has no issue of break out noise is shown experimentally as well as using FEA to give more than 15 dB TL over a frequency band of 50–550 Hz (Fig. 13); this model is better than the corresponding expansion chamber at low frequencies though at the expense of losing a high-frequency TL peak.

Based on comparing the two- and three-dimensional models for two-duct SAS, it is noticed that the TL frequency band can be artificially shifted to lower frequencies based on the plate thickness (shift being higher for higher plate thickness) if only a two-dimensional model is used. However, this fact can be used in practice by setting the lateral plate ends free so that the *in vacuo* resonance of the first lateral mode is nearly zero and cavity fluid loading is the only cause for shift in TL frequency band to higher frequencies. Another possible means of realizing low resonance frequency of the lateral plate mode is to laminate the thick plate over a very thin plate to reduce the boundary impedance. Such a shift to lower frequencies is a very desirable feature.

A hole in the plate (at $x=L$) connecting the fluid in the main duct and the lower cavity allows for static pressure equilibration and introduces an additional degree of freedom (with low resonance frequency) to couple with the plate-cavity system. Such a design provides another way for low-frequency transmission loss using structural acoustic silencers and hence needs to be explored for noise control purposes.

In general, to design the structural acoustic silencer with any additional modifications to the systems presented in this paper, one needs to look at the effective impedance experienced by the wave in the fluid in the main duct at the silencer location. To achieve high TL at low frequencies, this impedance should be minimum at low frequencies. In order to increase the TL level and bandwidth, the magnitude of the effective impedance at frequencies around the TL peak should be reduced over a wide band. The future challenge is therefore to come up with better models that have such an impedance mechanism.

ACKNOWLEDGMENTS

This research is funded by the Office of Naval Research and the National Science Foundation.

Boden, H., and Abom, M. (**1986**). ''Influence of errors on the two-microphone method for measuring acoustic properties in ducts,'' J. Acoust. Soc. Am. **79**, 541–549.

Choy, Y. S., and Huang, L. (**2002**). ''Experimental studies of a drumlike silencer,'' J. Acoust. Soc. Am. **112**, 2026–2035.

Chung, J. Y., and Blaser, D. A. (**1980a**). ''Transfer function method of measuring in-duct acoustic properties II. experiment,'' J. Acoust. Soc. Am. **68**, 914–921.

Chung, J. Y., and Blaser, D. A. (**1980b**). ''Transfer function method of measuring in-duct acoustic properties. I. theory,'' J. Acoust. Soc. Am. **68**, 907–913.

Cummings, A. (**2001**). ''Sound transmission through duct walls,'' J. Sound Vib. **239**, 731–765.

Dowell, E. H., and Voss, H. M. (**1963**). ''The effect of a cavity on panel vibration,'' AIAA J. **1**, 476–477.

Grosh, K., and Dodson, J. (**1999**). ''Cochlear-based transducer designs,'' in *Proceedings ASME IMECE*.
Huang, L. (**1999**). ''A theoretical study of duct noise control by flexible panels,'' J. Acoust. Soc. Am. **106**, 1801–1809.
Huang, L. (**2002**). ''Modal analysis of a drumlike silencer,'' J. Acoust. Soc. Am. **112**, 2014–2025.
Huang, L., Choy, Y. S., So, R. M. C., and Chong, T. L. (**2000**). ''Experimental study of sound propagation in a flexible duct,'' J. Acoust. Soc. Am. **108**, 624–631.
Keller, J. B., and Givoli, D. (**1989**). ''Exact non-reflecting boundary conditions,'' J. Comput. Phys. **82**, 172–192.
Munjal, M. L. (**1987**). *Acoustics of Ducts and Mufflers* (Wiley, New York).
Nelson, P. A., and Elliott, S. J. (**1992**). *Active Control of Sound* (Academic, New York).
Pierce, A. D. (**1989**). *Acoustics: An Introduction to its Physical Principles and Applications* (Acoustical Society of America, Woodbury, NY).
Pretlove, A. J. (**1965**). ''Free vibrations of a rectangular panel backed by a closed rectangular cavity,'' J. Sound Vib. **2**, 197–209.
Ramamoorthy, S., and Grosh, K. (**2000**). ''Broadband structural acoustic silencers,'' J. Acoust. Soc. Am. **108**, 2464.
Ramamoorthy, S., Grosh, K., and Dodson, J. M. (**2002**). ''Theoretical study of structural acoustic silencers for hydraulic systems,'' J. Acoust. Soc. Am. **111**, 2097–2108.
Ramamoorthy, S., Cook, W., Grosh, K., and Nawar, T. G. (**2001**). ''Broadband passive noise control in air/water using structural acoustic silencers,'' J. Acoust. Soc. Am. **110**, 2675.

Contrast source inversion methods in elastodynamics

George Pelekanos
Department of Mathematics and Statistics, Southern Illinois University, Edwardsville, Illinois 62026

Aria Abubakar
Schlumberger-Doll Research, 36 Old Quarry Road, Ridgefield, Connecticut 06877

Peter M. van den Berg
Center for Technical Geoscience, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, The Netherlands

(Received 5 May 2001; accepted for publication 25 August 2003)

In this paper a nonlinear inversion method is presented for determining the mass density of an elastic inclusion from the knowledge of how the inclusion scatters known incident elastic waves. The algorithm employed is an extension of the multiplicative regularized contrast source inversion method (MR-CSI) to elasticity. This method involves alternate determination of the mass density contrast and the contrast sources (the product of the contrast and the fields) in each iterative step. The simple updating schemes of the method allow the introduction of an extra regularization term to the cost functional as a multiplicative constraint. This so-called MR-CSI method (MR-CSI) has been proven to be very effective for the acoustic and electromagnetic inverse scattering problems. Numerical examples demonstrate that the MR-CSI method shows excellent edge preserving properties by robustly handling noisy data very well, even for more complicated elastodynamic problems. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1618751]

PACS numbers: 43.60.Pt, 43.10.Nq, 43.10.Sv [VWS] Pages: 2825–2834

I. INTRODUCTION

As described in Pao,[1] the essential properties of a medium are defined by the stiffness coefficients and the mass density. Direct or forward problems are concerned with the determination of the field when the source and medium properties are given. Inverse problems deal with the determination of the material properties when the sources are known and some data measuring the responses are given. Not much is known about solutions of inverse elasticity problems. Interested readers may find an introduction to the subject in the monograph by Claerbout.[2] Aymé-Bellegarda and Habashy[3] introduced a generalized noniterative inverse solution for 2D and 3D scatterers in a multilayered solid elastic medium; they used an exact dyadic contrast-source observation equation simplified in the Born framework, variations in the two Lamé constants and the density being arbitrary, while numerical results are given for a 2D variation in density.

We need not emphasize the importance of inverse problems since many of the interesting applications of the theory of elastic waves are inverse problems. They include the well-known problem of locating an earthquake source from seismograms, determining the nature of an earthquake source, finding the oil-trapping porous layer in the earth, and determining the ocean bottom from reflection data. Motivated by the importance of these problems, Pelekanos *et al.*[4] used an iterative technique to reconstruct the spatially dependent density of a two-dimensional object from measurements of the elastic field scattered when the object is illuminated by known sources. The inverse problem is formulated as an optimization problem, in which the cost functional is the sum of two terms: one is the defect in matching measured field data with the field scattered by a body of particular real density (error in data equation) and the second is the error in satisfying the equations of state, a system of integral equations for the field due to each excitation (error in the object equation). The density and the fields are updated by a linear iterative method in which the updating directions are weighted by parameters which are determined by minimizing the cost functional.

The method we implement here is an extension of the ideas presented in van den Berg and Kleinman,[5] and then extended by van den Berg *et al.*[6] and van den Berg and Abubakar.[7] This so-called contrast source inversion (CSI) method reconstructs the mass density contrast and the contrast sources alternatingly in each iterative step by minimizing the cost functional consisting of errors in the data and object equations. The method is very attractive, since it does not need to solve any full forward problem in each iteration. Later on, the CSI method has been modified to include a multiplicative regularization factor by van den Berg *et al.*[6] and extended by Abubakar *et al.*[8] to the full-vectorial electromagnetic 3D problem. This multiplicative technique allows the method to use a regularization factor without the necessity of determining an artificial weighting parameter. This regularization parameter is determined by the iterative process itself, which makes the method very suitable to invert experimental data as shown by Bloemenkamp *et al.*[9] In the present paper the improved version of the method, the so-called multiplicative regularized CSI (MR-CSI) method, described in van den Berg and Abubakar,[7] is applied to the even more challenging elasticity inverse scattering problem. Numerical examples demonstrate that the MR-CSI method shows excellent edge-preserving properties by robustly handling noisy as well as limited data very well.

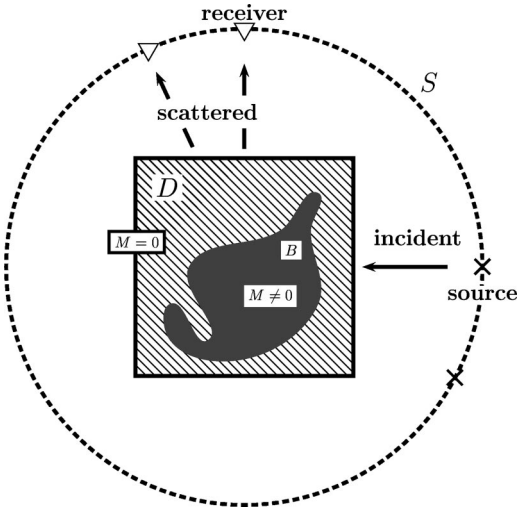


FIG. 1. The geometry of the scattering experiments.

II. NOTATION AND PROBLEM STATEMENT

Consider a bounded, simply connected domain D located in an unbounded homogeneous background medium with constant Lamé coefficients λ and μ. Let a scattering object (or objects) B, whose location and contrast are unknown, be embedded in D and assume that B is made by an inhomogeneous elastic material, also characterized by the constant Lamé coefficients λ and μ. The vector $\boldsymbol{x}$ denotes the vectorial position in $\mathbb{R}^2$. We irradiate the object by a number of known incident fields of one single frequency, $\boldsymbol{u}^{\mathrm{inc},(j)}(\boldsymbol{x}) = \boldsymbol{u}^{\mathrm{inc}}(\boldsymbol{x},\boldsymbol{x}_j')$, $j=1,...,J$, and source points $\boldsymbol{x}_j'$. The sources are located in a domain (or on a curve) S outside of and surrounding D, where the scattered field is measured as well. This scattering configuration is shown in Fig. 1. It has been shown (see Kupradze[10]) that the total field in D satisfies the following domain integral equation:

$$\boldsymbol{u}^{(j)}(\boldsymbol{x}) = \boldsymbol{u}^{\mathrm{inc},(j)}(\boldsymbol{x}) + (k^S)^2\mu \int_D \boldsymbol{\Gamma}(\boldsymbol{x},\boldsymbol{x}')\chi(\boldsymbol{x}') \times \boldsymbol{u}^{(j)}(\boldsymbol{x}')\mathrm{d}v(\boldsymbol{x}'), \quad \boldsymbol{x}\in D, \tag{1}$$

where

$$\chi(\boldsymbol{x}') = \frac{\rho(\boldsymbol{x}')-\rho_e}{\rho_e}, \tag{2}$$

is the normalized contrast between the mass density $\rho = \rho(\boldsymbol{x})$ of the scattering object and the mass density ρ_e of the homogeneous embedding. The Green's displacement tensor $\boldsymbol{\Gamma}(\boldsymbol{x},\boldsymbol{x}')$ is given by Morse and Feshbach[11] as

$$\boldsymbol{\Gamma}(\boldsymbol{x},\boldsymbol{x}') = \frac{i}{4}\frac{1}{\mu}\boldsymbol{I}H_0^{(1)}(k^S R) - \frac{i}{4}\frac{1}{\omega^2\rho_e}\boldsymbol{\nabla}\boldsymbol{\nabla}[H_0^{(1)}(k^P R) - H_0^{(1)}(k^S R)], \tag{3}$$

where $R=|\boldsymbol{x}-\boldsymbol{x}'|$, while k^P and k^S are the wave numbers for the P- and the S-wave, respectively and $\boldsymbol{I}$ is a unit matrix. Note that if $\boldsymbol{x}'$ is not in B, then the contrast function χ vanishes outside B.

If $\boldsymbol{u}^{(j)}$ solves the equation above, then the scattered field is obtained from the integral representation

$$\boldsymbol{u}^{\mathrm{sct},(j)}(\boldsymbol{x}) = (k^S)^2\mu \int_D \boldsymbol{\Gamma}(\boldsymbol{x},\boldsymbol{x}')\chi(\boldsymbol{x}')\boldsymbol{u}^{(j)}(\boldsymbol{x}')\mathrm{d}v(\boldsymbol{x}'), \quad \boldsymbol{x}\in S. \tag{4}$$

For convenience, in the solution of the inverse problem the P (dilatation) and SV (rotation) components of the two-dimensional scattered field are computed (Eringen and Suhubi[12]). The scattered P component is given by

$$u^{P,(j)}(\boldsymbol{x}) = \partial_1 u_1^{\mathrm{sct},(j)}(\boldsymbol{x}) + \partial_2 u_2^{\mathrm{sct},(j)}(\boldsymbol{x}), \quad \boldsymbol{x}\in S, \quad j=1,\ldots,J. \tag{5}$$

Consequently, the scattered SV component is given by

$$u^{SV,(j)}(\boldsymbol{x}) = \partial_1 u_2^{\mathrm{sct},(j)}(\boldsymbol{x}) - \partial_2 u_1^{\mathrm{sct},(j)}(\boldsymbol{x}), \quad \boldsymbol{x}\in S, \quad j=1,\ldots,J, \tag{6}$$

where ∂_1 and ∂_2 denote the spatial differentiations.

It is easily observed that in the equations above the contrast and fields occur as a product, hence, as in van den Berg and Kleinman,[5] we introduce the quantity

$$\boldsymbol{w}^{(j)} = \chi\boldsymbol{u}^{(j)}, \tag{7}$$

which is called the contrast source. Multiplying both sides of (1) by χ, and using (7), we obtain the object equations in terms of the contrast sources as follows:

$$\boldsymbol{w}^{(j)} - \chi\boldsymbol{Z}_D\boldsymbol{w}^{(j)} = \chi\boldsymbol{u}^{\mathrm{inc},(j)}, \quad \boldsymbol{x}\in D, \quad j=1,\ldots,J, \tag{8}$$

where

$$\boldsymbol{Z}_D\boldsymbol{w}^{(j)}(\boldsymbol{x}) = (k^S)^2\mu \int_D \boldsymbol{\Gamma}(\boldsymbol{x},\boldsymbol{x}')\boldsymbol{w}^{(j)}(\boldsymbol{x}')\mathrm{d}v(\boldsymbol{x}'), \quad \boldsymbol{x}\in D. \tag{9}$$

Using (7), the data equations become

$$\boldsymbol{Z}_S^P\cdot\boldsymbol{w}^{(j)} = u^{P,(j)}, \quad \text{on } S, \tag{10}$$

and

$$\boldsymbol{Z}_S^{SV}\cdot\boldsymbol{w}^{(j)} = u^{SV,(j)}, \quad \text{on } S, \tag{11}$$

where $\boldsymbol{Z}_S^P\cdot\boldsymbol{w}^{(j)}$ and $\boldsymbol{Z}_S^{SV}\cdot\boldsymbol{w}^{(j)}$ are set equal to the right-hand side of Eqs. (5) and (6), respectively.

III. CONTRAST SOURCE INVERSION METHOD

The application of the contrast source inversion in the inverse elasticity problem is based on the construction of sequences of sources $\boldsymbol{w}_n^{(j)}$ and contrast χ_n in order for a cost functional to be minimized. In our specific problem there are three error measurements involved; the first is the defect in matching the measured P data in $L^2(S)$, the second is the defect in matching the measured SV data also in $L^2(S)$, and the third is the error in the object equation in $L^2(D)$. We combine these error measures into a normalized cost functional

$$F_n(\boldsymbol{w}^{(j)},\chi) = F_S(\boldsymbol{w}^{(j)}) + F_{D,n}(\boldsymbol{w}^{(j)},\chi), \tag{12}$$

where

$$F_S(\boldsymbol{w}^{(j)}) = w_S \sum_{j=1}^{J} (\|u^{P,(j)} - \boldsymbol{Z}_S^P \cdot \boldsymbol{w}^{(j)}\|_S^2 + \|u^{SV,(j)} - \boldsymbol{Z}_S^{SV} \cdot \boldsymbol{w}^{(j)}\|_S^2), \tag{13}$$

and

$$F_{D,n}(\boldsymbol{w}^{(j)}, \chi) = w_{D,n-1} \sum_{j=1}^{J} \|\chi \boldsymbol{u}^{\mathrm{inc},(j)} - \boldsymbol{w}^{(j)} + \chi \boldsymbol{Z}_D \boldsymbol{w}^{(j)}\|_D^2. \tag{14}$$

The normalization factors in (13) and (14) are chosen as

$$w_S = \left[\sum_{j=1}^{J} (\|u^{P,(j)}\|_S^2 + \|u^{SV,(j)}\|_S^2) \right]^{-1}$$

and (15)

$$w_{D,n-1} = \left[\sum_{j=1}^{J} \|\chi_{n-1} \boldsymbol{u}^{\mathrm{inc},(j)}\|_D^2 \right]^{-1}.$$

The subscripts S and D are included in the L^2 norm to indicate the domain of integration. This is a quadratic functional in $\boldsymbol{w}^{(j)}$, but highly nonlinear in χ. Note that the term $\chi \boldsymbol{Z}_D \boldsymbol{w}^{(j)}$ is responsible for the nonlinearity of the inverse problem. We propose an iterative minimization of this cost functional using an alternating method which first updates $\boldsymbol{w}^{(j)}$ and then updates χ.

A. Updating the contrast sources

The method begins with the updating of the contrast sources $\boldsymbol{w}^{(j)}$ in the following manner:

Define the data errors to be

$$q_n^{P,(j)} = u^{P,(j)} - \boldsymbol{Z}_S^P \cdot \boldsymbol{w}_n^{(j)}, \quad n = 1,2,\ldots, \tag{16}$$

and

$$q_n^{SV,(j)} = u^{SV,(j)} - \boldsymbol{Z}_S^{SV} \cdot \boldsymbol{w}_n^{(j)}, \quad n = 1,2,\ldots. \tag{17}$$

The object error is

$$\boldsymbol{r}_n^{(j)} = \chi_n \boldsymbol{u}_n^{(j)} - \boldsymbol{w}_n^{(j)}, \quad n = 1,2,\ldots, \tag{18}$$

where

$$\boldsymbol{u}_n^{(j)} = \boldsymbol{u}^{\mathrm{inc},(j)} + \boldsymbol{Z}_D \boldsymbol{w}_n^{(j)}. \tag{19}$$

We now assume that $\boldsymbol{w}_{n-1}^{(j)}$ and χ_{n-1} are known and we update $\boldsymbol{w}^{(j)}$ by

$$\boldsymbol{w}_n^{(j)} = \boldsymbol{w}_{n-1}^{(j)} + \alpha_n \boldsymbol{v}_n^{(j)}, \tag{20}$$

where α_n is a constant and the update directions $\boldsymbol{v}_n^{(j)}$ are chosen as the Polak–Ribière conjugate gradient directions, namely

$$\boldsymbol{v}_0^{(j)} = \boldsymbol{0}, \quad \boldsymbol{v}_n^{(j)} = \boldsymbol{g}_n^{(j)} + \gamma_n \boldsymbol{v}_{n-1}^{(j)} \quad n \geqslant 1, \tag{21}$$

with

$$\gamma_n = \frac{\Sigma_j \langle \boldsymbol{g}_n^{(j)}, \boldsymbol{g}_n^{(j)} - \boldsymbol{g}_{n-1}^{(j)} \rangle_D}{\Sigma_j \|\boldsymbol{g}_{n-1,\boldsymbol{v}}^{(j)}\|_D^2}, \tag{22}$$

where $\boldsymbol{g}_{n,\boldsymbol{v}}^{(j)}$ is the gradient of the cost functional with respect to $\boldsymbol{w}^{(j)}$ evaluated at $\boldsymbol{w}_{n-1}^{(j)}$, χ_{n-1} while $\langle .,. \rangle_D$ denotes the inner product on $L^2(D)$. The expression for the gradient is found to be

$$\boldsymbol{g}_{n,\boldsymbol{v}}^{(j)} = w_S(\boldsymbol{Z}_S^{P^\star} q_{n-1}^{P,(j)} + \boldsymbol{Z}_S^{SV^\star} q_{n-1}^{SV,(j)}) + w_{D,n-1}(\boldsymbol{r}_{n-1}^{(j)} - \boldsymbol{Z}_D^\star(\overline{\chi}_{n-1} \boldsymbol{r}_{n-1}^{(j)})), \tag{23}$$

with $\boldsymbol{Z}_D^\star$ and $\boldsymbol{Z}_S^{P^\star}$, $\boldsymbol{Z}_S^{SV^\star}$ being the adjoint operators mapping from $L^2(D)$ and $L^2(S)$, respectively, into $L^2(D)$, and the overbar denotes complex conjugate.

Once the update directions are chosen, the constant parameter α_n is determined to minimize the cost functional

$$\begin{aligned} F_n(\boldsymbol{w}^{(j)}, \chi_{n-1}) &= w_S \sum_{j=1}^{J} (\|u^{P,(j)} - \boldsymbol{Z}_S^P \cdot \boldsymbol{w}_n^{(j)}\|_S^2 + \|u^{SV,(j)} - \boldsymbol{Z}_S^{SV} \cdot \boldsymbol{w}_n^{(j)}\|_S^2) \\ &\quad + w_{D,n-1} \sum_{j=1}^{J} \|\chi_{n-1} \boldsymbol{u}_n^{(j)} - \boldsymbol{w}_n^{(j)}\|_D^2 \\ &= w_S \sum_{j=1}^{J} (\|q_{n-1}^{P,(j)} - \alpha_n \boldsymbol{Z}_S^P \cdot \boldsymbol{v}_n^{(j)}\|_S^2 \\ &\quad + \|q_{n-1}^{SV,(j)} - \alpha_n \boldsymbol{Z}_S^{SV} \cdot \boldsymbol{v}_n^{(j)}\|_S^2) \\ &\quad + w_{D,n-1} \sum_{j=1}^{J} \|\boldsymbol{r}_{n-1}^{(j)} - \alpha_n (\boldsymbol{v}_n^{(j)} - \chi_{n-1} \boldsymbol{Z}_D \boldsymbol{v}_n^{(j)})\|_D^2, \end{aligned} \tag{24}$$

and is found to be

$$\alpha_n = \frac{\Sigma_j f_{n,j}}{\Sigma_j h_{n,j}}, \tag{25}$$

where

$$f_{n,j} = w_S(\langle q_{n-1}^{P,(j)}, \boldsymbol{Z}_S^P \cdot \boldsymbol{v}_n^{(j)} \rangle_S + \langle q_{n-1}^{SV,(j)}, \boldsymbol{Z}_S^{SV} \cdot \boldsymbol{v}_n^{(j)} \rangle_S) + w_{D,n-1} \langle \boldsymbol{r}_{n-1}^{(j)}, \boldsymbol{v}_n^{(j)} - \chi_{n-1} \boldsymbol{Z}_D \boldsymbol{v}_n^{(j)} \rangle_D, \tag{26}$$

and

$$h_{n,j} = w_S(\|\boldsymbol{Z}_S^P \cdot \boldsymbol{v}_n^{(j)}\|_S^2 + \|\boldsymbol{Z}_S^{SV} \cdot \boldsymbol{v}_n^{(j)}\|_S^2) + w_{D,n-1} \|\boldsymbol{v}_n^{(j)} - \chi_{n-1} \boldsymbol{Z}_D \boldsymbol{v}_n^{(j)}\|_D^2 \tag{27}$$

and where $\langle .,. \rangle_S$ denotes the inner product on $L^2(S)$.

After the contrast sources $\boldsymbol{w}_n^{(j)}$ are obtained, the total fields $\boldsymbol{u}_n^{(j)}$ can be found via (19)

$$\boldsymbol{u}_n^{(j)} = \boldsymbol{u}_{n-1}^{(j)} + \alpha_n \boldsymbol{Z}_D \boldsymbol{v}_n^{(j)}. \tag{28}$$

B. Updating the contrast

If the contrast sources $\boldsymbol{w}_n^{(j)}$ and the fields $\boldsymbol{u}_n^{(j)}$ are known, the contrast χ_n is now obtained by minimizing the second term in (12), i.e.,

$$F_{D,n}(\boldsymbol{w}_n^{(j)}, \chi_n) = w_{D,n-1} \sum_{j=1}^{J} \|\chi_n \boldsymbol{u}_n^{(j)} - \boldsymbol{w}_n^{(j)}\|_D^2. \tag{29}$$

The normalized error $F_{D,n}(\boldsymbol{w}_n^{(j)}, \chi_n)$ is minimized by choosing[5]

$$\chi_n = \frac{\Sigma_j \operatorname{Re}(\boldsymbol{w}_n^{(j)} \cdot \overline{\boldsymbol{u}_n^{(j)}})}{\Sigma_j |\boldsymbol{u}_n^{(j)}|^2}. \tag{30}$$

Note that this result is identical to the result obtained by updating the contrast as

$$\chi_n = \chi_{n-1} + \alpha_n^{\chi} d_n, \tag{31}$$

where $\alpha_n^{\chi} = w_{D,n-1}^{-1}$ and d_n is the preconditioned gradient of $F_{D,n}(\boldsymbol{w}_n^{(j)}, \chi_n)$, i.e.,

$$d_n = \frac{-w_{D,n-1} \Sigma_j \operatorname{Re}[(\chi_{n-1}\boldsymbol{u}_n^{(j)} - \boldsymbol{w}_n^{(j)}) \cdot \overline{\boldsymbol{u}_n^{(j)}}]}{\Sigma_j |\boldsymbol{u}_n^{(j)}|^2}. \tag{32}$$

In case we have *a priori* information that the mass density is a positive quantity, we remark that this positivity constraint is easily implemented by enforcing a negative value to zero after each update of the contrast.

C. Choice of the initial estimate

We now need to indicate starting values for $\boldsymbol{w}_0^{(j)}$. Clearly zero is not a good choice, since then the cost functional (12) will be undefined for $n=1$. To this end we choose as starting values the contrast sources that minimize the data error, that is the contrast sources obtained by backpropagation

$$\boldsymbol{w}_0^{(j)} = \alpha_0 (\boldsymbol{Z}_S^{P\star} u^{P,(j)} + \boldsymbol{Z}_S^{SV\star} u^{SV,(j)}), \tag{33}$$

where

$$\alpha_0 = \frac{\Sigma_j \| (\boldsymbol{Z}_S^{P\star} u^{P,(j)} + \boldsymbol{Z}_S^{SV\star} u^{SV,(j)}) \|_S^2}{\Sigma_j \| \boldsymbol{Z}_S^{P} \cdot (\boldsymbol{Z}_S^{P\star} u^{P,(j)} + \boldsymbol{Z}_S^{SV\star} u^{SV,(j)}) \|_S^2 + \| \boldsymbol{Z}_S^{SV} \cdot (\boldsymbol{Z}_S^{P\star} u^{P,(j)} + \boldsymbol{Z}_S^{SV\star} u^{SV,(j)}) \|_S^2}. \tag{34}$$

As far as the starting value of χ_0 is concerned, we obtain

$$\chi_0 = \frac{\Sigma_j \operatorname{Re}(\boldsymbol{w}_0^{(j)} \cdot \overline{\boldsymbol{u}_0^{(j)}})}{\Sigma_j |\boldsymbol{u}_0^{(j)}|_D^2}, \tag{35}$$

where

$$\boldsymbol{u}_0^{(j)} = \boldsymbol{u}^{\mathrm{inc},(j)} + \boldsymbol{Z}_D \boldsymbol{w}_0^{(j)}, \tag{36}$$

and $\boldsymbol{w}_0^{(j)}$ is given in (33).

This completes the description of the nonregularized version of the contrast source inversion (CSI) algorithm, where in each iteration an update of the contrast sources is followed by an update of the contrast. Note that this alternating scheme allows us to introduce a multiplicative constraint in a simple fashion.

D. Inclusion of a multiplicative constraint

Recent work with image enhancement has shown that minimization of the total variation of the image can significantly improve the quality of the reconstruction, see, e.g., Acar and Vogel,[13] Blomgren *et al.*,[14] Dobson and Santosa.[15,16] Van den Berg and Kleinman[17] incorporated the total variation (TV) in an inverse scattering problem by enhancing the modified gradient algorithm. In the latter approach a total variation term was added to the cost functional, resulting in a substantial improvement of the performance of the reconstruction method, both for ''blocky'' and smooth contrast configurations. The addition of the total variation to the cost functional has a very positive effect on the quality of the reconstructions for both blocky and smooth profiles, but a drawback is the presence of an artificial weighting parameter in the cost functional, which can only be determined through considerable numerical experimentation; see Hansen,[18] and *a priori* information of the desired reconstruction.

van den Berg *et al.*[6] have suggested including the total variation as a multiplicative constraint, and hence the original cost functional is the weighting parameter, i.e., determined by the inversion problem itself. This eliminates the choice of the artificial regularization parameters completely. In each iteration, a multiplicative cost functional is introduced as

$$\mathcal{F}_n(\boldsymbol{w}^{(j)}, \chi) = [F_S(\boldsymbol{w}^{(j)}) + F_{D,n}(\boldsymbol{w}^{(j)}, \chi)] F_n^{\mathrm{TV}}(\chi), \tag{37}$$

where the first factor is the original cost functional (12) of the CSI method, and where the second factor is the so-called TV factor. Recently, van den Berg and Abubakar[7] have considered the TV factor as a weighted norm on $L^2(D)$, in which the weighting favors flat parts and nonflat parts of the contrast profile almost equally. This weighted TV factor has been shown to be more effective than the one given in van den Berg *et al.*[6] Thus, in this paper we choose to use a TV factor similar to the one used in van den Berg and Abubakar.[7] This weighted total-variation factor is given by

$$F_n^{\mathrm{TV}}(\chi) = \frac{1}{V} \int_D \frac{|\nabla \chi(\boldsymbol{x}')|^2 + \delta_{n-1}^2}{|\nabla \chi_{n-1}(\boldsymbol{x}')|^2 + \delta_{n-1}^2} \, \mathrm{d}v(\boldsymbol{x}'), \tag{38}$$

where $V = \int_D \mathrm{d}v(\boldsymbol{x}')$ denotes the area (two-dimensional volume) of the test domain D.

The quantity δ^2 in (38) is introduced for restoring differentiability to the total variation and is chosen as

$$\delta_{n-1}^2 = F_{D,n-1} \tilde{\Delta}^2, \tag{39}$$

where $\tilde{\Delta}$ denotes the reciprocal mesh size of the discretized domain D, and $F_{D,n-1}$ is the normalized norm of the object error of the previous iteration, cf. (29). We have chosen δ^2 to be large in the beginning of the optimization and small towards the end; hence, optimization will reconstruct the contrast in the first iterations in a normal way before it will apply the minimization of variation to shape the image further.

The choice of the multiplicative cost functional (37) is based on two things: the objective of minimizing the error in the data and object equations and the observation that the weighted TV factor, when minimized, converges to 1. The structure of this cost functional is such that it will minimize the total variation with a large weighting parameter in the beginning of the optimization process, because the value of cost functional $F_n(\boldsymbol{w}^{(j)},\chi)$ is still large, and that it will gradually minimize more and more the error in the data and object equations when the weighted TV factor has reached a nearly constant value close to 1. If noise is present in the data, the data error term will remain at a large value during the optimization; therefore, the weight of the weighted TV factor will be more significant. Hence, the noise will at all times be suppressed in the reconstruction process and we automatically fulfill the need of a larger regularization factor when the data contains noise as suggested by Chan and Wong[19] and Rudin *et al.*[20]

By introducing this cost functional $\mathcal{F}_n$, the TV factor does not change the updating of the contrast sources $\boldsymbol{w}_n^{(j)}$ and the fields $\boldsymbol{u}_n^{(j)}$, because $F_n^{\mathrm{TV}}(\chi_{n-1})=1$ at the beginning of each iteration. The updating scheme for χ_n is given by

$$\chi_n=\chi_{n-1}+\alpha_n^{\chi} d_n, \tag{40}$$

where the update directions d_n are taken as Polak–Ribière conjugate gradient directions of the cost functional (37), *viz.*,

$$d_0=0, \quad d_n=g_n^{\chi}+\frac{\mathrm{Re}\langle g_n^{\chi}, g_n^{\chi}-g_{n-1}^{\chi}\rangle_D}{\langle g_{n-1}^{\chi}, g_{n-1}^{\chi}\rangle_D} d_{n-1}, \quad n\geqslant 1, \tag{41}$$

while the preconditioned gradient is determined as, cf. (32)

$$g_n^{\chi}=\frac{-w_{D,n-1}\Sigma_j \mathrm{Re}[(\chi_{n-1}\boldsymbol{u}_n^{(j)}-\boldsymbol{w}_n^{(j)})\cdot\overline{\boldsymbol{u}_n^{(j)}}]+F_n(\boldsymbol{w}^{(j)},\chi_{n-1})g_n^{\mathrm{TV}}}{\Sigma_j|\boldsymbol{u}_n^{(j)}|^2}, \tag{42}$$

where

$$g_n^{\mathrm{TV}}=\frac{1}{V}\nabla\cdot\left[\frac{\nabla\chi_{n-1}}{|\nabla\chi_{n-1}|^2+\delta_{n-1}^2}\right]. \tag{43}$$

Note that the gradient tends to the direction d_n of (32) of the original CSI method as the gradient, g_n^{TV}, tends to zero. The weighting of the gradients clearly depends on the errors in the cost functional F_n and the weighted TV factor F_n^{TV}. Similar to an additive regularization, the present multiplicative regularization decreases the chance that the gradient has a zero direction, which reduces the possibility to arrive at in a local minimum. Comparing this gradient with the one of the nonweighted TV factor as in van den Berg *et al.*,[6] we immediately observe that apart from a constant factor, these gradients are identical. Hence, this weighted TV factor combines the features of minimization of the TV in the L^2 norm and in the L^1 norm through its gradient.

With the Polak–Ribière update directions completely specified, the real-valued constant α_n^{χ} in (40) is found as

$$\alpha_n^{\chi}=\arg\min_{\text{real }\alpha}\{[\boldsymbol{F}_S(\boldsymbol{w}_n^{(j)})+F_{D,n}(\boldsymbol{w}_n^{(j)},\chi_{n-1}+\alpha\, d_n)] \times F_n^{\mathrm{TV}}(\chi_{n-1}+\alpha\, d_n)\}. \tag{44}$$

The minimization of the multiplicative cost functional (44) can be performed analytically. The cost functional is a fourth-degree polynomial in α^{χ}, *viz.*,

$$\begin{aligned}\mathcal{F}=\Big[&F_S(\boldsymbol{w}_n^{(j)})+F_{D,n}(\boldsymbol{w}_n^{(j)},\chi_{n-1})\\&+2\alpha w_{D,n}\,\mathrm{Re}\sum_j\langle d_n\boldsymbol{u}_n^{(j)},\chi_{n-1}\boldsymbol{u}_n^{(j)}-\boldsymbol{w}_n^{(j)}\rangle_D\\&+\alpha^2 w_{D,n}\sum_j\|d_n\boldsymbol{u}_n^{(j)}\|_D^2\Big]\\&\times[1+2\alpha\,\mathrm{Re}\langle b_{n-1}\nabla\chi_{n-1},b_{n-1}\nabla d_n\rangle_D\\&+\alpha^2\|b_{n-1}\nabla d_n\|_D^2],\end{aligned} \tag{45}$$

where

$$b_{n-1}=V^{-1/2}(|\nabla\chi_{n-1}|^2+\delta_{n-1}^2)^{-1/2}. \tag{46}$$

Differentiation with respect to α yields a cubic equation with one real root and two complex conjugate roots. The real root is the desired minimizer α_n^{χ}. In our numerical examples we use this weighted TV factor as the multiplicative regularization of the CSI method, and we name this method the MR-CSI method.

IV. NUMERICAL RESULTS

In this section some numerical examples of inversion using the MR-CSI method will be presented. The discretized versions of the various operators used in the algorithm can be found in Pelekanos *et al.*[21] Further, in order to be able to compare the reconstruction results using the data excited either by *P*- or *SV*-waves, we introduce the quantity error in contrast as follows:

$$\mathrm{ERR}_n=\frac{\|\rho(\boldsymbol{x})-\rho^{\mathrm{exact}}(\boldsymbol{x})\|_D^2}{\|\rho^{\mathrm{exact}}(\boldsymbol{x})\|_D^2}. \tag{47}$$

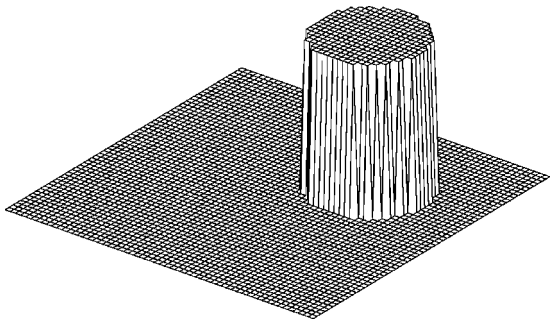


FIG. 2. The approximate original profile of the circular cylinder.

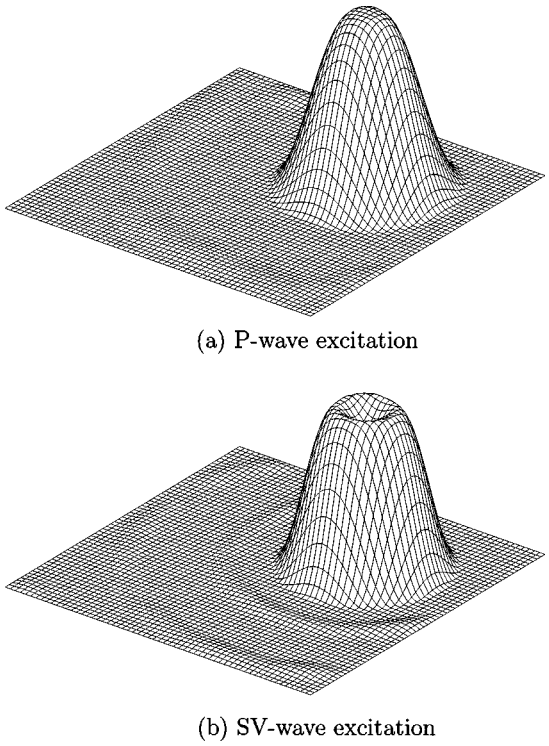
In all of our numerical examples, after generation of synthetic data by either the Bessel series solution (see White[22]) or the conjugate gradient FFT method (see Pelekanos *et al.*[21]), 10% additive random white noise is added. Moreover, the simple inclusion of *a priori* positivity constraint is used. We excite our object by either *P*- or *SV*-waves. The object will scatter both *P*- and *SV* waves. Subsequently, we reconstruct the object by measuring the contributions from both kinds of scattered waves.

A. Reconstruction of a circular cylinder

In our first numerical example the scatterer was taken to be a circular cylinder of radius $a=0.35$ m and density $\rho=2.0$, while the outer medium's density was $\rho_e=1$. Hence, the contrast is $\chi=1.0$. The circular cylinder is centered at $(-0.35,-0.35)$ m. The scatterer is located in the test domain D. This test domain was divided into 61×61 subsquares of 0.0344×0.0344 m^2. Note that this fine discretization grid is needed in order to adequately model the boundary of the circular cylinder. The wave numbers of the *P*- and *SV* waves are $k^P=3$ and $k^S=6$, respectively. This means that the side length of the test domain was equal to about one wavelength for the *P* waves and two wavelengths for the *SV* waves in the exterior medium.

The measurement surface S was chosen to be a circle with a radius of 2 m. Twenty-nine stations ($J=29$) were located uniformly on this circle, with each station serving successively as a line source and all stations acting as receivers. Throughout the description of the solution for both direct and inverse problems, it became apparent that the method used is essentially the same. There is a suspicion, however, that the use of ''measured'' data in the inverse problem, which are produced by the same numerical method (forward problem) can result in an ''inverse crime.'' Therefore in this first example we use the exact data obtained from the Bessel series solution; see White.[22] Furthermore, after generation of synthetic data we add 10% random additive white noise to the scattered fields for both *P*- and *SV* waves. The approximate model of the circular cylinder is given in Fig. 2.

For *P*-wave incidence excitation, the reconstruction after 512 iterations using the CSI method is shown in Fig. 3(a). Note that although the number of iterations is large, we do not solve any full-forward problem in each iteration. One iteration of the CSI method for this example takes only 10 s on a personal computer with a 600-MHz Pentium II processor. For *SV*-wave incidence, the reconstruction after 512 iterations using the CSI method is shown in Fig. 3(b). The top values of the reconstructed circular cylinder for *P*- and *SV*-wave incidence are approximately equal to $\chi=1.15$ and $\chi=1.05$, respectively. Further, we observe that the boundary of the circular cylinder is better reconstructed from data generated by *SV* incident waves.



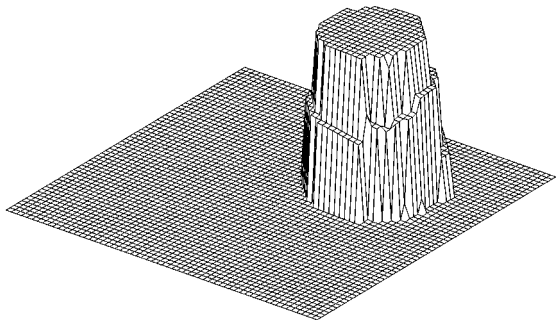
(a) P-wave excitation

(b) SV-wave excitation

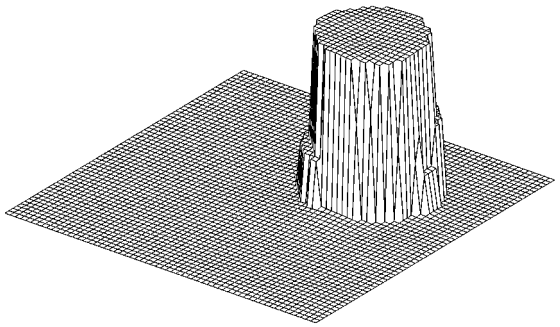
FIG. 3. The reconstruction results of the circular cylinder from excitation with *P* waves (a) and *SV* waves (b) using the CSI method. The data sets are generated using series of Bessel function solutions with 10% additive random white noise.

Next, we use the CSI method using the multiplicative regularization factor (the MR-CSI method). The reconstructions after 512 iterations are shown in Fig. 4(a) for *P*-wave incidence, while for *SV*-wave incidence in Fig. 4(b). The improvements are remarkable and the reconstruction time per iteration in the inversion algorithm is only increased by approximately 20%. Now, the top values of the reconstructed circular cylinder for both *P*- and *SV*-wave excitation are approximately equal to $\chi=1.05$. Nevertheless, the *SV*-wave incidence still gives a more accurate result.

In order to have some idea about the convergence behavior of the reconstruction results from both excitations, we present in Fig. 5 the error in contrast ERR_n, see (47), as a function of the number of iterations. The errors in contrast for *P*-wave incidence are given by the dotted line for the CSI method and the dashed line for the MR-CSI method. The errors in contrast for *SV* incident waves are given by the dashed-dotted line for the CSI method and the solid line for the MR-CSI method. Note that, even at the very beginning of the iterative process (after backpropagation), the errors in contrast of the reconstructed profile from *P*-wave excitations for both methods are always larger than the ones from *SV*-wave excitations. This is an indication that the bad re-



(a) P-wave excitation



(b) SV-wave excitation

FIG. 4. The reconstruction results of the circular cylinder from excitation with P waves (a) and SV waves (b) using the MR-CSI method. The data sets are generated using series of Bessel function solutions with 10% additive random white noise.

constructed boundary of the circular cylinder from P-wave incidence is due to the physical behavior of the incident waves; it is not due to the behavior of the inversion algorithm.

B. Reconstruction of three distinct square cylinders

In our second numerical example we consider the inversion of multiple scatterers. The scatterers were taken to be three square cylinders of diameter 3/4 m. The mass densities

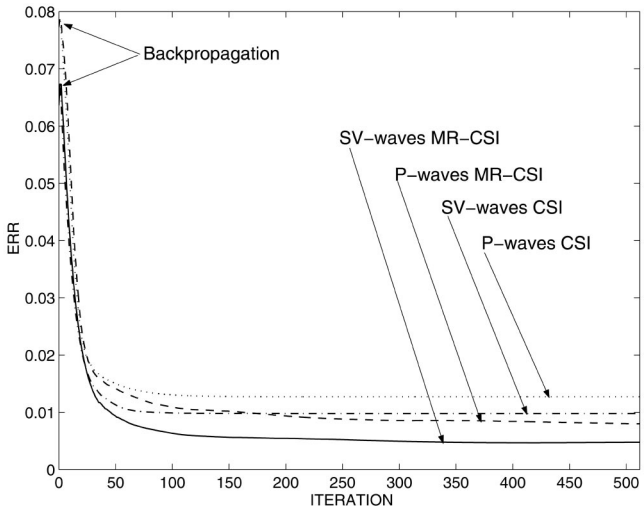


FIG. 5. The error in contrast ERR_n as the function of the number of iteration n for the reconstruction of the circular cylinder.

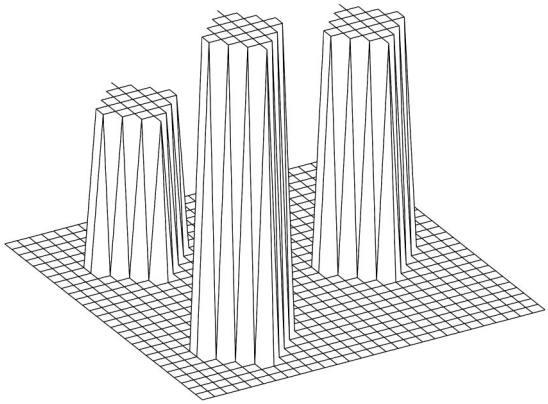


FIG. 6. The original profile of the three distinct square cylinders.

for the scatterers were $\rho=2.0$, $\rho=2.5$, and $\rho=3.0$, while the outer medium's density was $\rho_e=1.0$. Hence, the contrasts are $\chi=1.0$, $\chi=1.5$, and $\chi=2.0$. The scatterers are assumed to be located in a test domain D. This test domain was divided into 29×29 subsquares of 0.1034×0.1034 m^2. The wave numbers of the P- and SV-waves are $k^P=3$ and $k^S=6$, respectively. This means that the side length of the test domain was equal to about one and one half wavelengths for the P waves and three wavelengths for the SV waves in the background medium.

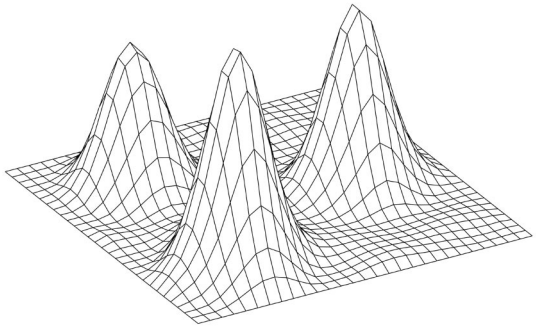
The measurement surface S was chosen to be a circle of radius 3 m. Twenty-nine stations ($J=29$) were located uniformly on this circle, with each station serving successively as a line source and all stations acting as receivers. The data are generated numerically using the CG-FFT method described in Pelekanos *et al.*[21] The actual profile which has been used to generate synthetic data is given in Fig. 6. After generation of synthetic data we added 10% random additive white noise to the scattered fields for both P- and SV-waves.

The reconstructed results after 512 iterations from P-wave incidence using the CSI method are presented in Fig. 7(a). One iteration of the CSI method for this example takes only 2.5 s on a personal computer with a 600-MHz Pentium II processor. However, poor reconstruction results are observed. The contrast value levels for the square cylinders are equal to $\chi=0.85$, $\chi=1.2$, and $\chi=1.45$. Inversion of the SV-wave data yields a minor improvement. The results are shown in Fig. 7(b). The contrast value levels for the square cylinders are approximately the same as the ones obtained from inverting the P-wave data.

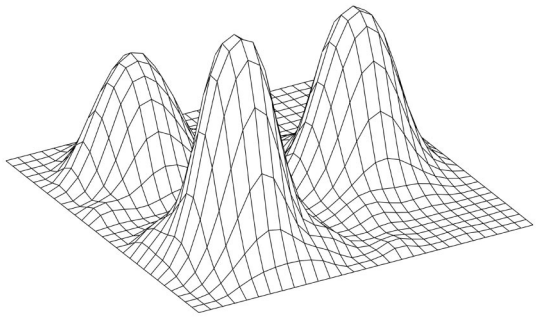
Again, we observe a remarkable improvement by using the MR-CSI method. The results for both excitations are given in Fig. 8. Now, the levels for the contrast values for the square cylinders for both excitation types are, $\chi=1.0$, $\chi=1.45$, and $\chi=1.95$. Note that also here reconstructions using incident SV waves are better compared to the ones obtained from P-wave excitations.

C. Reconstruction of concentric square cylinders

In our last example, an inhomogeneous scatterer is located in a test domain that is subdivided into 29×29 subsquares of 0.1034×0.1034 m^2. The wave numbers of the P- and SV waves are $k^P=3$, $k^S=6$, respectively. This means

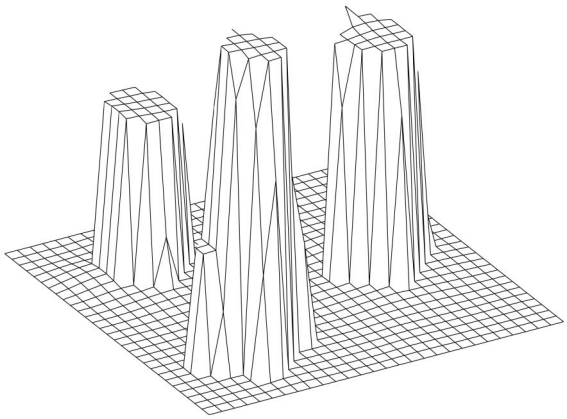


(a) P-wave excitation

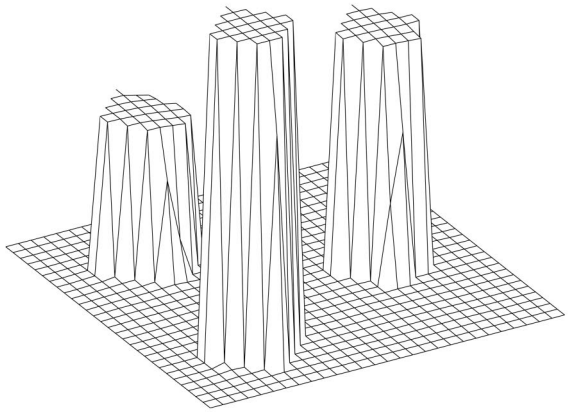


(b) SV-wave excitation

FIG. 7. The reconstruction results of the three distinct square cylinders from excitation with P waves (a) and SV waves (b) using the CSI method. The data sets are generated using CG-FFT method with 10% additive random white noise.



(a) P-wave excitation



(b) SV-wave excitation

FIG. 8. The reconstruction results of the three distinct square cylinders from excitation with P waves (a) and SV waves (b) using the MR-CSI method. The data sets are generated using CG-FFT method with 10% additive random white noise.

that the side length of the test square was equal to about one and one half wavelengths for the P waves and three wavelengths for the SV waves in the background medium. The actual profile of the scatterer consists of an inner square with mass density $\rho=1.9$, a middle square with $\rho=1.6$, and an outer square with $\rho=1.3$; see Fig. 9. The contrasts from the inner square to the outer are equal to $\chi=0.9$, $\chi=0.6$, and $\chi=0.3$. The side length of the inner square is equal to 0.75 m, the one of the middle square is equal to 1.5 m, and the one of the outer square is equal to 2 m.

The measurement setup is the same as in the previous examples. The data are generated numerically using the CG-FFT method, and after generation, 10% additive random white noise was added. The reconstruction results after 512 iterations for both P- and SV-wave incidence using the CSI method are shown in Figs. 10(a) and (b), while the ones using the MR-CSI method are shown in Figs. 11(a) and (b). Note that for the MR-CSI method the contrast values are very well reconstructed. Comparing now the results of both excitations, we observe again that inversion from SV-wave data yields better reconstructions compared to the ones obtained from P-wave excitations.

It is worthwhile to note that even though we only show reconstructions of objects with sharp edges, the MR-CSI method has also been shown to work properly with smooth contrast objects.[23]

V. CONCLUSIONS

We have developed a nonlinear inversion method for the two-dimensional problem of scattering of an elastic (P or SV) wave. This method, which is denoted as CSI, is based on the source-type integral equation which relates measured data to a source distribution in the scattering object. A cost functional is defined consisting of errors in the source-type equations and the state equations. The updates in the sources are found as a conjugate gradient step, after which the con-

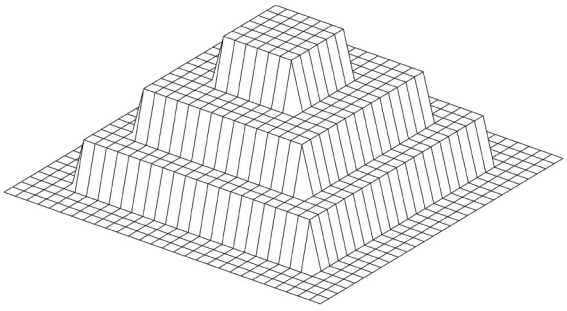
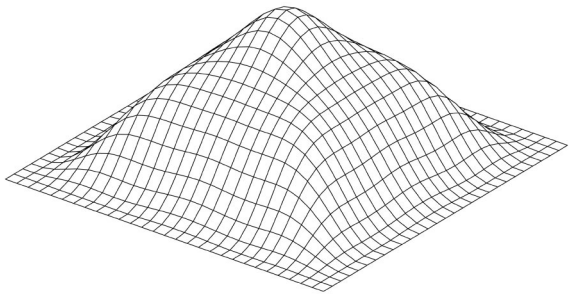
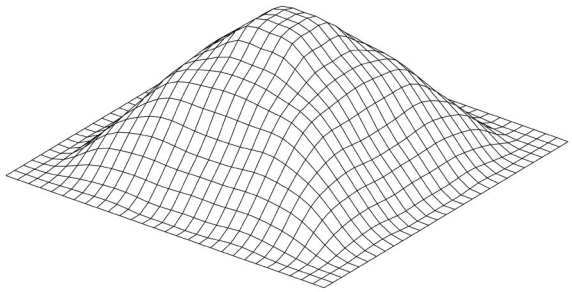


FIG. 9. The original profile of the concentric square cylinders.

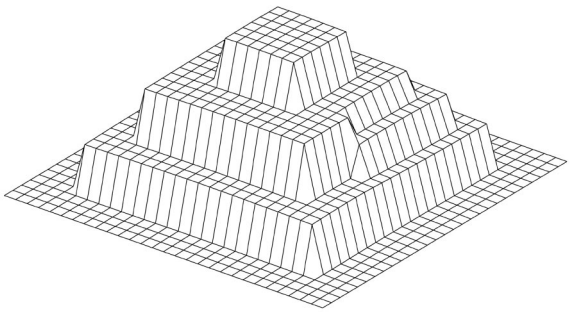


(a) P-wave excitation

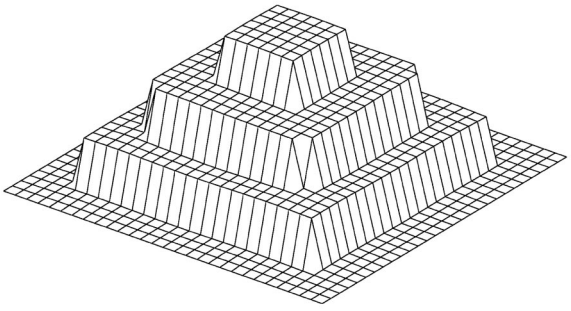


(b) SV-wave excitation

FIG. 10. The reconstruction results of the concentric square cylinders from excitation with *P* waves (a) and *SV* waves (b) using the CSI method. The data sets are generated using CG-FFT method with 10% additive random white noise.



(a) P-wave excitation



(b) SV-wave excitation

FIG. 11. The reconstruction results of the concentric square cylinders from excitation with *P* waves (a) and *SV* waves (b) using the MR-CSI method. The data sets are generated using CG-FFT method with 10% additive random white noise.

trast is updated by minimizing the error in the state equations in a very simple manner. The most interesting feature of the algorithm is that it does not solve any full forward problem at any stage of the iteration. This feature, as also expressed in Abubakar *et al.*,[8] gives hope that a three-dimensional inverse problem can be handled with a moderate computational power. In addition, we have discussed a new type of regularization that, together with the contrast source inversion method, is an extremely effective and robust with respect to noisy data algorithm. This new multiplicative regularized CSI method is denoted as the MR-CSI method.

Future work will concern further refinements to make the algorithm more applicable in real-world problems by removing the assumption that the same traction operator occurs in both media. This results in a very complicated hypersingular integral equation which also contains gradients of the displacement fields. This will give more degrees of freedom to the problem, and the object will be reconstructed from its Lamé parameters. Future work should also be directed toward extending the method to include measurements at more than one frequency in order for larger contrasts to be treated effectively.

[1] Y-H. Pao, ''Elastic waves in solids,'' J. Appl. Mech. **50**, 1152–1164 (1983).

[2] J.F. Claerbout, *Fundamentals of Geophysical Data Processing* (McGraw-Hill, New York, 1976).

[3] E.J. Aymè-Bellegarda and T.M. Habashy, ''Ultrasonic inverse scattering of multidimensional objects buried in multilayered elastic background structures,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 11–18 (1992).

[4] G. Pelekanos, R.E. Kleinman, and P.M. van den Berg, ''Inverse scattering in elasticity—A modified gradient approach,'' Wave Motion **32**, 57–65 (2000).

[5] P.M. van den Berg and R.E. Kleinman, ''A contrast source inversion method,'' Inverse Probl. **13**, 1607–1620 (1997).

[6] P.M. van den Berg, A.L. van Broekhoven, and A. Abubakar, ''Extended contrast source inversion,'' Inverse Probl. **15**, 1325–1344 (1999).

[7] P.M. van den Berg and A. Abubakar, ''Contrast source inversion method: State of art,'' Prog. Electromagn. Res. **34**, 189–218 (2001).

[8] A. Abubakar, P.M. van den Berg, and B.J. Kooij, ''A conjugate gradient contrast source technique for 3D profile inversion,'' *Special Issue on Problems of Random Scattering and Electromagnetic Wave Sensing*, IEICE Trans. Electron. **E83-C**, 1864–1874 (2000).

[9] R.F. Bloemenkamp, A. Abubakar, and P.M. van den Berg, ''Inversion of experimental multifrequency data using the Contrast Source Inversion method,'' *Special Issue on Testing Inversion Algorithms Against Experimental Data*, Inverse Probl. **17**, 1611–1622 (2001).

[10] V.D. Kupradze, *Progress in Solid Mechanics* (North-Holland, New York, 1963), Vol. III.

[11] P.M. Morse and H. Feshbach, *Methods of Mathematical Physics* (McGraw-Hill, New York, 1953), Part II.

[12] A.C. Eringen and E.S. Suhubi, *Elastodynamics* (Academic, New York, San Francisco, London, 1983), Vol. II.

[13] R. Acar and C.R. Vogel, ''Analysis of bounded variation penalty methods for ill-posed problems,'' Inverse Probl. **10**, 1217–1229 (1994).

[14] P. Blomgren, T.F. Chan, P. Mulet, and C.K. Wong, ''Total variation image restorations: Numerical methods and extensions,'' IEEE Proc. ICIP **97**, 384–387 (1997).

[15] D.C. Dobson and F. Santosa, ''An image-enhancement technique for electrical impedance tomography,'' Inverse Probl. **10**, 317–334 (1994).

[16] D.C. Dobson and F. Santosa, ''Recovery of blocky images for noisy and blurred data,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **56**, 1181–1198 (1996).

[17] P.M. van den Berg and R.E. Kleinman, ''A total variation enhanced modified gradient algorithm for profile reconstruction,'' Inverse Probl. **11**, L5–L10 (1995).

[18] P.C. Hansen, ''Analysis of discrete ill-posed problems by means of the L-curve,'' SIAM Rev. **34**, 561–580 (1992).
[19] T.F. Chan and C.K. Wong, ''Total variation blind deconvolution,'' IEEE Trans. Image Process. **7**, 370–375 (1998).
[20] L. Rudin, S. Osher, and C. Fatemi, ''Nonlinear total variation based noise removal algorithm,'' Physica D **60**, 259–268 (1992).
[21] G. Pelekanos, R.E. Kleinman, and P.M. van den Berg, ''A weak form of the conjugate gradient FFT method for two-dimensional elastodynamics,'' J. Comput. Phys. **160**, 597–611 (2000).
[22] R.M. White, ''Elastic wave scattering at a cylindrical discontinuity in a solid,'' J. Acoust. Soc. Am. **30**, 771–785 (1958).
[23] P.M. van den Berg and A. Abubakar, ''Inverse scattering algorithms based on contrast source integral representations,'' Inverse Probl. Eng. **10**, 559–576 (2002).

Multiple bursts, multiple looks, and stream coherence in the release from informational masking[a)]

Gerald Kidd, Jr.[b)] and Christine R. Mason
Hearing Research Center and Programs in Communication Disorders, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

Virginia M. Richards
Department of Psychology, 3815 Walnut Street, University of Pennsylvania, Philadelphia, Pennsylvania 19104

(Received 20 December 2002; revised 5 September 2003; accepted 8 September 2003)

In the simultaneous multitone masking paradigm introduced by Neff and Green [Percept. Psychophys. **41**, 409–415 (1987)] the masker typically is a small number of tones having frequencies and levels that are randomly drawn on every presentation. Large amounts of masking for a pure-tone signal often occur that are thought to reflect central, rather than peripheral, limitations on processing. Previous work from this laboratory has indicated that playing a rapid succession of randomly drawn multitone maskers in each observation interval dramatically reduces the amount of masking that is observed relative to a single burst (SB). In this multiple-bursts-different (MBD) procedure, the signal tone is the only constant frequency component during the sequence of bursts and tends to perceptually segregate from the masker. In this study, the number of masker bursts and the interburst interval (IBI) were varied. The goals were to determine how the release from masking relative to the SB condition depends on the number of bursts and to examine whether increasing the IBI would cause each burst to be processed independently. If the latter were true, it might disrupt the perception of signal stream coherence, thereby diminishing the MBD advantage. However, multiple independent looks could also lead to an improvement in performance. For those subjects showing large amounts of informational masking in the SB condition, substantial reduction in masked thresholds occurred as the number of masker bursts increased, while masking increased as IBI lengthened. The results were not consistent with a simple version of a multiple-look model in which the information from each burst was combined optimally, but instead appear to be attributable to mechanisms involved in the perceptual organization of sounds. [DOI: 10.1121/1.1621864]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Lj [NFV] Pages: 2835–2845

I. INTRODUCTION

The multitone masking paradigm developed by Neff and Green (1987) has been used by a number of investigators to examine the effects of random spectral variation on simultaneous masking (e.g., Neff *et al.*, 1993; Neff, 1995; Oh and Lutfi, 1998; Wright and Saberi, 1999; Micheyl *et al.*, 2000; Richards *et al.*, 2002; Durlach *et al.*, 2003a). Typically, the task of the listener is to detect a pure-tone signal of a fixed and known frequency embedded in a masker that is composed of a small number (e.g., 2–10) of tones having frequencies that are randomly selected on each presentation. The masking that is observed, which can be quite large on average but varies considerably across individual observers (Neff and Dethlefs, 1995; Oh and Lutfi, 1998; Durlach *et al.*, 2003a; Lutfi *et al.*, 2003b), does not appear to be attributable to overlapping patterns of excitation on the basilar membrane and in the auditory nerve, but instead to central processes.

Kidd *et al.* (1994) found that when multiple independent samples of these multitone maskers were played in rapid succession within an observation interval (multiple-bursts different, or MBD), the masking obtained relative to a single burst (SB) was significantly reduced—an effect that we will refer to as the ''MBD advantage.'' No reduction in masking was found when the SB masker was simply repeated in a burst sequence of equal length (multiple-bursts same, or MBS). In the SB condition, thresholds were measured for a 50-ms, 8-component multitone masker (as well as for maskers having fewer components). In the MBD condition, four independent 50-ms bursts were played in each observation interval. Although subjects differed considerably in their performance, thresholds were, on average, nearly 20 dB lower for the MBD condition than for the SB condition. Further, in a second experiment with a different group of subjects, the group-mean thresholds for MBD maskers composed of many bursts (from 8 to 32) were about 20 dB lower than for a 4-burst masker. All of the maskers had equal sequence duration, however, so that the duration of the individual bursts became shorter as the number of bursts increased. Although it is somewhat difficult to compare across the two experiments because of the differences in stimuli and the large differences among subjects, the overall pattern of results suggests that—at least for 1 to 8 bursts—increasing the number of bursts of the MBD stimulus significantly de-

[a)]Portions of this work were presented at the Midwinter Meeting of the Association for Research in Otolaryngology in St. Petersburg Beach, FL, February 2002.
[b)]Electronic mail: gkidd@bu.edu

creased the amount of informational masking. The interpretation of Kidd *et al.* (1994) for the reduction in informational masking found for the MBD masking condition was that the signal tone—which was constant across bursts within a sequence—formed a coherent auditory object that segregated from the masker. The perceptual segregation of a signal from a masker can be a very effective means for reducing informational masking (e.g., Neff, 1995; Oh and Lutfi, 1998; Durlach *et al.*, 2003a). Thus, the advantage found for the MBD masker relative to the SB masker, which may be as great as 40–50 dB for some subjects (Kidd *et al.*, 1994), was thought to reflect perceptual processes involved in the grouping and segregation of sounds.

An alternative explanation to that offered above is that the MBD advantage reflects multiple chances to detect the signal. Because the stimulus is arranged in N sequential bursts in an observation interval, the observer has N opportunities to detect the signal. Further, because the maskers in each burst are independent samples, it might be expected that considerable statistical power could be obtained by optimally combining the information from each burst. Thus, from a statistical decision theory perspective, some advantage in MBD re: SB might be expected simply based on the fact that there are more independent samples available to the observer. Quantitative predictions of the improvement in detectability owing to multiple observations have long formed a fundamental part of signal detection theory (e.g., Swets *et al.*, 1959; Green and Swets, 1966), and models based on statistical summation have been proposed as explanations for informational masking (e.g., Lutfi, 1993). One particular model in which the observations occur sequentially, and thus appears to be a candidate for application to the MBD masking experiment, is termed the ''multiple-look'' model (Viemeister and Wakefield, 1991).

The present study is an attempt to evaluate these two alternative explanations for the MBD advantage. While there is little doubt that a multiple-look model would predict some improvement in performance from multiple bursts re: SB, it is by no means certain that it can account for the large advantages found previously. Two approaches toward evaluating the feasibility of a multiple-look interpretation of the MBD advantage were taken here. First, both the number of bursts in a sequence and the length of time between the bursts (interburst interval, or IBI) were varied parametrically. To be consistent with the multiple-look model, performance should improve as the number of bursts increases,[1] but should be unaffected by changes in IBI. This is because the version of the multiple-look model that is considered here[2] treats each look as independent and equally detectable, and assumes that the information from the looks is optimally combined. Therefore, the time between bursts/observations does not affect the predictions. In practice, this assumes that there is no loss of information in memory. If, however, a significant amount of information is lost over time due to limitations on short-term memory, the assumption of the model that looks are equally detectable may not hold. Also, if the listener tends to weight the looks differentially over time, giving less weight to earlier looks (e.g., Viemeister and Wakefield, 1991; Buus, 1999) than to later looks, the information from the sequence of looks may not be optimally combined, again violating the assumptions of the simple multiple-look model. This point is considered in detail in the discussion. Also, the assumption of independence between looks may not be valid for very short IBIs where the potential for interactions between rapid stimulus bursts would be expected to be greater than between moderate or longer IBIs.

A second method for assessing the multiple-look model is to examine the psychometric functions obtained in the different conditions. Theoretically, given the assumptions noted above, it would be expected that the index of detectability, d', should vary as the square root of the number of looks.[3] Further, the magnitude of the expected advantage in signal-to-noise ratio for a fixed d' afforded by multiple looks depends on the slopes of the psychometric functions (e.g., Donaldson *et al.*, 1997). The shallower the slopes of the psychometric functions, which are assumed to be parallel for different numbers of looks, the greater the advantage. Because the slopes of psychometric functions for informational masking are known to be very shallow (e.g., Neff and Callaghan, 1987), the argument that the MBD advantage is due to multiple looks seems plausible. Thus, it would seem to be worthwhile to compare predictions of the reductions in threshold for constant d' as the number of bursts increases with the changes in the obtained thresholds. Analysis of both of these factors—the dependence of threshold on IBI and the predictions obtained from the psychometric functions—should provide a clearer understanding of the basis for the MBD advantage.

II. METHODS

A. Listeners

Six young-adult college students served as listeners. The ages ranged from 18 to 21 years. All listeners had normal hearing as determined by audiometric evaluation. All six listeners had served as subjects in previous informational masking experiments including multitone conditions similar to those tested here.

B. Stimuli

The stimuli were computer-generated at a 20-kHz rate and subsequently low-pass filtered (KEMO VBF 8, 96 dB/octave) at 7.5 kHz. The digital waveforms were played through 16-bit DACs and an array of programmable analog equipment (Tucker-Davis Technology) including attenuators to control stimulus levels, mixers to combine stimuli, and headphone buffers to drive the earphones. The multitone maskers were composed of 60-ms bursts (10/40/10-ms rise/steady-state/decay with cosine-squared ramps) of 8 frequencies. There was no temporal overlap between bursts. The masker frequencies were drawn randomly and independently, without replacement, on each burst from a uniform distribution on a logarithmic scale bounded by 200 and 5000 Hz excluding a protected region extending from 851–1175 Hz. The maskers were presented at 70 dB SPL (61 dB/tone). Within a block of trials the number of bursts in a given masker was 1, 2, 4, or 8 with a different random draw of frequencies present in each burst. The maskers in the two

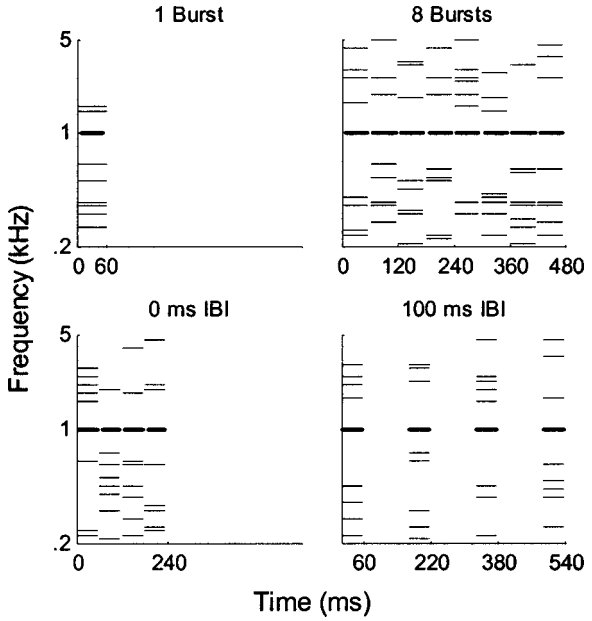


FIG. 1. Schematic examples of the multitone maskers and signal. In each panel, the 1000-Hz signal (bold) is shown along with a typical masker sample. The upper two panels illustrate different numbers of masker bursts (1 and 8) for the 0-ms IBI, while the lower panels illustrate 2 different IBIs (0 and 100 ms) for the 4-burst masker. Note the different scales for upper and lower panels.

intervals of each trial were also different sets of samples. There were five IBIs tested: 0, 50, 100, 200, and 400 ms. Some examples of the multitone maskers are illustrated schematically in Fig. 1.

In addition, a Gaussian noise masker was tested for comparison with the results from the multitone maskers. The Gaussian noise was digitally generated and filtered from 200–5000 Hz (slopes>96 dB/octave) with no protected region and was presented in a 1-burst condition and at all IBIs in a 4-burst condition. This masker was intended to produce primarily energetic masking (that is, masking produced by overlapping patterns of excitation in the auditory periphery; cf. Kidd *et al.*, 2002) and is included for comparison with the informational maskers. The temporal characteristics were the same as for the multitone maskers, and the overall level of presentation was also 70 dB SPL. The signal was a sequence of 1000-Hz tone bursts played synchronously with the masker bursts. Both signal and masker phases were chosen at random on each burst.

C. Procedures

The task was 2-interval, 2-alternative forced choice. The masker, when present, was held constant in level and the signal was varied adaptively in level in a 2-down, 1-up tracking routine that estimates the 70.7%-correct point on the psychometric function (Levitt, 1971). The starting level for the signal was set to 10–15 dB above the expected threshold so that it was clearly audible at the beginning of the adaptive track. The initial step size for the signal was 4 dB, which was reduced to 2 dB after four reversals. There was a minimum of 50 trials in each block of trials. If at least nine reversals were not obtained at the end of 50 trials, the block was extended until nine reversals were obtained. The thresholds that are reported are the means of the reversals after the first three or four were discarded (whichever resulted in an even number of reversals with a minimum of six). Response feedback was provided after every trial.

The subjects listened monaurally using their preferred ear through a calibrated TDH-50 earphone. They listened in individual, double-walled, sound-attenuating booths while seated in front of an LCD display which provided information about observation and response intervals and trial-by-trial feedback, and a keyboard, which allowed the registering of responses. Quiet signal thresholds were measured for the 8-burst signal and were, on average, 7.5, 9.7, 2.2, 1.0, 9.0, and −1.3 dB SPL, for each listener L1–L6, respectively. The data that are reported are means from a minimum of eight repetitions per subject per condition.

III. RESULTS

Individual masked thresholds from all six listeners for all multitone-masker conditions are presented in Fig. 2 and Table I.

The abscissa of Fig. 2 is the number of bursts and the ordinate is masked signal threshold. The parameter of the graph is IBI. Each panel shows the results for one listener. The 1-burst (or SB) condition is represented by a different

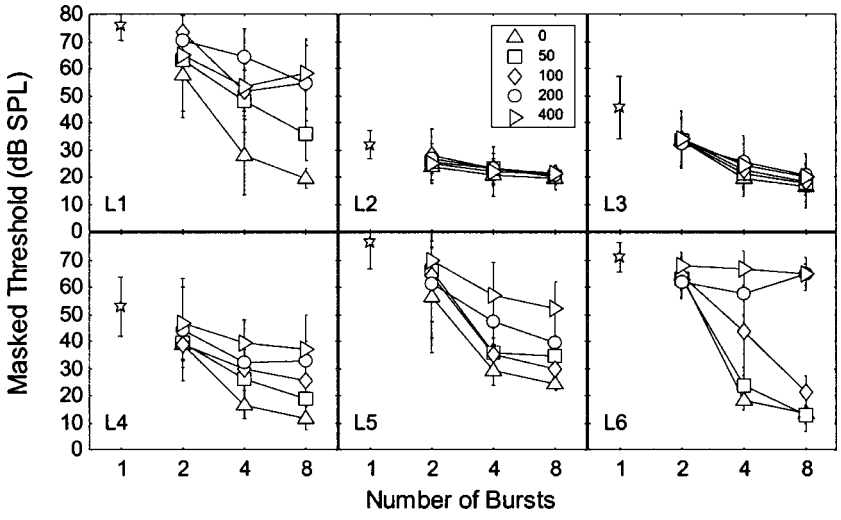


FIG. 2. Masked thresholds for all six listeners for all of the multitone masker conditions. The abscissa is the number of masker bursts in an interval and the ordinate is the sound-pressure level. The parameter of the graph is the interburst interval. The data points are means and standard deviations across repetitions. The star is for the single-burst condition. The figure key indicates the symbols for the different IBIs.

TABLE I. Individual average masked thresholds and standard deviations for listeners 1–6 (L1–L6) for all conditions. Also, group-mean thresholds and standard errors of the means are tabulated. Empty cells are for conditions not tested.

		Number of bursts					
IBI (ms)	Listener	1	1 (Gaussian)	2	4	4 (Gaussian)	8
0	L1	75.6 (5.1)	54.3 (3.3)	57.6 (15.7)	27.8 (14.5)	53.8 (6.6)	19.6 (3.7)
	L2	31.7 (5.3)	52.3 (2.2)	23.5 (2.5)	20.7 (2.9)	50.4 (2.3)	19.3 (3.9)
	L3	45.5 (11.4)	54.7 (3.8)	33.2 (8.7)	19.4 (6.4)	51.8 (2.7)	16.5 (8.1)
	L4	52.8 (10.9)	53.7 (2.7)	38.7 (8.3)	16.6 (5.1)	49.0 (2.3)	11.5 (4.2)
	L5	76.3 (9.8)	53.5 (2.6)	56.5 (20.7)	29.0 (5.4)	50.9 (1.5)	24.2 (2.5)
	L6	71.0 (5.5)	55.1 (3.3)	64.3 (8.4)	18.3 (3.7)	53.0 (2.0)	13.2 (3.2)
	AVG (se)	**58.8** (**7.5**)	**53.9** (**0.4**)	**45.6** (**6.6**)	**22.0** (**2.11**)	**51.5** (**0.7**)	**17.4** (**1.9**)
50	L1			62.9 (18.7)	48.0 (11.6)	53.3 (5.2)	35.5 (9.5)
	L2			25.6 (6.8)	22.8 (4.0)	50.5 (2.1)	19.8 (2.8)
	L3			33.4 (8.1)	21.2 (4.7)	52.3 (2.7)	17.6 (6.5)
	L4			39.6 (6.7)	26.3 (7.5)	50.6 (1.8)	18.6 (6.3)
	L5			66.4 (8.0)	36.0 (5.2)	51.8 (1.9)	34.7 (4.0)
	L6			63.3 (6.9)	23.6 (6.5)	50.0 (3.1)	12.9 (6.4)
	AVG (se)			**48.5** (**7.3**)	**29.7** (**4.3**)	**51.4** (**0.5**)	**23.2** (**3.9**)
100	L1			73.4 (11.0)	51.6 (10.2)	55.6 (4.2)	54.6 (13.7)
	L2			27.8 (10.0)	22.9 (2.7)	51.4 (1.3)	21.2 (2.7)
	L3			33.9 (10.2)	22.2 (4.4)	51.7 (3.7)	18.0 (3.4)
	L4			38.6 (8.1)	29.6 (5.0)	50.6 (2.8)	25.2 (5.6)
	L5			64.0 (16.7)	35.0 (4.0)	52.2 (1.8)	30.0 (2.3)
	L6			64.5 (5.4)	43.8 (16.3)	52.9 (3.0)	21.3 (5.9)
	AVG (se)			**50.4** (**7.8**)	**34.2** (**4.8**)	**52.4** (**0.7**)	**28.4** (**5.5**)
200	L1			70.3 (9.0)	64.0 (10.8)	55.0 (2.4)	54.3 (16.0)
	L2			26.7 (5.6)	22.7 (5.5)	52.3 (2.1)	20.6 (3.4)
	L3			32.4 (9.5)	25.2 (10.2)	53.5 (1.5)	20.8 (7.6)
	L4			44.3 (18.9)	32.4 (4.3)	50.2 (3.0)	32.6 (7.5)
	L5			60.9 (19.5)	47.4 (10.4)	53.5 (1.2)	39.2 (13.3)
	L6			61.7 (5.8)	57.6 (7.9)	54.0 (3.0)	64.7 (4.0)
	AVG (se)			**49.4** (**7.2**)	**41.6** (**7.1**)	**53.1** (**0.7**)	**38.7** (**7.3**)
400	L1			65.2 (7.2)	53.0 (8.6)	55.0 (4.6)	58.2 (13.0)
	L2			25.1 (2.7)	21.7 (9.1)	50.3 (2.0)	21.3 (3.2)
	L3			33.8 (10.4)	24.2 (8.1)	51.9 (2.3)	19.9 (5.1)
	L4			46.8 (13.5)	39.2 (8.8)	51.0 (3.5)	37.2 (12.4)
	L5			69.5 (7.5)	56.9 (12.1)	54.6 (1.4)	52.0 (10.0)
	L6			68.1 (3.0)	67.0 (6.2)	54.9 (1.9)	65.0 (5.9)
	AVG (se)			**51.4** (**7.8**)	**43.7** (**7.5**)	**52.9** (**0.9**)	**42.3** (**7.8**)

symbol (star) and is not connected to the other functions because it was only measured once and has no IBI associated with it.

First, it is apparent that subjects differ significantly with respect to the amount of masking obtained in these conditions. This is a common finding in studies of informational masking in general, and has often been reported in experiments employing multitone maskers and pure-tone signals in a detection task (e.g., Neff and Dethlefs, 1995; Oh and Lutfi, 1998; Durlach *et al.*, 2003a; Lutfi *et al.*, 2003b). The reasons for the large intersubject differences in informational masking are not well understood and continue to be of active research interest. Oxenham *et al.* (2003) note that highly trained musicians demonstrate less informational masking in tasks that require the subject to ''hear out'' a pure-tone signal, as was the case in the current study. Also, smaller intersubject differences appear to occur for other types of tasks demonstrating informational masking such as speech-on-speech masking (e.g., Brungart, 2001; Arbogast *et al.*, 2002).

Despite the differences between subjects noted above, all subjects showed a progressive decline in threshold as the number of bursts increased for the briefer IBIs, with the amount of improvement very much dependent on how much masking was produced in the 1-burst condition. The higher the threshold for a single burst, the greater the decline in threshold as the number of bursts increased. For example, the threshold for L2 in the 1-burst condition was about 31.7 dB SPL and only improved by about 12 dB as the number of bursts increased to 8 for the 0-ms IBI condition. In contrast, L1, L5, and L6 had 1-burst thresholds greater than 70 dB SPL, and for 0-ms IBI improved by more than 50 dB as the number of bursts increased to 8. In general, the advantage due to increasing the number of bursts was greatest at the shortest IBI and diminished as IBI increased. These three listeners still demonstrate more than 40 dB of improvement when the IBI is 50 ms, whereas the advantage for multiple bursts is greatly diminished for the longer IBIs. This is in contrast to L2 and L3, who show relatively low detection thresholds for the 1-burst case, modest improvement with increasing number of bursts, and little change with IBI. In cases where the thresholds were higher than the level of the tones comprising the masker, as occurred for L1, L5, and L6, the possibility cannot be ruled out that the subjects were simply responding to overall loudness which may well re-

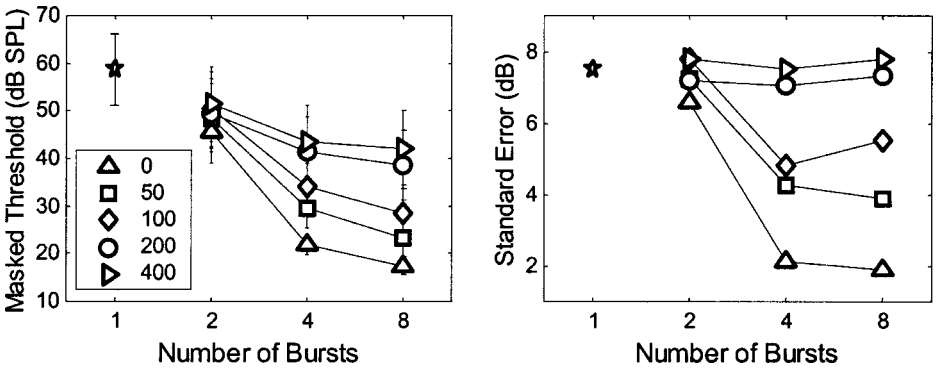


FIG. 3. The left panel shows group-mean masked thresholds and standard errors for all multitone masker conditions plotted in the same way as Fig. 2. The abscissa is the number of bursts and the parameter is the IBI. The right panel shows the standard error of the mean of the group thresholds for all of the multitone masker conditions plotted in the same manner as the left panel.

flect an upper limit on the amount of masking that is produced in these conditions.

Group-mean thresholds and standard errors of the mean are shown in the left panel of Fig. 3 plotted on the same coordinates as Fig. 2.

As a group, the mean decrease in threshold as the number of bursts increased from 1 to 8 was about 41.5 dB for 0 ms IBI, while at the 400-ms IBI, the reduction in threshold was about 16.5 dB. A related aspect of these findings is that the intersubject variability was greatest for the conditions in which the most masking occurred. This result is illustrated in the right panel of Fig. 3, which shows the standard error of the mean across subjects (same as the error bars shown in the left panel). Comparison of the two panels of Fig. 3 reveals a high degree of similarity in the overall form of the data. And, as might be expected based on such qualitative similarity, there was in fact a significant positive correlation ($r=0.90$, $p<0.001$) between the group-mean masked thresholds and the standard errors of the mean. This result reinforces the view that: (a) subjects differ considerably in their susceptibility to informational masking, and (b) conditions producing large amounts of informational masking will also be characterized by large intersubject differences.

The effect of varying IBI is illustrated in Fig. 4 as intersubject mean thresholds and standard errors of the means for each number of masker bursts. In this figure, the multitone data from Fig. 3 are replotted and the data from the Gaussian noise conditions are added.

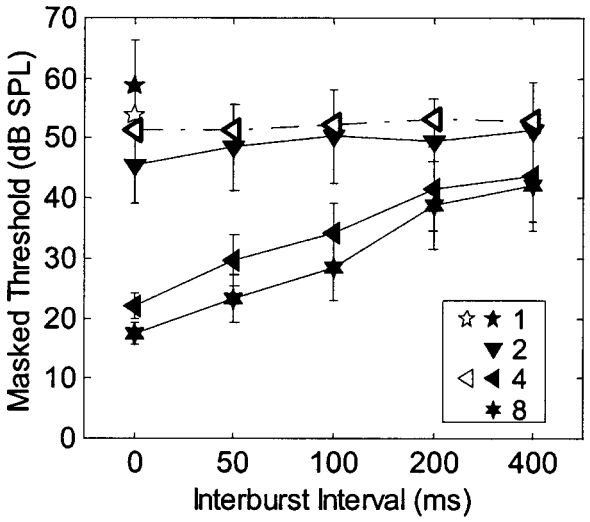


FIG. 4. The data from Fig. 3 are replotted along an abscissa that is IBI with the parameter being number of bursts (see the symbol key). In addition to the data from the multitone maskers, the results from the Gaussian noise masker are plotted (unfilled symbols). Single-burst results are arbitrarily plotted at an IBI of 0 ms.

It is apparent from these results that the effects of increasing IBI were greater for multitone maskers (filled symbols) having 4 or 8 bursts than for maskers having only 2 bursts. For 2-masker bursts, group-mean thresholds increased about 6 dB over the range of IBIs tested, while for 4- and 8-masker bursts the increase in threshold was more than 20 dB. An analysis of variance performed on the data from the multitone maskers (excluding the 1-burst condition) indicated that the main factors of number of bursts [$F(2,10)=27$, $p<0.001$] and IBI [$F(4,20)=7.0$, $p<0.001$] were highly significant, as was the 2-way interaction [$F(8,40)=3.9$, $p<0.002$].

It is of interest to compare the results obtained from the Gaussian noise masker (unfilled symbols) with the corresponding results from the multitone maskers because they presumably represent the consequences of two different types of masking, energetic and informational, respectively. First, on average, there was only a 2.5-dB decrease in threshold obtained by increasing the number of bursts from 1 to 4 at 0-ms IBI. The corresponding improvement for the multitone masker was about 37 dB. Also, the change in threshold with increasing IBI was much less for the noise masker, increasing only about 1.5 dB over the range of IBIs tested compared to an increase of 21.7 dB for the 4-burst multitone masker. If it is assumed that the listener could simply compare the last burst in the sequence in the first interval with the first burst of the sequence in the second interval, the effect of IBI would not be expected to be very large, because the level to which the thresholds increase should not exceed the 1-burst threshold. Hence, the difference in 1-burst and 4-burst thresholds determines the entire range of performance expected when IBI is varied. Overall, then, small effects of varying the number of bursts and the IBI were found for the subset of conditions tested with the Gaussian noise masker. In addition, the differences between listeners were very small, with standard errors of the mean less than 1 dB in all conditions.

The results from the multitone masker, which indicate that thresholds improve significantly with number of bursts (consistent with the multiple-look model) but worsen as IBI increases (inconsistent with the multiple-look model), can only partly answer the question of whether that model can

TABLE II. Individual and group-mean values of the slopes and proportions of variance accounted for, for the straight line fits (in coordinates of masked signal threshold in dB and log d') to the data for all conditions.

		Number of bursts											
		1		1 (Gaussian)		2		4		4 (Gaussian)		8	
IBI (ms)	Listener	Slope	Pvar	Slope	Pvar	Slope	Pvar	Slope	Pvar	Slope	Pvar	Slope	Pvar
0	L1	−0.001	0.01	0.029	0.64	0.005	0.13	0.004	0.22	0.033	0.66	0.024	0.73
	L2	0.016	0.63	0.057	0.76	0.035	0.63	0.047	0.98	0.090	0.92	0.063	0.93
	L3	0.008	0.51	0.051	0.70	0.017	0.55	0.026	0.72	0.062	0.82	0.016	0.62
	L4	0.007	0.43	0.072	0.94	0.015	0.68	0.041	0.56	0.076	0.70	0.053	0.80
	L5	−0.003	0.16	0.049	0.72	0.003	0.05	0.051	0.94	0.119	0.99	0.031	0.60
	L6	0.021	0.63	0.045	0.33	0.007	0.12	0.013	0.48	0.078	0.93	0.047	0.93
	AVG (se)	**0.008**(**0.004**)	**0.40**(**0.10**)	**0.05**(**0.006**)	**0.68**(**0.08**)	**0.014**(**0.005**)	**0.36**(**0.12**)	**0.030**(**0.008**)	**0.65**(**0.12**)	**0.076**(**0.012**)	**0.84**(**0.06**)	**0.039**(**0.007**)	**0.76**(**0.06**)
50	L1					0.001	0.00	0.010	0.34	0.031	0.65	0.014	0.57
	L2					0.018	0.85	0.045	0.80	0.088	0.81	0.064	0.89
	L3					0.016	0.85	0.038	0.67	0.059	0.84	0.024	0.79
	L4					0.008	0.57	0.018	0.53	0.055	1.00	0.021	0.84
	L5					0.015	0.35	0.015	0.85	0.096	0.91	0.035	0.78
	L6					0.020	0.72	0.010	0.68	0.034	0.66	0.011	0.23
	AVG (se)					**0.013**(**0.003**)	**0.56**(**0.14**)	**0.023**(**0.006**)	**0.64**(**0.07**)	**0.060**(**0.011**)	**0.81**(**0.06**)	**0.028**(**0.008**)	**0.68**(**0.10**)
100	L1					0.004	0.03	0.001	0.01	0.050	0.72	0.002	0.11
	L2					0.015	0.66	0.049	0.73	0.083	0.64	0.058	0.69
	L3					0.009	0.63	0.032	0.90	0.055	0.84	0.024	0.80
	L4					0.013	0.86	0.018	0.57	0.051	0.77	0.023	0.84
	L5					−0.002	0.03	0.019	0.78	0.077	0.84	0.012	0.62
	L6					0.014	0.81	0.001	0.01	0.076	0.95	0.007	0.29
	AVG (se)					**0.009**(**0.003**)	**0.50**(**0.15**)	**0.020**(**0.008**)	**0.50**(**0.16**)	**0.065**(**0.006**)	**0.79**(**0.04**)	**0.021**(**0.008**)	**0.56**(**0.12**)
200	L1					0.012	0.50	0.002	0.05	0.150	0.86	0.003	0.09
	L2					0.028	0.75	0.030	0.87	0.106	0.96	0.059	0.98
	L3					0.014	0.77	0.012	0.54	0.103	0.89	0.016	0.73
	L4					0.001	0.01	0.020	0.82	0.053	0.53	0.015	0.63
	L5					−0.002	0.09	0.007	0.29	0.068	0.80	0.009	0.40
	L6					0.022	0.90	0.009	0.28	0.073	0.74	0.014	0.68
	AVG (se)					**0.013**(**0.005**)	**0.50**(**0.15**)	**0.013**(**0.004**)	**0.48**(**0.13**)	**0.092**(**0.014**)	**0.80**(**0.06**)	**0.020**(**0.008**)	**0.58**(**0.13**)
400	L1					0.015	0.47	0.011	0.53	0.036	0.67	−0.005	0.16
	L2					0.048	0.95	0.011	0.23	0.105	0.69	0.047	0.85
	L3					0.013	0.57	0.027	0.73	0.031	0.35	0.019	0.69
	L4					0.006	0.30	0.016	0.81	0.051	0.49	0.010	0.46
	L5					0.016	0.42	0.007	0.31	0.086	0.61	0.005	0.08
	L6					0.036	0.96	0.010	0.33	0.074	0.77	0.012	0.77
	AVG (se)					**0.022**(**0.007**)	**0.61**(**0.11**)	**0.014**(**0.003**)	**0.49**(**0.10**)	**0.064**(**0.012**)	**0.60**(**0.06**)	**0.015**(**0.007**)	**0.50**(**0.13**)

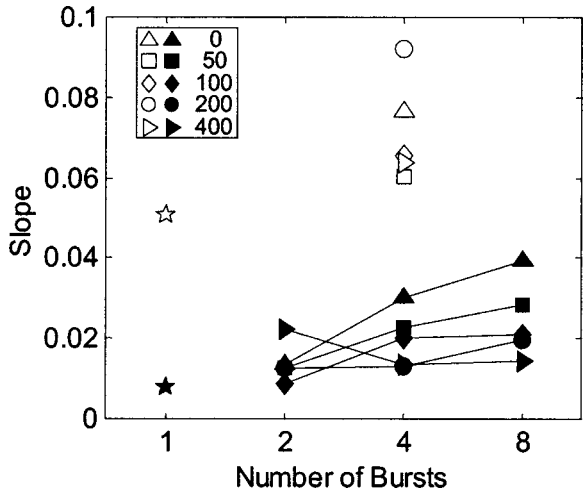


FIG. 5. Group-mean slope values and standard errors for the fits to the psychometric functions. The abscissa is the number of masker bursts and the ordinate is the slope obtained from the fits in logarithmic d' units (see the text). The parameter of the graph is the IBI. The filled symbols show the data for the MBD masker and the unfilled symbols show the data from the Gaussian noise masker.

account for the MBD advantage. This is because all of the data presented to this point were estimates obtained at a single level of performance; i.e., the 70.7%-correct point estimated by the adaptive tracking procedure. Before considering the extent to which a multiple-look model can explain these results, it is important to examine the underlying psychometric functions.

The psychometric functions were reconstructed from the trial-by-trial record of the adaptive tracks used to measure masked thresholds. The percent-correct values were converted to d', the logarithms of the d's were obtained, and the data (signal level in dB versus $\log d'$) were fit with weighted (by number of trials) linear least-squares functions. Table II is a summary of the fits to the data for all subjects and conditions.

As is apparent from inspection of Table II, the slopes of the reconstructed psychometric functions vary considerably from individual to individual and condition to condition, and the fits are often quite poor, especially, as would be expected, for the very shallow slopes. Ideally, an analysis of the slopes of these psychometric functions would not be based on estimates obtained from adaptive tracking runs. This is because adaptive procedures concentrate the levels of the adapting stimulus around a single point on the function (70.7%-correct here) and fewer trials are obtained at levels remote from that point, potentially reducing the precision of the reconstructed functions. Nonetheless, there are consistent trends apparent in the data. Figure 5 contains the group-mean slopes of the psychometric functions for the multitone masker (filled symbols) plotted as a function of the number of bursts for different IBIs.

First, the slope for 1 burst is the shallowest of all at about 0.008 $\log d'$ units per dB. For 0-ms IBI, the slope increased as the number of bursts increased, up to a value of about 0.04 for 8 bursts. And, continuing the inspection of the 8-burst data, the group-mean slopes decreased in an orderly way as IBI increased declining to a value of about 0.015 at an IBI of 400 ms. Also shown are the slope values obtained from the fits to the data from the Gaussian noise masker (unfilled symbols). The slopes for all of the Gaussian noise masker conditions are steeper than all of the corresponding multitone masker conditions; in fact, the two sets of slope values are completely nonoverlapping.

An analysis of variance performed on the slopes for the multitone maskers (excluding the 1-burst condition) revealed that IBI was a significant main factor [$F(4,20)=5.93$, $p<0.005$], whereas number of bursts was not [$F(2,10)=2.85$, $p=0.105$]. The interaction between IBI and number of bursts was also significant [$F(8,40)=3.12$, $p<0.01$]. The group-mean slopes for the multitone maskers were negatively correlated with the masked thresholds ($r=-0.88$, $p<0.001$ for all values including the 1-burst condition). The correlation between all of the individual slopes and thresholds was also significant at $r=-0.58$ ($p<0.001$). Group-mean slopes versus thresholds for the Gaussian noise masker were not significantly correlated ($r=-0.19$, $p>0.1$). By comparison, the correlation of group-mean slopes and thresholds using the multitone data for only 1 and 4 bursts was statistically significant ($r=-0.96$, $p<0.01$). Thus, a general trend is that the slopes vary in an orderly way, with shallow slopes corresponding to conditions producing large amounts of informational masking and steeper slopes found for conditions producing less informational masking. These results are in agreement with others using the multitone masking procedure (e.g., Neff and Callaghan, 1987; Durlach *et al.*, 2003c; Lutfi *et al.*, 2003a) and with a similar trend found in the slopes of performance-level functions for discrimination and identification studies (e.g., Kidd *et al.*, 1998, 2002; Arbogast *et al.*, 2002). Taken as a whole, these findings suggest that shallow psychometric function slopes are a general characteristic of informational masking. The slopes reported here for detecting a tone in a Gaussian noise masker range from 0.05–0.09, which are in good agreement with the slopes of psychometric functions reported by Buus (1999), which averaged about 0.076 for detecting a tone in 1, 2, 4, and 6 bursts of narrow-band noise. They are slightly shallower, though, than the range from 0.08–0.15 for normal hearing listeners reported by Carlyon *et al.* (1990) for unmasked detection of tone sequences and the group-mean slope of 0.15 reported by Viemeister and Wakefield (1991) for detecting tone pulses in noise. Overall, then, the slope analysis indicated that there was a significant negative correlation between slope and masked threshold in the multitone masking conditions. No significant correlation between these two variables was found for the Gaussian noise masker conditions. In all cases, the slopes were much steeper when the masker was Gaussian noise than when the masker was a multitone complex for either the 1-burst or MBD conditions. Despite the poor fits, the conclusion that psychometric function slopes are roughly inversely related to the amount of informational masking appears to hold for the conditions tested in this study and may prove to be a general property of informational masking.

IV. DISCUSSION

There were two main purposes of the current study. The first purpose was to examine parametrically the effect of varying the number of bursts and IBI in the MBD measurement paradigm. The results of the parametric examination of those variables are discussed below. The second purpose of this study was to evaluate the hypothesis that the MBD advantage can be attributed to multiple looks. For a number of reasons also discussed below, our conclusion, acknowledging certain qualifications, is that a simple multiple-look explanation for the MBD advantage is not supported by these data.

To summarize the empirical findings: first, for the multitone maskers, it was noted that subjects differed in the amount of informational masking produced in the various number-of-bursts/IBI conditions. The intersubject differences were not constant across conditions but were significantly correlated with the amount of masking that was produced. There was no similar correlation found for the Gaussian noise masker. The finding that subjects differ more in informational masking than energetic masking was expected based on earlier work by Kidd *et al.* (1994) using the multiple-bursts paradigm and the related single-burst studies of others (e.g., Neff and Dethlefs, 1995; Neff, 1995; Oh and Lutfi, 1998; Richards *et al.*, 2002; Durlach *et al.*, 2003a; Lutfi *et al.*, 2003b). Nonetheless, the systematic variation of the magnitude of the individual differences as a function of the amount of informational masking obtained across the various conditions tested here was remarkable. Second, the group-mean data indicated a clear and significant role of both number of bursts and IBI, as well as an interaction between the two. At short IBIs, increasing the number of bursts resulted in an enormous decrease in threshold for most listeners. This was particularly true for increasing the number of bursts from 1 to 4, but less of an advantage was found in increasing the number of bursts from 4 to 8. The latter may reflect a ''floor'' effect, whereby the informational component of masking has been nearly eliminated. If the remaining masking was primarily energetic, then little improvement with increasing number of bursts would be expected—consistent with the finding from the (largely energetic) Gaussian noise masker conditions tested here. And, assuming that energetic masking is determined primarily by peripheral factors, the differences between subjects would be expected to be much less than for conditions dominated by informational masking (e.g., Neff and Dethleffs, 1995; Oxenham *et al.*, 2003). In the group-mean data, at long IBIs, thresholds were relatively high for all numbers of bursts and the amount of improvement as the number of bursts increased was significantly diminished. This statement should be qualified by the observation that two of the listeners, L2 and L3, show smaller but substantial benefit of increasing the number of bursts for 0-ms IBI and little difference in that improvement for longer IBIs.

The group-mean slopes of the psychometric functions for the multitone masker conditions were found to vary in a systematic way that was significantly (negatively) correlated with the masked thresholds. This meant that the conditions producing the widest range of thresholds—the briefer IBIs—also had the widest range of psychometric-function slopes (i.e., deviated the most from parallel).

The conclusion that the multiple-look model cannot account for the MBD advantage is based on findings that are inconsistent with the assumptions of the simple form of the model considered here. For conditions in which large reductions in the group-mean masked thresholds occurred for multiple bursts relative to that obtained for a single burst, increasing IBI significantly increased thresholds—more than 25 dB in some cases for the IBIs tested here. As discussed in the Introduction, the increase in thresholds with increasing IBI is not consistent with the predictions of a multiple-look model that assumes that each look is equally detectable and optimally combined. Likewise, the finding that the slopes are not constant across conditions, but vary in a systematic way, also violates the assumptions of the model.

There are at least two potential explanations, which may not be mutually exclusive, for the finding that thresholds increase with increasing IBI. It is possible that there is a rather massive loss of information during the IBI due to decay of a short-term/working memory store. This could be a reflection of increasing ''memory noise'' over time. Or, it could be that this result occurred because of a loss of signal coherence as the bursts are separated in time, leading to a diminishing of the perceptual segregation of the signal from the masker. With respect to the explanation based on decay of short-term auditory memory, the issue is complex. Viemeister and Wakefield (1991) do not comment directly on how decay of memory might compromise the stored looks or how that might affect the ability of the observer to use that information. They describe the memory process involved in the multiple-look model (as applied to temporal integration for detection threshold) as a vector of observations that are each the output of a 3-ms sliding window (following filtering and rectification) and are stored in a buffer that is subsequently available for processing. They speculate that ''These looks are available for various types of computations and comparisons. The observer can scan this vector to find, for example, periodic envelope fluctuations or a gap in the input...'' (p. 859) and ''...(the looks) can be *selectively* used for further processing and decision making'' (p. 865). As an example, they point out that the classical temporal integration function proposed by Plomp and Bouman (1959) can be fairly well approximated by assuming that the looks are weighted according to an exponentially increasing function, and that looks nearer in time are weighted more heavily than looks remote in time and could reflect any of a number of possible mechanisms. For example, the weights might indicate the decay of short-term memory by assuming that the weight is proportional to the ''strength'' of the trace of the sound, which could diminish over time, or perhaps by assuming that memory noise increases over time, leading the observer to count earlier data as less reliable than later data. Buus (1999) has shown that differential weighting of looks can be used to account for temporal integration functions that exhibit segments having different slopes. The estimates of weights he obtained for detecting a tone in multiple bursts of a narrow band of noise, while statistically significant, varied only modestly over time. In a study by Carlyon *et al.* (1990),

the general assumption with respect to memory seems to be that spacing the tone pulses farther apart—up to 160 ms for pulse trains consisting of as many as 10 pulses—simply assured that each would be independent. Their data show some evidence of a decline in performance with increasing interpulse interval, but nothing approaching what was found here. However, as mentioned earlier, the amount that performance can be degraded as the length of time between stimuli is increased is presumably limited by performance for a single stimulus. For example, in Carlyon *et al.* (1990), the difference in thresholds between 1 pulse and 10 pulses was, on average, about 8 dB. Likewise in the current study, as the number of bursts increased from 1 to 4 for the Gaussian noise masker, the decrease in signal threshold was only about 2.5 dB.[4] For the informational (multitone) maskers used here the range of thresholds is several orders of magnitude greater. Also, the expectation about the role of loss of information due to decay of memory between stimuli depends on the type of memory process that applies to the task and stimuli. Perhaps the most widely accepted theory of how memory for auditory stimuli affects detection or discrimination performance is that proposed by Durlach and Braida (1969). Their theory posits two distinctly different forms of memory for sounds: a ''sensory-trace'' coding mode, in which an image of the stimulus is actively maintained through rehearsal over time until a judgment is made or another stimulus is available for comparison. Sensory-trace coding memory is highly susceptible to corruption or noise over time and performance characteristically decreases as a function of interstimulus interval (ISI). The second form of memory for sounds is referred to as ''context-coding'' memory, in which a verbal representation of some value or property of the stimulus is stored. Context-coding memory is relatively immune to decay over time and is limited by the range of stimulus values encountered in the experiment. In order for the extremely large effects of IBI found here to be consistent with a decay-of-memory explanation, one would presumably have to argue that listeners were limited to storing and comparing sensory traces and that the amount of decay or noise occurring over time would be much larger than that which has typically been observed in the past. Green *et al.* (1983) demonstrated that when either simple or complex sounds were presented in a way that required successive comparisons (of sound intensity in their study), performance declined in the expected way with increasing interstimulus interval (ISI) out to 8 s. The magnitude of the increase in thresholds they found—corresponding to a level difference of about 3.5 dB—was still much smaller than that which was found here. However, when the judgment could be made independently in the two intervals (determining a particular type of alteration in the shape of a known complex spectrum), no increase in threshold was observed over the same 8-s ISI. They speculated that, when the task involved a simultaneous comparison of level across frequency (profile analysis), each stimulus could be evaluated separately without the need to hold a sensory trace of the first stimulus of the pair during the ISI. However, the issue is not simply broadband versus narrow-band analysis or intensity discrimination versus profile analysis. Kidd *et al.* (1988) demonstrated that even spectral shape judgments could be susceptible to the decay of sensory traces if the reference spectra were randomized between trials, thereby forcing the subjects to directly compare successive sounds. However, when the sounds were randomized in spectral shape on every presentation, performance not only became worse overall, the thresholds were not greatly affected by ISI, suggesting that perhaps the listeners were adopting a context-coding mode of operation. The stimuli used in the Kidd *et al.* (1988) study were in some ways similar to the maskers used in the present experiments: they were composed of multiple tones that were randomized in a way that caused significant spectral variability from sample to sample. To the extent that the Kidd *et al.* (1988) findings have any bearing on the issue of how decay of memory might have influenced the current results, they would support the interpretation that the large increases in threshold found with increasing IBI were not simply due to sensory-trace memory noise. In any case, we are not aware of any evidence that would support the interpretation that decay of short-term memory during the burst sequence can account for the 25-dB effect found here.

The other interpretation of this result, that the progressive elevation in threshold as a function of IBI is due to a loss of auditory stream coherence, seems more likely if only because of the time frame involved. Dating from the early work of van Noorden (1975; see also the recent review by Moore and Gockel, 2002), it has been known that slowing down the rate of, or inserting sufficiently long silent intervals between, tone pulses forming an auditory stream will lead to the loss of the perception of coherence of the elements. Bregman *et al.* (2000) found that the rated difficulty of maintaining a coherent stream (a ''galloping rhythm'') for a given frequency difference between tones played in a sequence was generally monotonically related to the ISI (time between elements of the same frequency) for a range of ISIs (50–150 ms) overlapping the range of IBIs tested here. However, stream formation is known to build up over a longer time frame than tested here, and recent evidence has demonstrated that the build-up of streaming is highly dependent on attentional focus (Carlyon *et al.*, 2001). If in fact the MBD advantage is due to the perception of a stream emerging from a random background as the burst sequence progresses, it is not surprising that the coherence of the stream would both build up as the number of bursts increases and diminish as the time between the elements increases. Some caution is necessary, however, in distinguishing between these two explanations—decay of sensory-trace memory stores and loss of stream coherence—for the large effects of IBI found here for the MBD stimuli. This is because both phenomena depend on persistence of excitation—almost certainly at higher levels of the auditory system—over time in common frequency regions. Although the time course and magnitude of the effects seem very different, the experimental tasks and procedures used to examine them are also very different. Because of this, and because the results are presented in very different metrics, we cannot rule out the possibility that both are manifestations of the same memory process and the difference has to do with the degree of active effort of the

listener to maintain, rehearse, or otherwise influence the image.

A second important assumption of the multiple-look model that is violated by the current findings is that the psychometric functions are parallel across conditions—especially for a given number of bursts as IBI is varied. Given the discussion above, it might be possible to account for the data from the 0-ms IBI condition by relaxing the assumption that the looks contribute equally to detection. This could occur either because the looks are not equally detectable or they are not optimally combined. It is plausible, for example, that a weighting function in which more recent looks contribute more heavily to detection than earlier looks, combined with a very shallow underlying psychometric function, could predict the magnitude of the MBD advantage found here. However, this still does not reconcile the systematic variation in the slopes of the functions. A particularly clear explanation of the relationship between number of looks and the slope of the psychometric function is found in Donaldson *et al.* (1997; see their Fig. 1 and the derivation in Appendix, p. 3719). They studied temporal integration in listeners with cochlear implants and evaluated the extent to which a multiple-look model could account for that phenomenon. As they explain, the expectation regarding the improvement in performance due to multiple observations is that the slopes of the psychometric functions remain invariant (in coordinates of $\log d'$ and signal level in dB). The amount of improvement with increasing number of looks is thus inversely related to the slopes of the functions. Because of the necessity of this assumption, Buus (1999) fit his functions using one value for slope across number of bursts. Such an analysis is not possible here because, as discussed above when considering the data in Table II and Fig. 5, there is a significant and systematic change in the slopes of the functions as the number of bursts/IBI changed. In fact, the results are completely opposite of what one might expect based on a multiple-look model. Because the slopes covary with masked thresholds, the conditions that produced the greatest MBD advantage (brief IBIs) are the ones where the psychometric function slopes changed the most as the number of bursts was changed. The finding that the slopes of the psychometric functions vary in a systematic way suggests that different mechanisms—energetic and informational masking—affect the results to varying degrees in the different conditions.

A final point regarding the application of the multiple-look model to the MBD stimulus is that it is difficult to specify the initial stages of the model with the same precision that Viemiester and Wakefield (1991) did for temporal integration. For detecting a tone in quiet or in fixed-level Gaussian noise, the classical critical-band energy-detector model has been widely and successfully applied. For informational masking, though, performance must certainly be related to the energy in the frequency region near the signal, but masker energy remote from the signal is also crucial. Thus, the attempts to model informational masking for detecting a tone in a multitone masker are usually channel models having provisions for weights in nonsignal channels[5] (e.g., Lutfi, 1993; Oh and Lutfi, 1998; Gilkey *et al.*, 2001; Richards *et al.*, 2002). An even greater complication in modeling informational masking has recently been pointed out by Durlach *et al.* (2003a,b) in which it is argued that factors other than masker uncertainty, such as signal–masker similarity, can be important in informational masking experiments and must ultimately be taken into account in any comprehensive theory.

To summarize, these experiments extend the earlier report by Kidd *et al.* (1994) indicating that multiple bursts of the multitone masker originally devised by Neff and Green (1987) can greatly reduce the informational masking produced by single bursts of that masker. The effect appears to be due to the perception of the signal as a coherent auditory stream embedded in a random background, and can be greatly reduced simply by adding silent intervals between bursts. The multiple-look model, which has been used to explain temporal integration and has helped to solve the resolution-integration paradox (Viemiester and Wakefield, 1991), does not appear to provide a satisfactory account of the present results.

ACKNOWLEDGMENTS

This work was supported by NIH/NIDCD research grants Nos. DC04545 and DC02012, as well as by the Boston University Hearing Research Center. The authors are grateful to Nathaniel I. Durlach, Robert A. Lutfi, Associate Editor Neal F. Viemeister, and an anonymous reviewer for comments on an earlier version of this article. We are grateful to Dr. Lutfi for pointing out the application of the CoRE model discussed in the footnote below. We also wish to thank Brian Fligor for help with the data collection.

[1]Here, we are equating a stimulus burst with a single observation or ''look.'' This is a simplification that is likely not strictly true. If one assumes a temporal window much shorter than the burst duration, such as the 3-ms duration used by Viemeister and Wakefield (1991), each burst would be ''looked at'' several times. However, excepting perhaps a small effect from internal noise that is uncorrelated over time, no detection performance advantage would be expected from these highly correlated observations and thus the simplification of equating number of looks with number of stimulus bursts seems reasonable.

[2]The multiple-look model we consider in this article is a simple version similar to that proposed by Viemeister and Wakefield (1991) for temporal integration of detection threshold. In this model, it is assumed that the looks are independent, equally detectable, and optimally combined. Other forms of the model may produce different predictions.

[3]Lutfi (1989) presents evidence suggesting that d' increases as the cube root, rather than the square root, of the number of stimulus elements or observations in some psychophysical discrimination tasks.

[4]The decrease in threshold for the signal in Gaussian noise was about 2.5 dB, on average, when the number of bursts was increased from 1 to 4. This is less than would be predicted from optimum combination of information across bursts. Using the slope value obtained from averaging across all observers and conditions (0.0683), which is reasonable given that there was no significant effect of condition for Gaussian noise, then the predicted decrease in threshold would be about 4.4 dB. That value is very close to what Carlyon *et al.* (1990) report for increasing the number of signal-in-quiet pulses from 1 to 4 at a short interpulse interval (5 ms). It is not known why the effect found here was somewhat less than expected.

[5]The component relative entropy (CoRE) model proposed by Lutfi (1993) provides an excellent account of the relationship between masked thresholds and number of masker components in the multitone masking experiment reported originally by Neff and Green (1987) and later extended by Oh and Lutfi (1998). The form of the masking function (assuming masker power is held constant as the number of components is varied) typically rises initially as the number of masker tones increases, reaches a maximum

around 10 to 20 components, then falls toward a point anchored by the threshold found in broadband noise. The key element of the model driving this function is the variance of the outputs of a set of attended-to auditory filters. When only a few masker components are present, the expected variance of the outputs of the filters, based on statistical summation across trials, is relatively high, and thus the predicted informational masking is also high. As the number of masker components increases, the variance in the outputs of the filters decreases as components increasingly fall in the same filters. A similar argument could be made in applying the CoRE model to the MBD advantage found here. For the 1-burst masker, the variance in the outputs of the auditory filters would be expected to be relatively high, producing substantial informational masking. As the number of bursts increases, the total number of masker components in a given observation interval increases as the product of number of bursts and number of components. Thus, the expected variance of the outputs of the auditory filters declines and the predicted thresholds decrease. Referring to the results plotted in the left panel of Fig. 3, choosing reasonable CoRE model parameters could closely approximate the curve showing masked threshold for this group of listeners as the number of masker bursts in an interval (and thus the total number of masker components per interval) increases. However, because the only parameter of the model that can be varied to attempt to account for the effect of IBI is the number of attended-to channels, the model is not successful in predicting the changes in thresholds due to IBI shown in Fig. 3. As is apparent in Fig. 3, increasing IBI interacts with the number of bursts causing the masking curves to diverge as IBI/number-of-bursts increases. If we allow the number of attended-to bands to vary in the CoRE model, the predicted masking curves converge with increasing IBI/number-of-bursts rather than diverge. Thus, although the model can predict the MBD advantage for a subset of the findings presented in this study (e.g., brief IBI), it is not successful in generalizing to the family of IBIs tested here.

Arbogast, T. L., Mason, C. R., and Kidd, Jr., G. (**2002**). ''The effect of spatial separation on informational and energetic masking of speech,'' J. Acoust. Soc. Am. **112**, 2086–2098.

Bregman, A. S., Ahad, P. A., Crum, P. A. C., and O'Reilly, J. (**2000**). ''Effects of time intervals and tone durations on auditory stream segregation,'' Percept. Psychophys. **62**, 626–636.

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

Buus, S. (**1999**). ''Temporal integration and multiple looks, revisited: Weights as a function of time,'' J. Acoust. Soc. Am. **105**, 2466–2475.

Carlyon, R. P., Buus, S., and Florentine, M. (**1990**). ''Temporal integration of trains of tone pulse by normal and cochlearly impaired listeners,'' J. Acoust. Soc. Am. **87**, 260–269.

Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (**2001**). ''Effects of attention and unilateral neglect on auditory stream segregation,'' J. Exp. Psychol. Hum. Percept. Perform. **27**, 115–127.

Donaldson, G. S., Viemeister, N. F., and Nelson, D. A. (**1997**). ''Psychometric functions and temporal integration in electric hearing,'' J. Acoust. Soc. Am. **101**, 3706–3721.

Durlach, N. I., and Braida, L. D. (**1969**). ''Intensity perception. I. Preliminary theory of intensity resolution,'' J. Acoust. Soc. Am. **46**, 372–383.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, Jr., G. (**2003a**). ''Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity,'' J. Acoust. Soc. Am. **114**, 368–379.

Durlach, N. I., Mason, C. R., Kidd, Jr., G., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003b**). ''Note on informational masking,'' J. Acoust. Soc. Am. **113**, 2984–2987.

Durlach, N. I., Mason, C. R., Arbogast, T. L., and Kidd, G., Jr. (**2003c**). ''Informational masking: Psychometric functions for 'frozen' and 'mixed' maskers,'' Presented at the Midwinter Research Meeting of the Association for Research in Otolaryngology, Daytona, FL.

Gilkey, R. H., Mason, C., and Kidd, Jr., G. (**2001**). ''Using reproducible noise to illuminate the nature of informational masking,'' J. Acoust. Soc. Am. **109**, 2468.

Green, D. M., Kidd, Jr., G., and Picardi, M. C. (**1983**). ''Successive versus simultaneous comparison in auditory intensity discrimination,'' J. Acoust. Soc. Am. **73**, 639–643.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York) (reprinted by R. A. Kreiger, 1973).

Kidd, Jr., G., Mason, C. R., and Hanna, T. E. (**1988**). ''Evidence for sensory-trace comparisons in spectral shape discrimination,'' J. Acoust. Soc. Am. **84**, 144–149.

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

Kidd, Jr., G., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). ''Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **104**, 422–431.

Kidd, Jr., 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.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Lutfi, R. A. (**1989**). ''Informational processing of complex sound. I. Intensity discrimination,'' J. Acoust. Soc. Am. **86**, 934–944.

Lutfi, R. A. (**1993**). ''A model of auditory pattern analysis based on component-relative-entropy,'' J. Acoust. Soc. Am. **94**, 748–748.

Lutfi, R. A., Kistler, D. J., Callaghan, M. R., and Wightman, F. L. (**2003a**). ''Psychometric functions for informational masking,'' J. Acoust. Soc. Am. **113**, 2288.

Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (**2003b**). ''One factor underlies individual differences in auditory informational masking within and across age groups,'' Percept. Psychophys. **65**, 396–406.

Micheyl, C., Arthaud, P., Reinhart, C., and Collet, L. (**2000**). ''Informational masking in normal-hearing and hearing-impaired listeners,'' Acta Oto-Laryngol. **120**, 242–246.

Moore, B. C. J., and Gockel, H. (**2002**) ''Factors influencing sequential stream segregation, Acustica, **88**, 320–332.

Neff, D. L. (**1995**). ''Signal properties that reduce masking by simultaneous, random-frequency maskers,'' J. Acoust. Soc. Am. **98**, 1909–1920.

Neff, D. L., and Callaghan, B. P. (**1987**). ''Psychometric functions for multicomponent maskers with spectral uncertainty,'' J. Acoust. Soc. Am. Suppl. 1 **81**, S53.

Neff, D. L., and Green, D. M. (**1987**). ''Masking produced by spectral uncertainty with multicomponent maskers,'' Percept. Psychophys. **41**, 409–415.

Neff, D. L., and Dethlefs, T. M. (**1995**). ''Individual differences in simultaneous masking with random-frequency, multicomponent maskers,'' J. Acoust. Soc. Am. **98**, 125–134.

Neff, D. L., Dethlefs, T. M., and Jesteadt, W. (**1993**). ''Informational masking for multicomponent maskers with spectral gaps,'' J. Acoust. Soc. Am. **94**, 3112–3126.

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

Oxenham, A. J., Fligor, B., Mason, C. R., and Kidd, Jr., G. (**2003**). ''Informational masking and musical training,'' J. Acoust. Soc. Am. **114**, 1543–1549.

Plomp, R., and Bouman, M. A. (**1959**). ''Relation between hearing threshold and duration for tone pulses,'' J. Acoust. Soc. Am. **31**, 749–758.

Richards, V. M., Tang, Z., and Kidd, Jr., G. (**2002**). ''Informational masking with small set sizes,'' J. Acoust. Soc. Am. **111**, 1359–1366.

Swets, J. A., Shipley, E. F., McKey, M. J., and Green, D. M. (**1959**). ''Multiple observation of signals in noise,'' J. Acoust. Soc. Am. **31**, 514–521.

van Noorden, L. P. A. S. (**1975**). ''Temporal coherence in the perception of tone sequences,'' Doctoral thesis, Eindhoven.

Viemeister, N. F., and Wakefield, G. H. (**1991**). ''Temporal integration and multiple looks,'' J. Acoust. Soc. Am. **90**, 858–865.

Wright, B. A., and Saberi, K. (**1999**). ''Strategies used to detect auditory signals in small sets of random maskers,'' J. Acoust. Soc. Am. **105**, 1765–1775.

Interactions between test- and inducer-tone durations in induced loudness reduction[a)]

Bärbel Nieder[b)]
Institute of Hearing, Speech and Language and Department of Speech-Language Pathology and Audiology (106A FR), Northeastern University, Boston, Massachusetts 02115

Søren Buus
Institute of Hearing, Speech and Language and Communications and Digital Signal Processing Center, Department of Electrical and Computer Engineering (440 DA), Northeastern University, Boston, Massachusetts 02115

Mary Florentine
Institute of Hearing, Speech and Language and Department of Speech-Language Pathology and Audiology (106A FR), Northeastern University, Boston, Massachusetts 02115

Bertram Scharf
Institute of Hearing, Speech and Language and Psychology Department (125 NI), Northeastern University, Boston, Massachusetts 02115

(Received 12 September 2002; revised 14 August 2003; accepted 15 August 2003)

A tone usually declines in loudness when preceded by a more intense inducer tone. This phenomenon is called ''loudness recalibration'' or ''induced loudness reduction'' (ILR). The present study investigates how ILR depends on level, loudness, and duration. A 2AFC procedure was used to obtain loudness matches between 2500-Hz comparison tones and 500-Hz test tones at 60 and 70 dB SPL, presented with and without preceding 500-Hz inducer tones. For 200-ms test and comparison tones, the amount of ILR did not depend on inducer level (set at 80 dB SPL and above), but ILR was greater with 200- than with 5-ms inducers, even when both were equally loud. For 5-ms tones, ILR was as great with 5- as with 200-ms inducers and about as great as when test and inducer tones both lasted 200 ms. These results suggest that (1) neither the loudness nor the SPL of the inducer alone governs ILR, and (2) inducer duration must equal or exceed test-tone duration to yield maximal amounts of ILR. Further analysis indicates that the efferent system may be partly responsible for ILR of 200-ms test tones, but is unlikely to account for ILR of 5-ms tones. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1616580]

PACS numbers: 43.66.Cb, 43.66.Mk [MRL] Pages: 2846–2855

I. INTRODUCTION

Generally, the loudness of a moderate-level *test* tone decreases when preceded by a louder *inducer* tone at the same or a near frequency. This process of context-dependent loudness changes has been called *loudness recalibration* (Marks, 1994) or *induced loudness reduction* (ILR, Scharf *et al.*, 2002), a more neutral and descriptive term. Its effect can be quantified as the *amount of ILR*, which is the reduction of the loudness level of the softer tone that is induced by the louder tone. That is, the amount of ILR is measured in loudness-matching experiments as the decrease in SPL of a comparison tone necessary to maintain equal loudness after the inducer tone has been presented.

The present experiment investigates how the amount of ILR depends on the duration and level of the inducer and test tones. Knowledge of the stimulus parameters that govern ILR is important because it contributes to an understanding of how the loudness of environmental sounds may depend on their acoustic context. Such knowledge also clarifies how the outcome of laboratory studies on loudness may depend on the range of stimuli used in an experiment. For example, equal-loudness contours can vary markedly depending on the range of levels presented to the listeners (Gabriel *et al.*, 1997). Likewise, a previous study from this laboratory (Buus *et al.*, 1999) showed that stimulus range could affect measurements of temporal integration of loudness. At moderate levels, the level differences between equally loud short and long tones were larger when they were presented as part of a series that included tones at high levels than when presented only with tones at lower levels.

These range effects may well be due to ILR. Certainly, properties of ILR appear consistent with the range effect obtained in Buus *et al.*'s (1999) measurements of temporal integration of loudness. Mapes-Riordan and Yost (1999) found a higher amount of ILR at 70 and 60 dB SPL than at 40 dB SPL and threshold. This is consistent with Schneider and Parker's (1990) finding that loudness functions derived from nonmetric scaling of loudness-interval comparisons revealed

[a)]Portions of this work were presented at the 141st Meeting of the Acoustical Society of America in June 2001 [Nieder *et al.*, J. Acoust Soc. Am. **109**, 2349(A) (2001)] and at the 25th Annual Midwinter Research Meeting of the Association for Research in Otolaryngology in January 2002 [Nieder *et al.*, Abs. 25th Ann. Mid-Winter Mtng. Assoc. Res. Otolaryngol., p. 194 (2002)].

[b)]Electronic mail: baerbelnieder@web.de

moderate-level tones to be softer when presented only with higher-level tones than when they were presented only together with lower-level tones. These results indicate that ILR reduces the slope of the loudness function (on log-log scales of loudness and intensity). Accordingly, the increase in level required to compensate for the reduction in loudness caused by the shortened duration of a brief moderate-level tone must be larger when the range of test levels includes high levels than when it does not, which agrees with the range effect observed by Buus *et al.* (1999).

The auditory processes that underlie ILR have yet to be uncovered, but efferent effects demonstrated in physiological studies are reminiscent of ILR. The basilar-membrane and auditory-nerve responses can be reduced by electrical stimulation of the medial efferents (Guinan and Stankovic, 1996; Russell and Murugasu, 1997). Moreover, Guinan and Stankovic (1996) showed that efferent stimulation had the largest effect (as measured by the equivalent reduction in stimulus level) at moderate levels and often reduced the slope of rate-intensity functions. These effects are similar to effects of the top-down gain control hypothesized by Parker and Schneider (1994) and may indicate that the efferent system is at least partly responsible for ILR.

When Marks (1988) discovered ILR in magnitude-estimation experiments, he suspected that it might be caused by context-dependent changes in judgments, but subsequent experiments made that seem unlikely. Analogous experiments on pitch and duration did not show such context effects (Marks, 1992). Moreover, ILR is frequency selective (Marks and Warner, 1991; Marks, 1994). High-level inducer tones appear to reduce the loudness of lower-level test tones only when the frequencies of the inducer and test tones are similar (see Scharf *et al.*, 2002). Finally, ILR can be demonstrated by loudness-matching experiments (e.g., Mapes-Riordan and Yost, 1999) and by loudness functions derived from loudness-interval comparisons (Schneider and Parker, 1990; Parker and Schneider, 1994). Taken together, these studies indicate that ILR is not caused by judgment bias, but is a sensory effect that reflects properties of auditory intensity coding.

The present experiment seeks to examine the mechanisms that might underlie context-dependent intensity coding in the auditory system and to provide a basis for understanding how ILR may affect measurements of temporal integration of loudness. To this end, we conducted two experiments that investigate how the level, loudness, and duration of the inducer and the duration of the test tone affect the amount of ILR. To our knowledge, these questions have not been addressed before. Experiment 1 used 200-ms test tones to measure the amount of ILR obtained with different inducer levels and durations. Experiment 2 used 5-ms test tones to examine the importance of the relative durations of the inducer and the test tones. Using both short and long tones may reveal the extent to which ILR depends on the efferent system, because tones shorter than 25 ms hardly activate the efferent neurons (Liberman and Brown, 1986).

II. EXPERIMENT 1: EFFECT OF INDUCER LEVEL AND DURATION ON 200-MS TEST TONES

A. Method

1. Overview

Listeners matched the loudness of 500-Hz test tones and 2500-Hz comparison tones in a two-interval, two-alternative forced-choice paradigm with interleaved adaptive tracks. In the baseline condition, only the test and comparison tones were presented. In the ILR condition, test tones were preceded by a more intense 500-Hz inducer. In both conditions, the listener indicated whether the test or the comparison tone was louder; the level of the comparison tone was varied according to the listener's response to determine the point of equal loudness. The amount of ILR was taken as the loudness-matched level of the 2500-Hz comparison tone in the baseline condition minus its loudness-matched level in the ILR condition. Thus, positive values indicate that the inducer caused a reduction in the loudness of the test tone—i.e., that ILR took place.

2. Stimuli

The test tones and inducers always had a frequency of 500 Hz; the comparison tones had a frequency of 2500 Hz. These frequencies have been used in most ILR experiments and were chosen to ensure that the inducer would have little effect on the loudness of the comparison tones (cf. Marks and Warner, 1991). Both test and comparison tones lasted 200 ms. The inducers lasted either 5, 200, or 500 ms. The shorter inducer durations were chosen to match the durations used by Buus *et al.* (1999). The 500-ms inducer duration was chosen to be comparable to those used in previous ILR studies. The 5-ms inducers were presented at 80, 95, and 110 dB SPL. The 200-ms inducers were presented at 80 and 95 dB SPL, and the 500-ms inducers were presented only at 80 dB SPL.

All durations are stated as equivalent rectangular durations (ERD). Because the ERD of the 6.67-ms raised-cosine rise and fall is equal to 5 ms, the 5-ms stimuli consisted only of the rise and fall, whereas the 200-ms stimuli had a 195-ms steady-state portion and the 500-ms tones had a 495-ms steady-state portion. The durations measured between the half-amplitude points were 1.67 ms longer than those stated. These envelope shapes ensured that almost all the energy of the test and inducer tones was contained within the 108-Hz-wide critical band centered around 500 Hz (cf. Buus, 1997). Even for the 5-ms tone burst, the energy within the critical band was only 0.8 dB less than the overall energy.

At the beginning of each ILR block, 12 inducer tones were presented separated by 210-ms intervals of silence (see Fig. 1) in an attempt to accelerate ILR.[1] The last inducer was followed by 710 ms of silence after which the test and comparison tones were presented with a 560-ms interstimulus interval. The listener's response initiated the next trial after a 110-ms delay. In subsequent trials, the silent interval between the single inducer and the test tone was 1410 ms; otherwise, the timing was kept the same as in the first trial. The relatively long pause between the inducer and the test tone was chosen to help listeners focus on the loudness of the

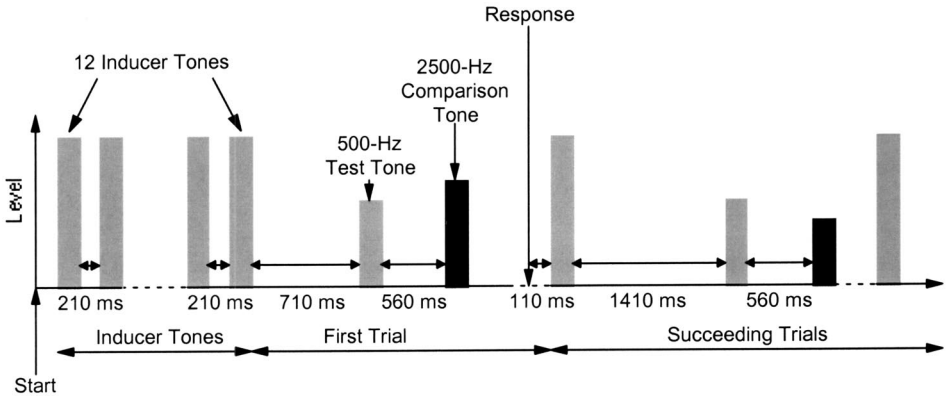


FIG. 1. Time sequence of ILR trials. Each block of ILR trials started with 12 inducers after which the test and comparison tones were presented. The response of the listener initiated a new trial after an 110-ms delay. In the subsequent trials, the delay between the ILR and the test tone was longer than in the first trial. An identical time sequence was used for the baseline trials, but the inducer was replaced by 200 ms of silence.

test tone and to ensure that ILR would not be diminished by having too short an interval. Arieh and Marks (2003) report that the amount of ILR increases with increasing interval up to 670 ms. The timing of the baseline condition was identical to that of the ILR condition, with inducer tones replaced by 200-ms intervals of silence.

3. Procedure

Each trial comprised two observation intervals. The 500-Hz test tone was always presented in the first interval, at either 60 or 70 dB SPL. The 2500-Hz comparison tone was presented in the second interval. This order of the tones was chosen to ensure that the interval between the inducer and test tones in the ILR condition would not vary, because the size of the interval may have a marked effect on the amount of ILR (Mapes-Riordan and Yost, 1998). Of course, because the test tone always preceded the comparison tone, an order effect may have skewed the loudness matches (e.g., Békésy, 1930). Given that the order was the same in the baseline and ILR conditions, any such bias would be common to both conditions. Hence, the amount of ILR, which is the difference between the matching comparison-tone levels obtained in the two conditions, would have been only minimally affected.

The level of the comparison tone was adjusted according to a simple up-down method (Jesteadt, 1980). If the listener indicated that the comparison tone was the louder one, its level was reduced; otherwise it was increased. The step size was 5 dB until the second reversal after which it was reduced to 2 dB. This procedure converges at the level corresponding to the 50% point on the psychometric function (Levitt, 1971). Each block of trials comprised four interleaved adaptive tracks, two for the 60-dB test tone and two for the 70-dB test tone. On each trial, the track for that trial was selected at random among those that had not yet ended. A track ended after nine reversals. The equal-loudness level for each track was calculated as the average of the last four reversals. This procedure with interleaved tracks was chosen to force the listener to compare only the two stimuli presented in each trial. The variation in test-tone levels across tracks ensures that across-trial comparisons are not useful [for further discussion, see Buus *et al.* (1997, 1998)].

Note that only the level of the comparison tone was varied, which is likely to bias the loudness matches somewhat because listeners tend to respond in a manner that causes the level of a variable tone to migrate towards moderate loudness (e.g., Scharf, 1959; Stevens and Greenbaum, 1966; Florentine *et al.*, 1996). Whereas this bias can be reduced by using a roving-level procedure that varies the level of both test and comparison tones (Buus *et al.*, 1998, 1999), there is a considerable risk that such a procedure would result in the presentation of high-level tones at the test frequency, thereby reducing the loudness of lower-level test tones even in the baseline condition. Avoiding contamination of the baseline measurements seemed more important than avoiding the bias caused by the possible migration of the comparison level towards moderate loudness. After all, the bias is common to the baseline and ILR conditions, and it is their difference that is taken as the amount of ILR.

The starting levels of the comparison tones were determined from pilot data. Two blocks of the baseline condition and two blocks of the ILR condition with a 200-ms inducer at 80 dB SPL were tested for each listener at the beginning of the experiment. For these blocks, the starting levels were 30 and 60 dB SPL for the two tracks with 60-dB test tones and 40 and 70 dB SPL for the two tracks with 70-dB test tones. For all further blocks, the starting levels were chosen for each listener on the basis of these pilot data. For each test-tone level, one track started 15 dB above and another started 15 dB below the average comparison level obtained in the initial baseline and ILR blocks. This allowed us to present the same starting levels for the baseline and the various ILR conditions. If the nominal starting level exceeded 85 dB, the starting level was set to 85 dB SPL.

To assess the consistency within a session, most listeners first completed two blocks of baseline matches and then two blocks of ILR matches (yielding a total of four tracks for each test tone level) within a single session lasting approximately 25 min. The inducer was always the same for the two ILR blocks within a session. The ILR conditions were presented in mixed order and each was tested in two sessions. Therefore, the loudness-matched level of the 2500-Hz comparison tone for each listener, test-tone level, and condition was generally based on a total of eight tracks. Successive sessions were separated by at least 3 h (typically 1 or more

days) to ensure that ILR induced in one session had minimal influence on the results obtained in the next session.

4. Apparatus

A PC-compatible computer with a signal processor (TDT AP2) generated the stimuli, recorded the listeners' responses, and executed the adaptive procedure. The sample rate was 48 kHz. The output of the 16-bit D/A converter (TDT DD1) was attenuated (TDT PA4), low-pass filtered (TDT FT5, $f_c = 20$ kHz, 135 dB/oct), attenuated again (TDT PA4), and led to a headphone amplifier (TDT HB6), which fed one earphone of the Sony MDR-V6 headset. The listeners were seated in a sound attenuating booth (Acoustic Systems), and the stimuli were presented monaurally to the ear preferred by the listener. For routine calibration, the output of the headphone amplifier was led to a 16-bit A/D converter (TDT DD1) such that the computer could sample the waveform, calculate its spectrum and rms voltage, and display the results before each session. The SPLs reported below presume a frequency-independent output of 116 dB SPL for an input of 1 V rms.

5. Listeners

Twelve listeners, two males and ten females, 18 to 50 years old with a mean of 24 years, were tested in all conditions. One additional listener was enrolled in the study, but was excluded because his baseline matches were unreliable. In successive sessions he set the 2500-Hz comparison tone progressively lower in the baseline loudness matches and eventually had it set 40 dB lower than the 500-Hz tone for equal loudness. All listeners had audiometric thresholds in the tested ear within 15 dB HL (ANSI, 1989) at octave frequencies between 0.25 and 8 kHz with the exception of listener 11, who had a threshold of 25 dB HL at 8 kHz. Some of the listeners had previous experience making equal-loudness judgments. Two are authors. All other listeners were paid for their participation.

6. Data analysis

Pilot data were included in the final results if the starting levels used to obtain them differed less than 3 dB from the starting levels determined from them. An analysis at the end of experiment 1 revealed that the baseline had not recovered after 3 h on two experimental days for one listener. On both days, the baseline was significantly lower in the second session than in the first. Therefore, these sessions were repeated and the data from the two sessions in which the baseline had not recovered were discarded.

To examine the statistical significance of the effects of the different inducers and test-tone levels, a two-way repeated-measures analysis of variance (ANOVA) (level and duration combined in one factor×test-tone level) was performed (DATA DESK 6.0.2, Data Description, Inc., Ithaca, NY, 1997). The dependent variable was the average amount of ILR ($L_{2500\ Hz}$[baseline condition] − $L_{2500\ Hz}$[ILR condition]) for each listener and condition. Scheffé *post hoc* tests for contrast (DATA DESK 6.0.2, 1997) were performed when appropriate to explore sources of significant effects and interactions. For all statistical tests, differences were considered significant when $p \leqslant 0.05$.

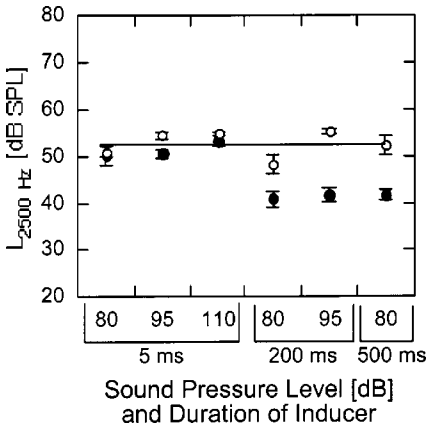


FIG. 2. Loudness matches by a typical listener. The level of the 2500-Hz comparison tone necessary to match the loudness of 500-Hz test tones at 70 dB SPL is shown for the baseline (open circles) and various ILR conditions (filled circles) as indicated on the *x* axis. Each point is the average of eight adaptive loudness matches. The horizontal line represents the average level of the 2500-Hz comparison tone in the baseline condition. The vertical bars show plusminus one standard error of the mean.

B. Results

Figure 2 shows loudness matches by a typical listener for each level and duration of the inducer.[2] The data are shown for test tones at 70 dB SPL. Each symbol represents the average level (across eight tracks) at which the variable 2500-Hz tone was judged to be as loud as the 500-Hz test tone. Open circles are for the baseline condition and filled circles for the ILR conditions. The horizontal line indicates the mean baseline across all conditions. The deviations of the open circles from the horizontal line indicate the listener's variability across sessions. To minimize the effect of day-to-day variations in the listeners' loudness judgments, the amount of ILR is calculated as the difference (in dB) between baseline and ILR loudness matches averaged across the two sessions in which the given condition was tested. In other words, the vertical distance between the open and the filled circles indicates the amount of ILR for the condition indicated on the *x* axis. Generally, ILR is apparent in all conditions, but it is small for the 5-ms inducers.

The mean amounts of ILR for all 12 listeners are plotted against the duration and level of the inducer in Fig. 3. Gray bars are for 60-dB test tones and black bars for 70-dB test tones. The standard deviations vary from 1.6 to 5.3 dB, reflecting large differences among listeners. Three main effects of condition are apparent: First, inducers lasting 5 ms produce less ILR than those lasting 200 or 500 ms. Second, for a given inducer duration, the amount of ILR is independent of inducer level. Third, with 200- and 500-ms inducers, the amount of ILR is smaller for 60- than for 70-dB test tones, but with 5-ms inducers the difference is slight. These observations were supported by an ANOVA and Scheffé *post hoc* tests.

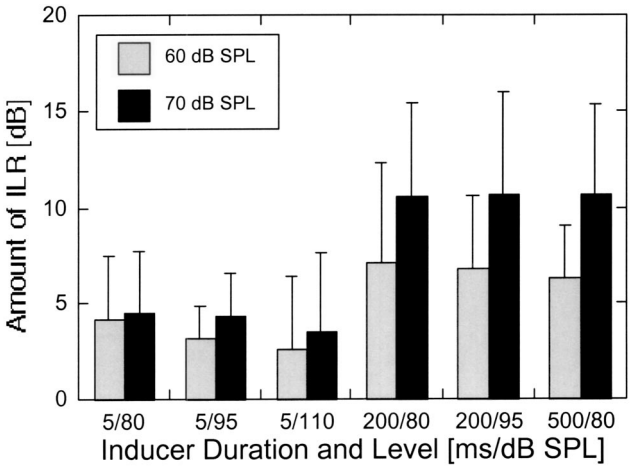


FIG. 3. Average amount of ILR for 12 listeners is plotted against the ILR condition. Gray columns are for test tones at 60 dB SPL and black columns are for test tones at 70 dB SPL. The vertical bars represent the standard deviation across the 12 listeners.

The outcome of the ANOVA is shown in Table I. The Const term in the top line assesses how much the average ILR deviates from zero in relation to the variance calculated across the means for individual listeners. The finding that this term is significant indicates that the overall amount of ILR across all listeners and conditions is greater than zero. The next three lines show that all the effects noted in Fig. 3 were highly significant.

C. Discussion

The present data are in good agreement with data for comparable stimulus conditions published by Mapes-Riordan and Yost (1999). The amount of ILR produced by their 1-s inducer at 80 dB SPL was 11 dB for 500-ms test tones at 60 and 70 dB SPL. The amount of ILR obtained with the present 200-ms inducer, also at 80 dB, was 7 dB for 60-dB test tones and 11 dB for 70-dB test tones. The smaller amount of ILR for the 60-dB test tones can be accounted for by a procedural difference. Whereas Mapes-Riordan and Yost presented the 60- and 70-dB test tones in separate blocks, we presented them concurrently in the same block. It is likely that presenting 60- and 70-dB test tones in one block with interleaved tracks caused the loudness of the 60-dB test tones to be reduced by ILR produced by the 70-dB test tones, even in the baseline condition. Likewise, higher-level comparison tones probably induce a reduction in the loudness of the lower-level comparison tones. Because the loudness reduction in the baseline condition probably is about the same for the lower-level test and comparison tones, the two effects cancel each other so that no loudness reduction is apparent in the baseline data. Nevertheless, any loudness reduction of the lower-level tones in the baseline condition ought to decrease the amount of ILR that can be obtained on these tones in the ILR condition. Accordingly, the difference between the comparison-tone levels in the baseline and ILR conditions for 60-dB tones in the present experiment is unlikely to indicate the full amount of ILR that would have been obtained had a 60-dB tone been tested separately from the 70-dB test tones.

Schneider and Parker's (1990) experiment also included some stimulus conditions comparable to the present ones and they obtained an ILR similar to that observed for 70-dB tones in the present study. For example, for 500-Hz test tones presented at 68 dB SPL in a series with 90-dB tones, which acted as inducers, their results indicate an ILR of 12 dB.

The present results can also be compared to data from experiments on ipsilaterally induced loudness adaptation (ILA). Scharf (2001) and Scharf *et al.* (2002) have argued that ILA in large part is produced by the same auditory processes as ILR. Like ILR, ILA measures the effect of a stronger sound on the loudness of a weaker one. The usual ILA paradigm is to have listeners judge at regular intervals the loudness of a continuous tone that is incremented intermittently or to which is added intermittently a stronger tone at a different frequency (e.g., Weiler *et al.*, 1981; Canévet *et al.*, 1983; Charron and Botte, 1988). The increment or added tone serves the function of the inducer in ILR experiments.

We consider first the study of Canévet *et al.* (1983) who found that the loudness reduction for a 60-dB test tone at 1000 Hz was greater with the increment at 80 than at 65 dB SPL. Although this finding indicates an effect of inducer level, it does not contradict the present finding that ILR does not increase with inducer level at 80 dB and above. It could mean that ILR increases with inducer level up to about 80 dB SPL and then remains at a plateau at higher levels, at least for test tones at or below 70 dB SPL. When the increment was at 80 dB SPL, Canévet *et al.* (1983) as well as Weiler *et al.* (1981) measured more than a 50% decline in loudness induced by a 5-s, 20-dB increment. According to the standard sone function (e.g., Scharf, 1978), halving the loudness corresponds to a 10-phon reduction in loudness level (i.e., a

TABLE I. Two-way analysis of variance for repeated measures of loudness matching. The dependent variable is the average amount of ILR for each listener and condition. The factors are condition (Cnd; six levels of inducer duration/SPL: 5/80, 5/95, 5/1 10, 200/80, 200/95, 500/80) and test-tone level (TLev, two levels: 60 and 70 dB SPL).

Source	df	Error df	Sums of squares	Mean square	*F*-ratio	Probability
Const	1	11	5571	5571	57.14	⩽0.0001
Cnd	5	55	905.0	181.0	15.07	⩽0.0001
TLev	1	11	206.3	206.3	21.51	0.0007
Cnd*TLev	5	55	93.69	18.74	6.868	⩽0.0001
Total	143		3194			

decrease in the level of the comparison tone by 10 dB), which is similar to the amounts of ILR reported in the various studies mentioned above, including ours.

Still more evidence for the similarity between ILA and ILR is provided by Charron and Botte (1988) who used magnitude estimation to measure the loudness reduction of a continuous 60-phon tone produced by increments (inducers) at levels between 75 and 90 dB. They found that the loudness reduction of a 1000-Hz tone was over 40%. The size of the reduction was the same at all levels as was the ILR obtained with inducer levels at and above 80 dB SPL in the present experiment.

Charron and Botte (1988) also varied the frequency of the test tone and found that, for the most part, the loudness reduction decreased as the frequency of the test tone moved away from that of the inducer tone. This finding is analogous to the frequency selectivity of ILR demonstrated by Marks (1994; Marks and Warner, 1991). However, as the inducer level increased above 75 dB, Charron and Botte (1988) found that the maximum loudness reduction occurred at frequencies above the 1000-Hz inducer frequency. With a 95-dB inducer, the maximum reduction of 50% was reached at a test-tone frequency near 1400 Hz. Whereas this finding does not disagree with the findings of no effect of inducer level above 80 dB SPL for same-frequency test and inducer tones, this shift in the locus of maximum loudness reduction is intriguing and will be considered further in Sec. IV.

The close similarity between ILA and ILR is somewhat surprising given the many differences between the procedure and stimulus durations used to measure them. In Charron and Botte (1988), the inducer lasted 24 s, and the test tone continued for 40 s after the end of the inducer; in Canévet *et al.* (1983) the inducer lasted 1 or 5 s and the test tone continued for another 29 or 25 s; in both studies, no silent interval separated the test tones from the inducer, which was presented only once (Charron and Botte, 1988) or six times (Canévet *et al.*, 1983). In the present experiments, the inducer and test tones lasted 200 ms, a silent interval of nearly 1.5 s separated them, and they were repeated many times. Apparently, ILR is a robust phenomenon that occurs under a variety of stimulus configurations.

Nevertheless, some stimulus configurations produce relatively little ILR as evidenced by the small amounts of ILR obtained with 5-ms inducers in the present study. Even when this brief inducer is as loud as a 200-ms inducer, it produces much less ILR on the 200-ms test tone. Similarly, Canévet *et al.* (1983) found that a 5-s increment reduced loudness more than a 1-s increment. This finding may seem at variance with our finding of no difference between 200- and 500-ms inducers, but it is possible that the critical variable is the relation between the durations of the test and inducer tones. The results of both studies can be summarized as showing that loudness reduction increases with inducer duration as long as it is less than the test-tone duration.

III. EXPERIMENT 2: EFFECT OF INDUCER DURATION ON 5-MS TEST TONES

To investigate the importance of the relative durations of the inducer and test tones, experiment 2 examined whether the 5-ms inducers yielded little ILR because they were so brief in absolute terms or because they were so brief relative to the 200-ms test tones. To this end, the influence of 5- and 200-ms inducers on the loudness of 5-ms test tones was investigated. Listeners can readily judge the loudness of such short tones as indicated by the consistent loudness balance judgments obtained in measurements of temporal integration of loudness at 500, 1000, and 4000 Hz (Buus *et al.*, 1999).

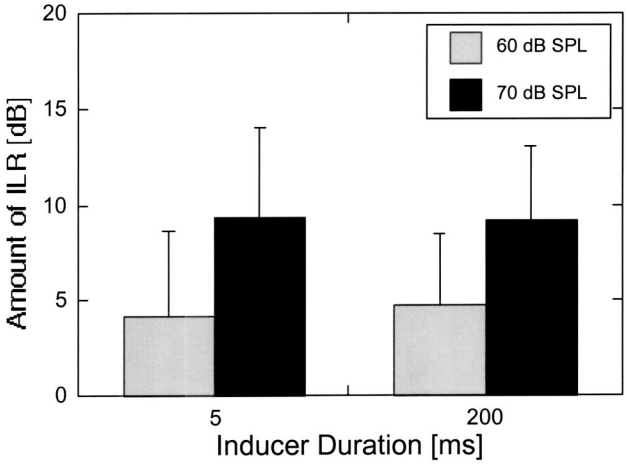


FIG. 4. Average amounts of ILR for 12 listeners tested with 5-ms test tones. The data are plotted in the same manner as Fig. 3.

A. Method

The procedure, stimuli, and apparatus were identical to those in experiment 1 except for the following three changes. First, the test and comparison tones both lasted 5 ms. Second, only 5- and 200-ms inducers at 80 dB SPL were used. Third, the trial timing was changed slightly to make the intervals multiples of 100 ms. At the beginning of each ILR block, 12 inducers were presented separated by 200-ms intervals of silence. The last inducer was followed by 700 ms of silence. The test and comparison tones were separated by an interstimulus interval of 600 ms. The listener's response initiated the next trial after 100 ms. In trials subsequent to the first, the silent interval between the inducer and the test tone was 1400 ms. As in experiment 1, the timing of the baseline condition was identical to that of the ILR condition, with the inducers replaced by 200-ms silent intervals.

Twelve listeners participated in experiment 2. Eleven had also participated in experiment 1. All listeners had thresholds at octave frequencies between 0.25 and 8 kHz that were within 15 dB HL (ANSI, 1989) in the test ear.

B. Results

Figure 4 shows the average amount of ILR for the 12 listeners plotted against the ILR condition. The amount of ILR is shown by gray bars for the 60-dB test tones and by black bars for the 70-dB test tones. The average amount of ILR is smaller for 60- than for 70-dB test tones and is about the same whether the inducer lasts 5 or 200 ms. A two-way ANOVA (inducer duration×test-tone level) for repeated

measures showed that the amount of ILR was significantly larger than zero in all conditions [$F(1,11)=50.1$, $p \leqslant 0.0001$]. The amount of ILR was significantly larger for the 70- than for the 60-dB test tones [$F(1,11)=38.1$, $p \leqslant 0.0001$]. Scheffé *post hoc* tests confirmed significant differences between the amounts of ILR for 60- and 70-dB test tones for both inducer durations. Neither the effect of inducer duration nor the interaction of inducer duration and test-tone level was significant.

C. Discussion

The data from experiment 2 show that a 5-ms inducer reduces the loudness of a 5-ms test tone as much as does a 200-ms inducer. Clearly, then, an inducer lasting only 5 ms is not too brief to give rise to ILR. Rather, it appears that ILR is small when the duration of the inducer is much less than that of the test tone as shown by the 3–4 dB ILR obtained for 5-ms inducers and 200-ms test tones in experiment I and the results obtained by Canévet *et al.* (1983) as noted in Sec. II C for experiment I.

Data relevant to the question of inducer duration may also be culled from measurements of loudness enhancement, the apparent increase in the loudness of a tone burst that follows an inducer with a delay of not more than 100 ms. Scharf *et al.* (2002) proposed that loudness enhancement arises from induced loudness reduction. Enhancement occurs when an intense tone is presented in temporal vicinity of the test tone because the intense tone reduces the loudness of the same-frequency comparison tone while leaving the loudness of the temporally closer test tone unchanged. On the basis of this interpretation of enhancement, we can infer from various enhancement studies that ILR is considerable for inducers and test tones lasting 20 ms (Elmasian and Galambos, 1975), 10 ms (Zwislocki and Sokolich, 1974), and even 0.2 ms (Bauer *et al.*, 1975). Although the inducers were brief in all these studies, they apparently were able to produce considerable ILR on the equally brief comparison tones. Moreover, Zwislocki and Sokolich found that inducers lasting 50 and 10 ms were equally effective in changing the loudness of a 10-ms tone. These results reinforce the conclusion that substantial ILR can be obtained as long as the inducer duration is equal to or greater than the test-tone duration.

Results of measurements of ILA, which—as suggested in the discussion of experiment 1—resembles ILR in many ways, provide further hints about the effect of inducer duration on loudness reduction. At least two studies show that ILA does not increase when the inducer duration is lengthened beyond that of the test tone. Scharf *et al.* (1983) measured the same amount of ILA for a 50-dB, 5-s tone accompanied by increments/inducers lasting 5, 15, or 25 s. At much longer durations, Botte and Mönikheim (1994) found the same amount of loudness reduction whether their 65-dB inducer lasted 5 or 25 min, or was composed of five successive 5-min tones. Altogether the results suggest that ILR (and ILA) varies little as a function of inducer duration provided the inducer is no briefer than the test tone, but decreases when the inducer is made briefer than the test tone.

IV. GENERAL DISCUSSION

The final discussion addresses several issues. The first section summarizes how ILR depends on level and duration of the inducer tone. This summary seeks to combine the results from the present experiments and others on ILR with results from experiments on ILA and temporary loudness shift. The next section considers the likely mechanisms underlying ILR. The final section seeks to clarify how ILR may have affected previous measurements of temporal integration of loudness obtained with a roving-level procedure.

A. Stimulus variables that govern ILR

The effect of level is complex because it involves the level of the inducer, the level of the test tone, and their relative values. With respect to the relative levels of the inducer and test tone, ILR is generally absent when the difference is small—less than approximately 5 dB (e.g., Mapes-Riordan and Yost, 1999). As for the inducer level, it is also unclear whether level is most appropriately measured as SPL, sensation level, or loudness level—although experiments 1 and 2 indicate that the loudness of the inducer is unlikely to be an important factor. In experiment 1, the amount of ILR for 200-ms test tones did not increase with the loudness (and SPL) of 200-ms inducers, which produced much more ILR than 5-ms inducers that were as loud as or louder than the longer inducers. In experiment 2, the amount of ILR for 5-ms test tones was about the same for 5- and 200-ms inducers, even though the brief inducers were considerably softer than the longer ones. These findings suggest that SPL (or perhaps sensation level) is the best measure to use.

Results from experiments on ipsilateral ILA combined with the current results suggest that the amount of ipsilateral ILR for test tones at 60 dB increases when the inducer level goes from 65 to 80 dB (Canévet *et al.*, 1983), but not at higher inducer levels (experiment 1; Charron and Botte, 1988). These findings were obtained with test tones at fixed frequencies. As noted earlier, ILA produced with increments at levels above 80 dB SPL becomes greater as the test-tone frequency increases, reaching a maximum at a frequency separation of about half an octave; moreover, the maximum ILR increases with inducer level (Charron and Botte, 1988).

This effect probably occurs because the maximum of the excitation pattern of a fixed-frequency tone migrates toward the base as the level increases (e.g., Ruggero *et al.*, 1997). Accordingly, the maximal effect of a high-level inducer ought to occur at a place whose best frequency (as measured with lower-level tones) is somewhat above the inducer frequency. This indicates that the level-independent ILR obtained in the present experiment is likely to result from a tradeoff between two factors. One is that the maximal ILR across all possible test frequencies increases with inducer level. The other is that the relative efficiency of high-level inducers at the test frequency decreases with level because the distance between the places of maximal excitation for the test and inducer tones increases with inducer level. Perhaps these two factors counteract one another to produce a constant ILR for test tones at the inducer frequency.

The effects of stimulus timing on ILR are clear. Stimulus duration has little effect on ILR provided the inducer is at

least as long as the test tone. This rule holds over a wide range of durations from the order of milliseconds to minutes. Once the inducer duration is shorter than the test-tone duration, the amount of ILR diminishes. The duration of the interval separating the test tone from the inducer also affects ILR. This interval was kept constant in the present study, but other studies suggest that ILR does not begin until the interval is at least a few hundred milliseconds long and then may increase with interval duration up to around 700 ms (Arieh and Marks, 2003; see also Mapes-Riordan and Yost, 1998; for further discussion, see Scharf *et al.*, 2002).

It should be noted that the interval probably does not need to be silent. One set of unpublished data showed no change in ILA whether a 50-dB, 10-s test tone was presented immediately after a 70-dB, 10-s inducer/increment or was delayed 1 or 5 s. This finding indicates that the loudness reduction is likely to depend on the time lapsed between the offset of the inducer and the listener's loudness judgment, which was not requested until several seconds after the offset of the inducer. In ILR, however, the listener is asked to judge a relatively short tone burst and, therefore, must judge the stimulus when it occurs. Thus, the relevant time interval is the silent interval between the inducer and the test tone, but the role of the silence probably is limited to ensuring that the listener judges the loudness of the test tone with an appropriate delay relative to the offset of the inducer.

B. Mechanisms of ILR

Several aspects of how the medial efferents affect signal transduction at the level of the cochlea appear to agree with findings on ILR and ILA, and it seems possible that at least part of the loudness reduction is mediated by the efferent system. The medial olivocochlear neurons project to the outer hair cells (Liberman and Brown, 1986) and stimulating these neurons reduces the vibration amplitude of the basilar membrane (Russell and Murugasu, 1997) and auditory-nerve activity (Guinan and Stankovic, 1996). As noted in the Introduction, the effects of stimulating efferent neurons appear to parallel the findings for ILR. Moreover, the great majority of efferent units show stronger responses to binaural stimulation than to monaural stimulation in the preferred ear (Liberman, 1988), which can explain Scharf *et al.*'s (1986) finding of at least partial additivity of ipsilateral and contralateral ILA.

Liberman (1988) reported an efferent aftereffect that may well produce ILR. Ten-minute exposures to relatively intense (94 to 114 dB SPL) broadband noises sensitized the efferent units as indicated by lowered thresholds and increased discharge rates produced by 50-ms tones. This sensitization persisted for several minutes after termination of the noise and would almost certainly result in a diminished afferent response to a tone burst presented during this time. In ILR experiments, the inducer tones could sensitize the efferent system. If so, test tones at a given level would cause more efferent activity in the ILR condition than in the baseline condition in which the efferent system presumably is not sensitized. Consequently, the afferent response to the test tone would be reduced in the ILR condition, which ought to reduce the loudness of the test tone. The ILR produced in this manner appears consistent with the reduced slope of the loudness function observed when ILR is present and is reminiscent of the top-down gain control hypothesized by Parker and Schneider (1994).

Although the similarities between efferent response properties and ILR are enticing, the efferent system is unlikely to be the sole cause of ILR. If it were, one would expect little ILR to occur with 5-ms inducer tones because the efferent neurons respond very weakly to tones with durations less than 25 ms (Liberman and Brown, 1986). Moreover, the response latency of the efferent system typically exceeds 10 ms for tones in quiet, although it may approach 5 ms when preceding noise stimulation has sensitized the efferent system (Liberman, 1988). These latencies are considerably longer than the afferent response latency measured in the same manner. Accordingly, most of the efferent response would occur too late to have a substantial effect on the afferent response (and, presumably, the loudness) of a 5-ms test tone. This indicates that the efferent system is unlikely to be responsible for ILR of 5-ms test tones, even if its properties are well suited for explaining many aspects of ILR for longer test tones. Thus, it appears that other auditory mechanisms must contribute to ILR, especially for brief test tones, but the nature of these mechanisms remains obscure.

The idea that sensitization of the efferent system is primarily responsible for ILR of long tones also provides an explanation for our finding that substantial ILR can only be obtained when the inducer duration equals or exceeds the test-tone duration. Because brief tones are ineffective stimuli for the efferent system, it seems unlikely that brief inducers would be capable of sensitizing it. Accordingly, brief inducers, however intense, would not alter the efferent response to the longer test tones and would be expected to produce little ILR in agreement with the results of experiment 1.

C. Effects of ILR on temporal integration

Two aspects of ILR help explain the range effects observed in recent studies on temporal integration of loudness (Buus *et al.*, 1999). First, ILR appears to reduce the slope of the loudness function for moderate-level tones (Schneider and Parker, 1990; Parker and Schneider, 1994). Consequently, restoring any loudness lost by shortening the duration of a tone requires a larger increase in level when ILR is present than when it is not. This effect increases the level difference between equally loud short and long tones. Second, brief inducers yield large ILR only with brief test tones, whereas long inducers yield substantial ILR with both long and short test tones. Hence, ILR ought to be greater for brief than for long tones when long and short tones are presented at various levels in a roving-level procedure. This follows because the moderate-level brief tones undergo ILR from all the intense test tones—whether brief or long—that occur at random, whereas the long test tones undergo ILR only from the long intense test tones. This effect would increase the loudness ratio between equal-SPL long and short tones at moderate levels, which increases the level difference between equally loud short and long tones regardless of the slope of the loudness function. Altogether, ILR offers a com-

pelling explanation for the range effects observed in experiments that use roving-level procedures to assess temporal integration of loudness.

V. SUMMARY AND CONCLUSIONS

The present experiment examined how induced loudness reduction (ILR) depends on the level, loudness, and duration of the inducer and on the duration of the test tones. Comparison tones at 2500 Hz were matched in loudness to test tones at 500 Hz in the presence (ILR condition) or absence (baseline condition) of more intense inducer tones. The inducer tones, when present, were also at 500 Hz. The amount of ILR was taken as the level difference between the comparison tones in the baseline and ILR conditions. The major conclusions are the following.

(1) ILR of 60- and 70-dB test tones is independent of inducer level at and above 80 dB SPL.
(2) The present results taken together with data from the literature indicate that the inducer-tone duration must equal or exceed the test-tone duration to obtain near-maximal ILR.
(3) Specifically, the present results show that ILR for 200-ms test tones is much greater with 200- and 500-ms inducers than with 5-ms inducers, even if the 5-ms inducers are as loud as or louder than the longer inducers. For 5-ms test tones, 5-ms inducers produce as much ILR as 200-ms inducers and as much ILR as is obtained with long test tones and long inducers.
(4) Persistent sensitization of the efferent system may be at least partly responsible for ILR of long tones, but other mechanisms must be responsible for ILR of brief tones because the ILR obtained with 5-ms test and inducer tones is inconsistent with temporal response properties of the efferent system.

ACKNOWLEDGMENTS

This research was supported by NIH/NIDCD Grant No. R01DC02241. Thanks to the reviewers and the Associate Editor for their helpful comments.

[1]This choice in accord with data of Mapes-Riordan and Yost (1999) and Arieh and Marks (2003), which indicate that ILR takes several presentations of the inducer tone before its effect has developed to its full extent. We arbitrarily chose to present 12 inducer tones at the beginning of each ILR block as this seemed reasonable according to the available data. The listeners were told that 12 louder tones would be presented at the beginning of the ILR condition, but they were not told the purpose of these tones.

[2]The average baseline matches between the 500- and 2500-Hz tones differed widely among individuals, as indicated by standard deviations of 10.0 (for 60-dB test tones) and 8.7 dB (for 70-dB test tones). Although these inter-listener differences are surprisingly large, they are not unreasonable. Equal-loudness contours obtained with headphones by Ross (1967) show ranges of about 11 dB among three listeners for the level difference between 500-Hz tones at about 60 and 70 dB SPL and equally loud tones at 3200 Hz. The corresponding standard deviations are 5.5 and 5.9 dB. Three differences between our study and Ross's study may account for the somewhat larger standard deviations in our data. First, estimates of the standard deviation based on data for only three listeners are uncertain and are often lower than the standard deviations obtained with larger populations of listeners. Second, Ross's data represent averages among several frequency pairs using only matches that passed careful screenings for transitivity and symmetry. Such screening and multi-frequency averaging is likely to reduce anomalies that may exist for a particular frequency pair. Finally, Ross used custom-fitted earphone cushions for each listener and measured the SPL produced at the eardrum in each individual to minimize variability due to variation of the headphone response within and across individuals. Our study did not attempt to control for differences in the response of the headphone within and across listeners. In any event, we have great confidence in our data because the listeners obviously were quite consistent both within and across sessions. The standard errors across repetitions averaged 1.2 dB. In addition, it is noteworthy that no obvious relation is apparent between the comparison level in the baseline matches and the amount of ILR obtained in individual listeners.

ANSI (**1989**). ANSI S3.6-1989, ''Specifications for audiometers'' (American National Standards Institute, New York).
Arieh, Y., and Marks, L. E. (**2003**). ''Time course of loudness recalibration: Implications for loudness enhancement,'' J. Acoust. Soc. Am. **114**, 1550–1556.
Bauer, J. W., Elmasian, R. O., and Galambos, R. (**1975**). ''Loudness enhancement in man. I. Brainstem-evoked response correlates,'' J. Acoust. Soc. Am. **57**, 165–171.
Békésy, G. v. (**1930**). ''Über das Fechnersche Gesetz und seine Bedeutung für die Theorie der akustischen Beobachtungfehler und die Theorie des Hörens'' (''On Fechner's Law and its importance for the theory of acoustic observer bias and the theory of hearing''), Ann. Phys. (Leipzig) **7**, 229–359.
Botte, M.-C., and Mönikheim, S. (**1994**). ''New data on the short-term effects of tone exposure,'' J. Acoust. Soc. Am. **95**, 2598–2605.
Buus, S. (**1997**). ''Auditory masking,'' in *Encyclopedia of Acoustics, Vol. 3*, edited by M. J. Crocker (Wiley, New York), pp. 1427–1445.
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.
Buus, S., Müsch, H., and Florentine, M. (**1998**). ''On loudness at threshold,'' J. Acoust. Soc. Am. **104**, 399–410.
Canévet, G., Scharf, B., and Botte, M.-C. (**1983**). ''Loudness adaptation, when induced, is real,'' Br. J. Audiol. **17**, 49–57.
Charron, S., and Botte, M.-C. (**1988**). ''Frequency selectivity in loudness adaptation and auditory fatigue,'' J. Acoust. Soc. Am. **83**, 178–187.
Elmasian, R., and Galambos, R. (**1975**). ''Loudness enhancement: Monaural, binaural, and dichotic,'' J. Acoust. Soc. Am. **58**, 229–234.
Florentine, M., Buus, S., and Poulsen, T. (**1996**). ''Temporal integration of loudness as a function of level,'' J. Acoust. Soc. Am. **99**, 1633–1644.
Gabriel, B., Kollmeier, B., and Mellert, V. (**1997**). ''Influence of individual listener, measurement room and choice of test-tone levels on the shape of equal-loudness level contours,'' Acta Acust. **83**, 670–683.
Guinan, J. J., and Stankovic, K. M. (**1996**). ''Medial efferent inhibition produces the largest equivalent attenuations at moderate to high sound levels in the cat auditory-nerve fibers,'' J. Acoust. Soc. Am. **100**, 1680–1690.
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.
Liberman, M. C. (**1988**). ''Response properties of cochlear efferent neurons: monaural vs. binaural stimulation and the effects of noise,'' J. Neurophysiol. **60**, 1779–1798.
Liberman, M. C., and Brown, M. C. (**1986**). ''Physiology and anatomy of single olivocochlear neurons in the cat,'' Hear. Res. **24**, 17–36.
Mapes-Riordan, D., and Yost, W. A. (**1998**). ''Temporal properties of loudness recalibration,'' Proc. 16th Int. Cong. Acoust., pp. 2361–2362.
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. (**1988**). ''Magnitude estimation and sensory matching,'' Percept. Psychophys. **16**, 511–525.
Marks, L. E. (**1992**). ''The slippery context effect in psychophysics: Intensive, extensive, and qualitative continua,'' Percept. Psychophys. **51**, 187–198.

Marks, L. E. (**1994**). ''Recalibrating the auditory system: The perception of loudness,'' J. Exp. Psychol. Hum. Percept. Perform. **20**, 382–396.
Marks, L. E., and Warner, E. (**1991**). ''Slippery context effect and critical bands,'' J. Exp. Psychol. Hum. Percept. Perform. **17**, 986–996.
Parker, S., and Schneider, B. (**1994**). ''The stimulus range effect: Evidence for top-down control of sensory intensity in audition,'' Percept. Psychophys. **56**, 1–11.
Ross, S. (**1967**). ''Matching functions and equal-sensation contours for loudness,'' J. Acoust. Soc. Am. **42**, 778–793.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Russell, I. J., and Murugasu, E. (**1997**). ''Medial efferent inhibition suppresses basilar membrane responses to near characteristic frequency tones of moderate to high intensities,'' J. Acoust. Soc. Am. **102**, 1734–1738.
Scharf, B. (**1959**). ''Critical bands and the loudness of complex sounds near threshold,'' J. Acoust. Soc. Am. **33**, 365–370.
Scharf, B. (**1978**). ''Loudness,'' in *Handbook of Perception: Vol. 4. Hearing*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), pp. 187–242.
Scharf, B. (**2001**). ''Sequential effects in loudness,'' in *Fechner Day 2001*, edited by E. Sommerfeld, R. Kompass, and T. Lachmann (Pabst, Berlin), pp. 254–259.
Scharf, B., Botte, M.-C., and Canévet, G. (**1983**). ''Loudness adaptation induced interaurally and monaurally,'' in *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer-Verlag, Berlin), pp. 257–263.
Scharf, B., Buus, S., and Nieder, B. (**2002**). ''Loudness enhancement: Induced loudness reduction in disguise?'' J. Acoust. Soc. Am. **112**, 807–810.
Scharf, B., Canévet, G., and Buus, S. (**1986**). ''Additivity of induced loudness adaptation,'' J. Acoust. Soc. Am. Suppl. 1 **80**, S123.
Schneider, B., and Parker, S. (**1990**). ''Does stimulus context affect loudness or only loudness judgments,'' Percept. Psychophys. **48**, 409–418.
Stevens, S. S., and Greenbaum, H. B. (**1966**). ''Regression effect in psychophysical judgement,'' Percept. Psychophys. **1**, 439–446.
Weiler, E. M., Sandman, D. E., and Pederson, L. M. (**1981**). ''Magnitude estimates of loudness adaption at 60 dB SPL,'' Br. J. Audiol. **15**, 201–204.
Zwislocki, J. J., and Sokolich, W. G. (**1974**). ''On loudness enhancement of a tone burst by a preceding tone burst,'' Percept. Psychophys. **16**, 87–90.

Spatial unmasking of nearby pure-tone targets in a simulated anechoic environment

Norbert Kopčo and Barbara G. Shinn-Cunningham[a)]
Hearing Research Center, Boston University, Boston, Massachusetts 02215

(Received 31 December 2002; revised 12 August 2003; accepted 15 August 2003)

Detection thresholds were measured for different spatial configurations of 500- and 1000-Hz pure-tone targets and broadband maskers. Sources were simulated using individually measured head-related transfer functions (HRTFs) for source positions varying in both azimuth and distance. For the spatial configurations tested, thresholds ranged over 50 dB, primarily as a result of large changes in the target-to-masker ratio (TMR) with changes in target and masker locations. Intersubject differences in both HRTFs and in binaural sensitivity were large; however, the overall pattern of results was similar across subjects. As expected, detection thresholds were generally smaller when the target and masker were separated in azimuth than when they were at the same location. However, in some cases, azimuthal separation of target and masker yielded little change or even a small increase in detection threshold. Significant intersubject differences occurred as a result both of differences in monaural and binaural acoustic cues in the individualized HRTFs and of different binaural contributions to performance. Model predictions captured general trends in the pattern of spatial unmasking. However, subject-specific model predictions did not account for the observed individual differences in performance, even after taking into account individual differences in HRTF measurements and overall binaural sensitivity. These results suggest that individuals differ not only in their overall sensitivity to binaural cues, but also in how their binaural sensitivity varies with the spatial position of (and interaural differences in) the masker. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1616577]

PACS numbers: 43.66.Pn, 43.66.Ba, 43.66.Qp [LRB] Pages: 2856–2870

I. INTRODUCTION AND BACKGROUND

When listening for a target sound in the presence of a masking sound, a listener's ability to detect the target is influenced by the locations of both target and masker. When target and masker are at the same distance, it is generally easier to detect or recognize the target when it is spatially separated from the masker compared to when the target and masker are at the same position. This ''spatial unmasking'' effect has been studied for many types of stimuli, including speech (e.g., see Freyman *et al.*, 1999; Shinn-Cunningham *et al.*, 2001), click-trains (e.g., see Saberi *et al.*, 1991; Good *et al.*, 1997), and tone complexes (e.g., see Kidd *et al.*, 1998).

For broadband noise maskers, spatial unmasking arises primarily from acoustic ''better-ear'' effects (moving a sound source in space changes the levels of the signal reaching the ears of the listener) and ''binaural'' effects. ''Better-ear'' effects lead to unmasking because the target-to-masker ratio (TMR) generally increases at one ear when target and masker are in different directions compared to when they are in the same direction. Binaural unmasking can occur when the interaural time and intensity differences in the target and masker differ.

There have been many studies of how binaural differences affect tone detectability in noise [see Durlach and Colburn (1978) for a review of this classic literature]. However, most of these studies were performed under headphones using interaural differences that do not occur naturally. There are only a few studies that have measured how tone detection is affected by the spatial locations of target and masker (examples include Ebata *et al.*, 1968; Gatehouse, 1987; Santon, 1987; Doll and Hanna, 1995). Moreover, results of these studies are inconsistent, finding spatial unmasking ranging from as little as 7 or 8 dB [Santon (1987) and Doll and Hanna (1995), respectively] to as much as 24 dB (Gatehouse, 1987). These apparent discrepancies may be caused by differences in the spatial configurations tested. However, none of these studies analyzed how the TMR at the ears changed with spatial configuration and did not factor out how better-ear (versus binaural) factors may have contributed to the observed spatial unmasking.

Previous studies of spatial unmasking for pure-tone targets considered sources relatively far from the listener and looked only at unmasking resulting from changes in source direction, ignoring any effects of source distance. For sources more than about a meter from the listener, the only significant effect of changing source distance is a change in signal level that is equal at the two ears. However, changes in source distance for sources within reach of the listener produce changes in signal level that differ at the two ears, resulting in exceptionally large interaural level differences (ILDs; see Brungart and Rabinowitz, 1999; Shinn-Cunningham *et al.*, 2000), even at low frequencies for which ILDs are essentially zero for relatively distant sources. In addition, for near sources, relatively small positional changes

[a)]Author to whom correspondence should be addressed: Department of Cognitive and Neural Systems, Boston University, 677 Beacon St., Room 311, Boston, MA 02215. Electronic mail: shinn@cns.bu.edu

can lead to large changes in the energy of the target and masker reaching the two ears. A few previous studies hint that, in some conditions, binaural performance can be worse than monaural performance using the better ear, particularly when there are large ILDs in the stimuli (e.g., see Bronkhorst and Plomp, 1988; Shinn-Cunningham *et al.*, 2001). Given that large ILDs can arise when sources are within reach of the listener, studies of binaural unmasking for nearby sound sources may shed light on these reports.

The current study examined spatial unmasking of pure tone sources within reach of a listener in a simulated anechoic environment. Individually measured head-related transfer functions (HRTFs) were used to simulate sources. This approach allowed realistic spatial acoustic cues to be presented to the subjects while still allowing detailed analyses of the stimuli reaching the subjects during the experiment. The main goals of the study were to (1) measure how target threshold depends on target and masker azimuth and distance for nearby sources, (2) characterize better-ear effects by analyzing how the TMR varies with the spatial configurations tested, (3) evaluate the binaural contribution to spatial unmasking, particularly for spatial configurations in which large ILDs arise, and (4) investigate the degree to which results can be accounted for by a model of binaural interaction.

II. SPATIAL UNMASKING OF NEARBY PURE TONE TARGETS

A. Methods

1. Subjects

Four graduate students with prior experience in psychoacoustic experiments (including author NK) participated in the study. One subject was female and three were male. Subject ages ranged from 25 to 28 years. All subjects had normal hearing as confirmed by an audiometric screening.

2. HRTF measurement

Individualized HRTF measurements were made with subjects seated in the center of a quiet classroom (rough dimensions of 5×9×3.5 m; broadband T_{60} of approximately 700 ms). Subjects were seated with their heads in a headrest so that their ears were approximately 1.5-m above the floor. Measurements were taken for sources in the right front horizontal plane (at ear height) for all six combinations of azimuths (0°, 45°, 90°) and distances (0.15 m, 1 m) relative to the center of the head (defined as the intersection of the interaural axis and the median plane) as shown in Fig. 1.

The Maximum-Length-Sequence (MLS) technique (e.g., see Vanderkooy, 1994) was used to measure HRTFs. Two identical 32 767-long maximum length sequences were concatenated and presented through a small loudspeaker using a 44.1-kHz sampling rate (details regarding the equipment are described below). The response to the second sequence was recorded.[1] This measurement was repeated ten times and the raw measurements averaged in the time domain. This average response was then used to estimate a 743-ms-long head-related impulse response.

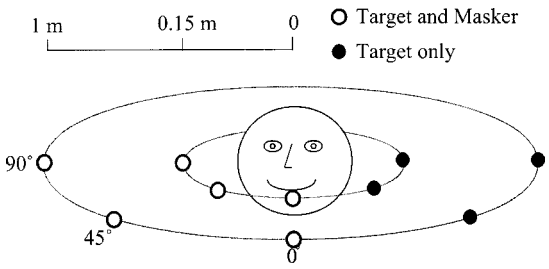


FIG. 1. Spatial positions used in the study. HRTFs were measured at the positions denoted by open symbols. Target detection thresholds were measured for all spatial combination of six masker positions (open symbols) and ten target positions (filled and open symbols; targets simulated at the filled symbols used the corresponding HRTFs from the contralateral hemifield with left- and right-ear signals reversed).

HRTFs were measured using a Tucker-Davis Technologies (TDT) signal processing system under computer control. For each measurement, the concatenated MLS sequence was read from a PC hard-drive and sent to a TDT D/A converter (TDT PD1), which drove a Crown amplifier connected to a BOSE mini-cube loudspeaker. At the start of the measurement session, the subject was positioned so that the center of his/her head was at a location marked on the floor of the room. The subject's head position was read from a Polhemus FastTrak electromagnetic tracker worn on the head to ensure that the center of the head was within 1-cm of the correct location in the room, marked on the floor. The experimenter used other angular and distance markings on the floor to hand-position the loudspeaker to the appropriate azimuth and distance prior to each measurement. Miniature microphones (Knowles FG-3329c) mounted in earplugs and inserted into the entrance of the subjects' ear canals (to produce blocked-meatus HRTF recordings) measured the raw acoustic responses to the MLS sequence. Microphone outputs drove a custom-built microphone amplifier that was connected to a TDT A/D converter (TDT PD1). These raw results were stored in digital form on the computer hard-drive for off-line processing to produce the estimated HRTFs.

No correction for the measurement system transfer function was performed, but the amplitude spectrum of the transfer-function of this measurement system was examined and found to vary by less than 2 dB and to cause no significant interaural distortion for frequencies between 400 and 1500 Hz (the frequency region important for the current study). The useful dynamic range of the measurements (taking into account the ambient acoustic and electrical noise) was at least 50 dB for all frequencies greater than 300 Hz.

HRTFs measured as described above include room echoes and reverberation. To eliminate room effects, time-domain impulse responses were multiplied by a 6-ms-long cos-squared time window (rise/fall time of 1 ms) to exclude all of the reverberant energy while retaining all of the direct-sound energy. The resulting ''pseudo-anechoic'' HRTFs were used to simulate sources (and in all subsequent analyses).

HRTFs were measured only for sources in the right hemifield. To simulate sources in the left hemifield, HRTFs from the corresponding right-hemifield position were used, exchanging the left and right channels (i.e., left/right symmetry was assumed; given that only pure tone targets were

simulated in the left hemifield, this approximation should introduce no significant perceptual artifacts in the simulated stimuli).

The measured HRTFs reflect the radiation characteristics of the loudspeaker used, which is not a uniformly radiating point source. For sources relatively far from the head, any differences in the measurement caused by the directivity of the source should be minor. For sources 15-cm from the center of the head, the effect of the source directivity may be significant. Therefore, the current study focuses on how distance influenced the signals reaching the ears for the particular source used (the Bose loudspeaker in question). The issue of how well the current results may generalize to other nearby sources is considered further in Sec. III, where empirical HRTF measurements are compared with theoretical predictions from a spherical head model that assumes a perfect point source.

In a similar vein, HRTFs measured for sources close to the head are much more sensitive to small displacements in the source (re: the intended source location) than more distant sources. However, given that all acoustic analyses and predictions of performance were made using the same measured HRTFs used to simulate the headphone-presented stimuli, any conclusions regarding which acoustic factors influence performance are justified, even if other measurement techniques might yield slightly different estimates of near-source HRTFs for the positions reported here.

3. Stimulus generation

Target stimuli consisted of 165-ms-long pure tones of 500 or 1000 Hz gated on and off by 30-ms cos-squared ramps. The 500-Hz target frequency was chosen so that results could be compared with previous studies of binaural masking level differences (BMLDs) and spatial unmasking of tones, most of which include a 500-Hz target condition. The 1000-Hz target was included in order to examine what happens for a higher target frequency where target and masker ITDs are still likely to have a large impact on detection but ILDs are larger than at 500 Hz. The target was temporally centered within a broadband, 250-ms-long masker. On each trial, the masker token was randomly chosen from a set of 100 pregenerated samples of broadband noise that were digitally low-pass filtered with a 5000 Hz cutoff frequency (ninth-order Butterworth filter, as implemented in the signal-processing toolbox in Matlab, the Mathworks, Natick, MA).

In most cases, target and masker were simulated as arising from different locations in anechoic space by convolving the stimuli with appropriate individualized head-related impulse responses (time-domain representation of the HRTFs). The simulated spatial configurations included all combinations of target at azimuths ($-90°$, $-45°$, $0°$, $45°$, $90°$) and distances (0.15 m, 1 m) and masker at azimuths ($0°$, $45°$, $90°$) and distances (0.15 m, 1 m). A total of 60 spatial configurations was tested (10 target locations $\times$ 6 masker locations; see Fig. 1). In a subset of trials, traditional BMLDs were measured using the same stimuli without HRTF processing.

For nearby sources, keeping the masker presentation level constant would result in the received level (at the subject's ears) varying widely with masker position. In order to keep the received level of masker relatively constant, the levels of the HRTF-processed masker stimuli were normalized to keep constant the rms energy falling within the equivalent rectangular band (ERB; Moore, 1997) centered on the target frequency at the ear receiving the more intense masker signal (the right ear for all of the tested configurations). In other words, the virtual stimuli actually simulated a masker whose distal energy level was adjusted up or down (depending on the masker spatial location) until the proximal stimulus level was constant at the more intense ear. In our analysis, the amounts by which the distal masker was adjusted were added back to the raw thresholds to predict the amount of spatial unmasking that would have occurred if the distal masker level had been constant.[2]

For the 500-Hz center frequency, the rms levels were adjusted using a 100-Hz-wide ERB. For the 1000-Hz target, the ERB width was set to 136 Hz. The masker signals were preprocessed in Matlab so that the right- (more-intense-) ear rms masker level in the ERB would be 64 dB SPL when played via headphones. BMLDs were measured with the low-pass-filtered noise spectral level fixed at 64 dB SPL.

Stimulus files, generated at a sampling rate of 44.1 kHz, were stored on the hard disk of the control computer (IBM PC compatible). On each trial, appropriate target and masker signals were presented through TDT hardware. Left- and right-ear target and masker signals were processed through four separate D/A converters (TDT PD1). Target signals were scaled to the appropriate presentation level by a programmable attenuator (TDT PA4), summed with the fixed-level masker signals (TDT SM3), and amplified through a headphone buffer (TDT HB6). The resulting binaural stimuli were presented via Etymotic Research ER-1 insert earphones. No filtering was done to compensate for the transfer characteristics of the playback system. A handheld RS 232 terminal (QTERM) was used to gather subject responses and provide feedback.

4. Experimental procedure

Behavioral experiments were performed in a single-walled sound-treated booth.

Each trial consisted of three intervals, each of which contained a noise burst. Either the second or third interval (randomly chosen, with equal probability, on each trial) also contained the tone-burst target. Subjects performed a two-alternative, forced-choice task in which they were asked to identify which interval, the second or third, contained the target tone. Correct-answer feedback was provided at the end of each trial.

A three-down–one-up adaptive procedure was used to estimate detection thresholds (Levitt, 1971), defined as the 79.4% correct point on the psychometric function. Each run started with the target at a clearly detectable level and continued until 11 ''reversals'' occurred. The target level was changed by 4 dB on the first reversal, 2 dB on the second reversal, and 1 dB on all subsequent reversals. For each adaptive run, detection threshold was estimated by taking the average target presentation level over the last six reversals.

TABLE I. Binaural masking level differences for individual subjects. Note that subjects S1 and S3 performed detection experiments for both 500- and 1000-Hz targets; S2 and S4 only performed the experiments for one target frequency (500 and 1000 Hz, respectively). Symbols give the convention used in the figures when plotting individual subject results.

Target frequency	Individual subject results				Across-subject average
	S1 ○	S2 ▽	S3 □	S4 △	
500 Hz	15.6	11.0	14.5	NA	13.7
1000 Hz	13.1	NA	7.5	8.7	9.8

At least three separate runs were performed for each subject in each condition. Final threshold estimates were computed by taking the average threshold across the repeated adaptive threshold estimates. Additional adaptive runs were performed as needed for every subject and condition to ensure that the standard error in this final threshold estimate was less than or equal to 1 dB for each condition and spatial configuration tested.

The study was divided into two parts, one measuring thresholds for the 500-Hz target and one for the 1000-Hz target. Three subjects performed each part (two of the four subjects performed both). For each target, subjects performed multiple sessions consisting of ten runs. Subjects were allowed to take short breaks between runs within one session, with a minimum 4-h break required between sessions. Each subject performed one initial practice session consisting of four practice runs and six runs measuring detection thresholds for NoSo and NoSπ conditions (where NoSo represents a sinusoidal diotic signal, i.e., with zero interaural phase difference, in the presence of a diotic noise; NoSπ represents a sinusoidal signal with interaural phase difference equal to π in the presence of a diotic noise). Subjects then performed 18 additional sessions (180 runs; 3 runs each of every combination for 6 target positions and 10 masker positions). In each of these sessions, a full set of thresholds was determined for one masker position (the order of the ten target positions was randomized within each session). These sessions were grouped into three blocks of six with each block containing a full set of thresholds. The order of masker positions was separately randomized for each block and subject. Any additional runs were performed after completion of the initial 19 sessions. Each subject performed approximately 20 h of testing per target frequency.

B. Results

1. Binaural masking level difference

Table I shows the BMLD (see Durlach and Colburn, 1978), defined as the difference in target detection threshold in the NoSo and NoSπ conditions. Results are consistent with those from previous, similar experiments. BMLDs are larger for the 500-Hz target (where BMLDs ranged from 11 to 16 dB) than the 1000-Hz target (where BMLDs ranged from 7 to 14 dB).

2. Spatial unmasking

The amount of ''spatial unmasking'' is defined as the change in the energy a target emits at threshold for a particular target location compared to when the target is at the same position as the masker. In order to estimate the target detection threshold when the emitted level of the masker is held constant, the amount by which the masker was normalized (to equate the masker level at the more intense ear) was first added back to the raw target detection thresholds. To estimate spatial unmasking (i.e., the amount by which detection thresholds improve with spatial separation of target and masker), the average of all thresholds when target and masker were at the same location was computed and this value was subtracted from all the renormalized thresholds.

Figures 2 and 3 plot the amount of spatial unmasking for 500- and 1000-Hz targets, respectively. Each panel shows the amount of spatial unmasking (improvement in target threshold relative to when target and masker are at the same location) for one masker location (shown graphically in the inset legend in each panel). The abscissa shows the target azimuth. Thick lines and filled symbols show results for the near target; thin lines and open symbols show results for the far target. Symbols show individual subject results and solid lines give the across-subject mean. Dashed lines represent the estimates of the better-ear contribution to spatial unmasking (averaged across subjects), discussed in detail in Sec. IV.

For the spatial configurations tested, the amount of spatial unmasking spans a range of over 50 dB [e.g., compare the thresholds for a 500-Hz target at (0°, 1 m), the center of the thin line in Fig. 2(d), to the thresholds for the 500-Hz target at (90°, 15 cm), the rightmost point of the thick line in Fig. 2(a)]. While subjects generally show similar patterns of results, intersubject differences are large. For instance, in Fig. 2(a) when the masker is at (0°, 1 m) and the 500-Hz target is at 15-cm, subject S1 (filled circles) consistently shows as much as 10 dB more unmasking than the other subjects (other filled symbols). However, this same subject consistently shows the least unmasking in other cases [e.g., in Fig. 2(f) when the masker is at (90°, 15 cm) and the target is at 1-m; compare open circles to the other open symbols].

Despite the large intersubject differences, overall trends are similar across subjects and for both 500- and 1000-Hz targets, and are summarized below.

To a first-order approximation, changing either target or source distance influences spatial unmasking in a straightforward way predicted by a simple change in the stimulus levels at the ears. For instance, looking within any single panel in Fig. 2 or 3 shows that positioning the target near the subject (thick lines) improves target detectability compared to when the target is far from the subject (thin lines; i.e., within any single panel thick lines are grossly similar to thin line results shifted upward by 10–20 dB). Similarly, comparison of the upper panels (a, b, and c) to the lower panels (d, e, and f) shows that positioning the masker near the subject (lower panels) degrades target detectability compared to when the masker is farther from the subject (upper panels; i.e., results in the upper panels are grossly similar to results in the lower panels shifted upward by 10–15 dB). However, closer inspection shows that the detailed pattern of spatial unmasking varies in a more complex way with both target and masker distance than a simple shift in threshold.

Spatial unmasking resulting from a fixed angular separation of target and masker is larger for nearby targets than

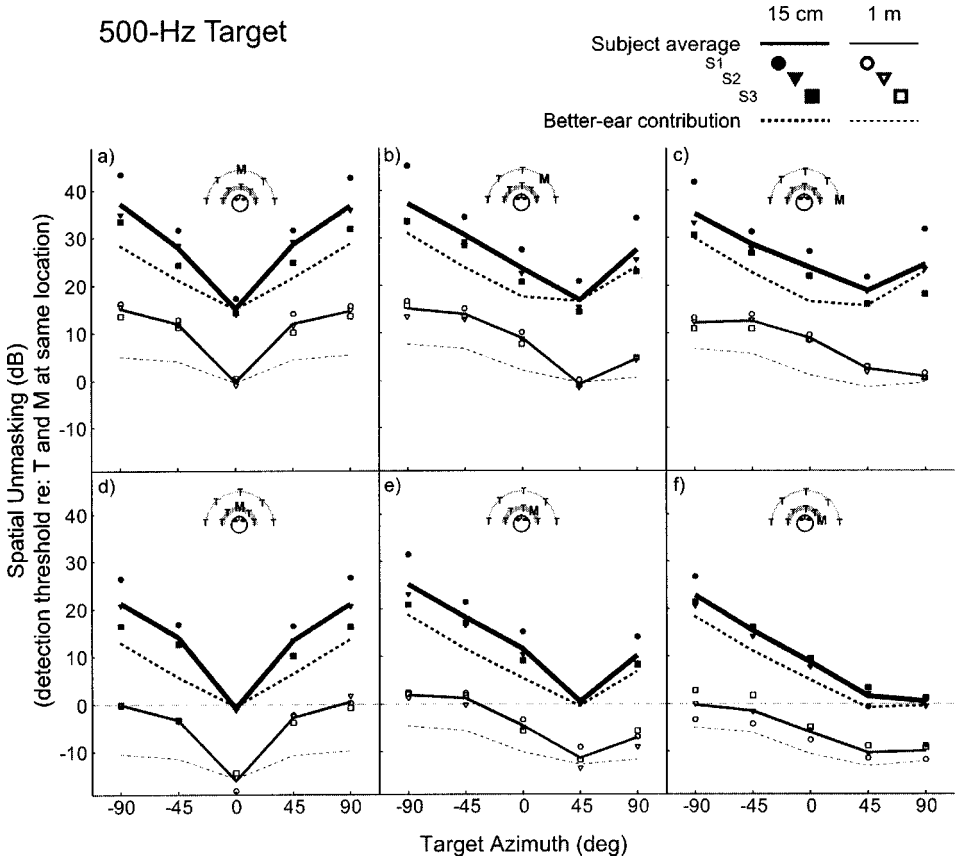


FIG. 2. Spatial unmasking for the 500-Hz target. Each panel plots spatial unmasking (the difference between target detection threshold when target and masker are at the same spatial location and when target and masker are in the spatial configuration denoted in the plot) as a function of target azimuth for a fixed masker location. Across-subject averages are plotted for target distances of 15-cm (thick solid lines) and 1-m (thin solid lines). Individual subject results are plotted as symbols. Dashed lines show the estimated better-ear contribution to spatial unmasking. The spatial configurations of target and masker represented in each panel are denoted in the panel legend.

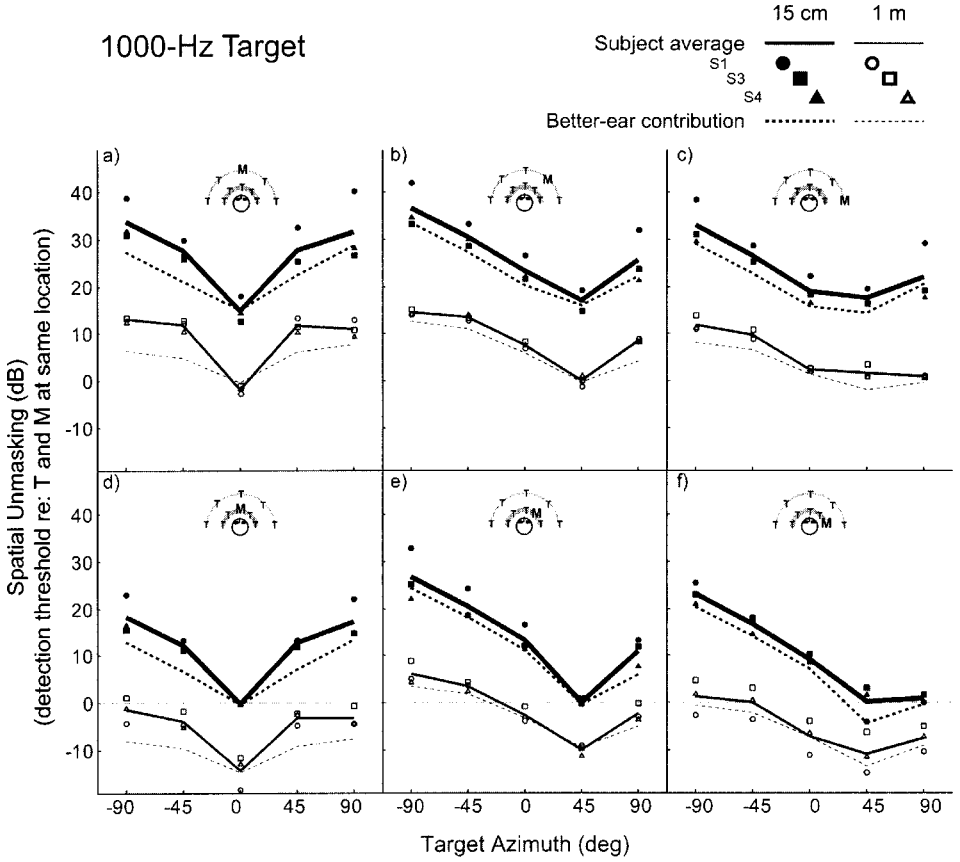


FIG. 3. Spatial unmasking for the 1000-Hz target. See caption for Fig. 2.

for distant targets. For example, in Fig. 3e, the difference between thresholds for the −90° and 45° targets is more than 25 dB for nearby targets (thick line) but less than 20 dB for distant targets (thin line).

Similarly, spatial unmasking resulting from a fixed angular separation of target and masker is larger for nearby maskers than for distant maskers. For example, as discussed above, for a 1000-Hz target when the masker is at (45°, 15 cm) [Fig. 3(e)], spatial unmasking for a 15-cm target (thick line) decreases by more than 25 dB when the target azimuth changes from −90° to +45°. However, when the masker is at (45°, 1 m) [Fig. 3(b)], this same angular displacement of the 15-cm target (thick line) produces a change in spatial unmasking of roughly 20 dB (compare the leftmost point and the point producing the least spatial unmasking, where the target is at 45°).

Angular separation of target and masker can actually make performance worse when target distance differs from masker distance. Usually, separating target and masker in azimuth improves target detectability compared to when the target and masker are in the same direction, but not in every case. When the masker is at 0° (panels a and d in both Figs. 2 and 3) the least amount of spatial unmasking occurs (thresholds are highest) when the target is at 0° (the same direction as the masker); when the masker is at 45° (panels b and e in Figs. 2 and 3) the least unmasking arises when the target is in the 45° masker direction. However, when the masker is at 90° (panels c and f in Figs. 2 and 3), angular separation of target and masker does not always increase the amount of unmasking. Specifically, for a masker at (90°, 1 m) [Figs. 2(c) and 3(c)] there is less spatial unmasking when the 15-cm target (thick line) is at 45° than when it is at 90°. Similarly, for a masker at (90°, 15 cm) [Figs. 2(f) and 3(f)] the amount of spatial unmasking for a 1-m target (thin line) is either equal [500-Hz target; Fig. 2(f)] or greater [1000-Hz target; Fig. 3(f)] when the target is at 90° compared to 45°.

Finally, independent of target or masker distance, the same angular separation of target and masker tends to produce less spatial unmasking as the masker laterality increases. For example, in Fig. 2(d) when the masker is at (0°, 15 cm) and the 500-Hz target is at a distance of 15-cm (thick line), a 90° angular separation of target and masker yields nearly 20 dB of unmasking. However, in Fig. 2(f), when the masker is at (90°, 15 cm) and the target is at 15-cm (thick line), the same angular separation of target and masker produces only 10 dB of unmasking.

C. Discussion

Intersubject differences in spatial unmasking may be partially explained by intersubject differences in the size of the BMLD. For instance, subject S1 has the largest BMLDs and exhibits the most spatial unmasking. However, intersubject differences in spatial unmasking could also be caused by differences in the acoustic parameters in the individually measured HRTFs. Analysis of acoustic differences in the measurements and the binaural contribution to spatial unmasking, which are considered further in Sec. IV, suggest that intersubject differences in spatial unmasking are affected both by subject-specific differences in acoustic cues and in different sensitivities to binaural cues.

Many of the current results follow easily predicted patterns. Moving the target closer to the subject improves detection performance (as expected on the basis of an increase in the level of the target reaching the listener); conversely, moving the masker closer degrades detection performance (as expected when the level of the masker at the ears increases). Separating target and masker in angle improves detection performance for most spatial configurations. However, there are other effects that are less intuitive. Unmasking varies more with target azimuth for a 15-cm masker than for a 1-m masker and for a 15-cm target than for a 1-m target. The masker laterality influences the effectiveness of a given angular separation of target and masker, decreasing with masker laterality. Finally, when target and masker are at different distances and the masker is at 90°, the amount of unmasking can actually decrease when the target is at 45° compared to when the target is in the same direction as the masker (this is essentially a case where there is ''spatial masking,'' i.e., where performance is actually worse when the sources are spatially separated compared to when they are at the same location).

Apparent discrepancies in the amount of spatial unmasking observed in previous studies are actually consistent with the current results. For example, the current study found more spatial unmasking for 1-m sources when the masker is at 0° compared to when the masker is at 90°. Thus, the relatively large amount of spatial unmasking observed by Gatehouse (1987) compared to that found by Santon (1987) and Doll and Hanna (1995) may be caused by the fact that Gatehouse fixed the masker in front of the listener and varied target azimuth, whereas Santon and Doll and Hanna fixed the target in front of the listener and varied masker azimuth.

III. HRTF MEASUREMENTS

The acoustic factors that influence spatial unmasking can be characterized by analysis of the HRTFs used in the simulations. Three acoustic characteristics of the HRTFs influence the performance in a spatial unmasking task: the magnitude spectra of, the interaural level differences (ILDs) in, and the interaural time differences (ITDs) in the signals reaching the two ears. The magnitude spectra of the HRTFs determine the intensity of the sound at the ears and thus the amount of spatial unmasking resulting from better-ear effects. ITDs and ILDs determine the amount of binaural unmasking. In this section, these parameters are analyzed for the individually measured HRTFs.

Individual HRTFs for the four human subjects are compared both to values measured for a KEMAR acoustic manikin (using the same measurement techniques used for the individual subjects) and those predicted from a spherical model of the head assuming a perfect point source. While the literature contains descriptions of both KEMAR (Brungart and Rabinowitz, 1999) and spherical-head model (Duda and Martens, 1998; Shinn-Cunningham *et al.*, 2000) HRTFs for sources near the listener, the current analysis compares these ''generic'' models to human measurements to determine

whether the models capture the acoustic effects that are important for predicting the amount of spatial unmasking as a function of nearby target and masker locations. As noted in Sec. II, the current measurements do not try to compensate for the radiation characteristics of the loudspeaker used; as such, any consistent discrepancies between predictions from a spherical-head model and measured results (from KEMAR and the human subjects) may reflect influences of the radiation characteristics of the loudspeaker used (which is not a point source) or other differences between the assumptions of the spherical-head model and properties of the physical sources and heads measured.

A. Methods

KEMAR HRTFs were measured using a procedure identical to that used for the human listeners (see description in Sec. II). HRTF predictions for a spherical head model (Brungart and Rabinowitz, 1999; Shinn-Cunningham *et al.*, 2000) were computed using a head with radius of 9-cm and diametrically opposed ears. These results are compared to the HRTFs measured for the four subjects who participated in the spatial unmasking experiment.

For all of the HRTFs, the magnitude spectra, ILD, and ITD were determined for the equivalent rectangular band (ERB) centered at a given frequency. Magnitude spectra were calculated as the rms energy in the HRTF falling within each ERB filter (100-Hz width centered at 500 Hz and 136-Hz width centered at 1000 Hz). ILDs were computed as the difference in the magnitude spectra for the left and right ears. ITD was first estimated as a function of frequency by taking the difference between the right- and left-ear HRTF phase angles at each frequency f and dividing by $2\pi f$. The ITD in each ERB filter was then estimated as the average of the ITD values for the frequencies falling within each ERB filter.

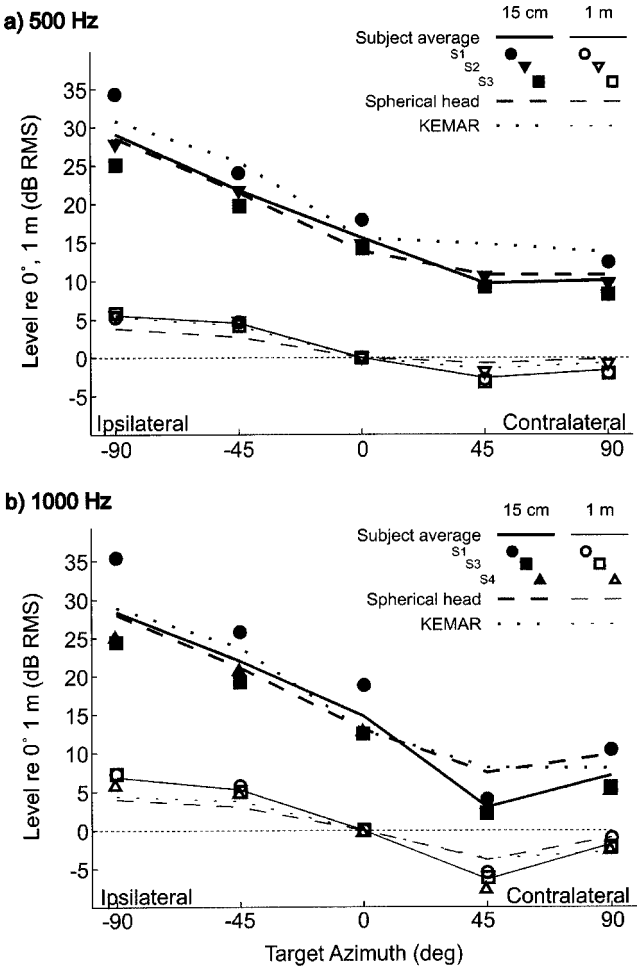


FIG. 4. Left-ear HRTF spectrum levels in ERB filters, relative to the left-ear HRTF for a source at (0°, 1 m). Results are shown for individual listeners, KEMAR, and the spherical head model as a function of source position. (a) 500 Hz. (b) 1000 Hz.

B. Results

1. Intensity effects

Figure 4 shows the magnitude of the ERB-filtered HRTFs at 500 [Fig. 4(a)] and 1000 Hz [Fig. 4(b)] for the left ear relative to a source at (0°, 1 m). (Recall that HRTFs were measured only for sources to the right of the listener and that this analysis assumes left-right symmetry.) Results are shown as a function of the target azimuth for individual human subjects (symbols), the across-human-subject average (solid line), KEMAR (dotted line), and a spherical head model (dashed line). Distant sources are represented by open symbols and thin lines; near sources are shown by filled symbols and thick lines.

Not surprisingly, for both frequencies the spectral gain is larger for near sources (thick lines) than far sources (thin lines). However, in addition to an overall shift in level, the dependence of the HRTF level on source azimuth differs for the two distances. Specifically, for the 15-cm distance (thick lines), the gain to the ipsilateral ear (i.e., the gain for sources at negative azimuths) grows rapidly with source eccentricity compared to the 1-m distance, while the gain to the contralateral ear (positive azimuths) changes similarly with source angle for both distances (compare thick and thin lines).

Overall, intersubject differences are modest for the more distant source (consider the open symbols in each panel). However, there are larger intersubject differences for the 15-cm source positions (filled symbols). For instance, at both frequencies [Figs. 4(a) and (b)], the 15-cm HRTF gain for subject S1 (filled circles) is generally 5–10 dB larger than for the other subjects, except at 45° where all HRTFs are similar.

For a 15-cm source at both 500 Hz [Fig. 4(a)] and 1000 Hz [Fig. 4(b)], KEMAR (thick dotted lines) and spherical-head gains (thick dashed lines) generally fall within the range of values observed for the four human subjects (filled symbols) measured in this study. However, in Fig. 4(b) for a 1-m source, KEMAR measurements (thin dotted lines) and model predictions (thin dashed lines) slightly underestimate the 1000-Hz gain to the ipsilateral ear compared to the individual subject results (lines fall below symbols for azimuths of −45° and −90°). At 500-Hz [Fig. 4(a)], the 1-m KEMAR measurements (thin dotted lines) fall within the range of results obtained from the human subjects (open symbols); however, the spherical head model results (thin dashed lines) fall below the subject measurements (open symbols) for ipsilateral sources (sources at −45° and −90°).

While, intuitively, we expect the level of the signal

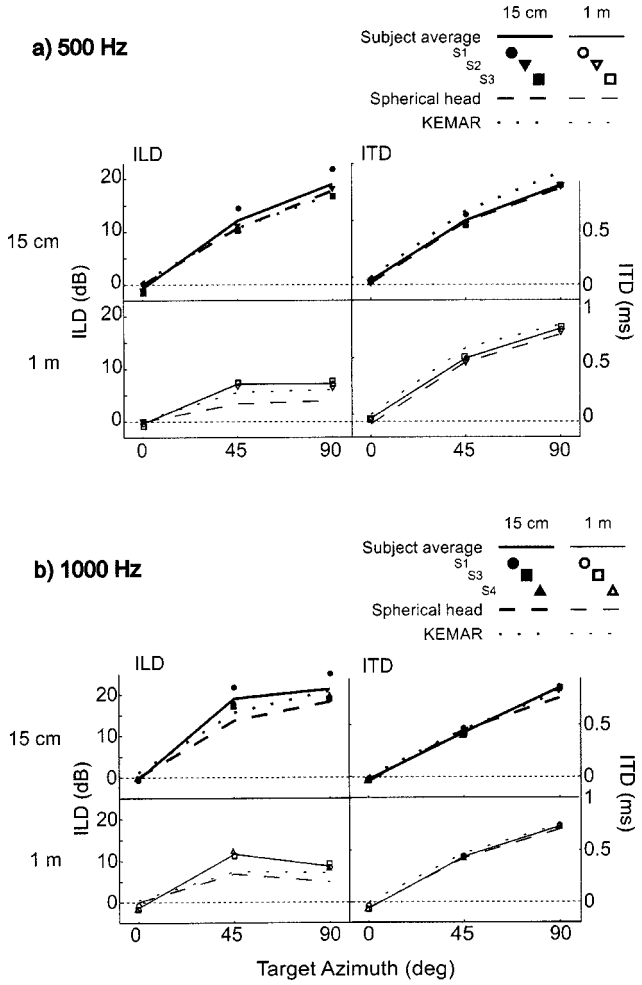


FIG. 5. ILDs and ITDs in HRTFs for individual subjects, KEMAR manikin, and the spherical head model. (a) 500 Hz. (b) 1000 Hz.

reaching the ears to vary monotonically with lateral angle of the source, human HRTF measurements show that this is not strictly true. In particular, the 1000-Hz human measurements [symbols and solid lines in Fig. 4(b)] show that less energy reaches the contralateral ear when a source is at 45° than when it is at 90° for both source distances (thick and thin lines are nonmonotonic with azimuth) Similarly, at 500 Hz [Fig. 4(a)] the gain to the contralateral ear is comparable for 45° and 90° sources rather than decreasing for the 90° source (thick and thin lines). This nonmonotonicity [which may in part be a consequence of the acoustic ''bright spot;'' e.g., see Brungart and Rabinowitz (1999)] is underestimated in both the spherical-head model (dashed lines) and KEMAR (dotted lines) HRTFs, especially at 1000 Hz [compare lines to human subject results for sources at 45°, especially in Fig. 4(b)].

2. Interaural differences

Figure 5 shows the ILDs and ITDs in the measured HRTFs at 500 and 1000 Hz [Figs. 5(a) and (b), respectively] for the spatial positions used in the study. As in Fig. 4, results for individual subjects (symbols), the across-human-subject average (full lines), KEMAR (dotted lines), and a spherical head model (dashed lines) are shown as a function of target azimuth. Results for near sources are shown in the top of each subplot with heavy lines and filled symbols. Thin lines and open symbols plot results for far sources (bottom row of each half of the figure). The left column shows ILD results and the right column shows ITD results.

ILDs were calculated directly from the measurements plotted in Fig. 4. As a result, there are large intersubject differences in the ILDs (left panels in Fig. 5) that are directly related to the intersubject differences in the monaural spectral gains. For instance, subject S1 has much larger ILDs at both 500 and 1000 Hz for the 15-cm source [filled circles in the left columns of Figs. 5(a) and (b)] than any of the other subjects (other filled symbols).

As expected, for both frequencies [Figs. 5(a) and (b)] ILDs are much larger for sources at 15-cm (thick lines in top left panels) compared to 1-m (thin lines in the bottom left panels) with ILDs at 500 and 1000 Hz approaching 20 dB for the nearby sources at 90° (rightmost point in the top left panels). The spherical-head (dashed lines) and KEMAR (dotted lines) results tend to underestimate ILDs for lateral sources, although for the 500-Hz, 15-cm sources [Fig. 5(a), top left panel], both spherical-head and KEMAR results are within the range of human observations. Discrepancies between human and model results are most pronounced for a 1000-Hz source at a distance of 1-m [Fig. 5(b), bottom left panel] and are greater for the spherical-head predictions (dashed lines) than KEMAR measurements (dotted lines).

ITDs [the right panels in Figs. 5(a) and (b)] vary primarily with source angle and change only slightly with distance and frequency. For most of the measured locations, both spherical-head and KEMAR results are in close agreement with human measurements.

C. Discussion

Both spherical-head and KEMAR HRTFs provide reasonable approximations to how acoustic parameters in human HRTFs vary with source location. In general, both KEMAR and the spherical head measurements fall within the range spanned by the individual subject measurements. However, both spherical-head predictions and KEMAR measurements slightly overestimate the gain at the contralateral ear when a source is at 45° (especially at 1000 Hz) and tend to modestly underestimate the ILD for sources off midline, particularly at the 1-m distance. These small differences cannot be attributed to loudspeaker characteristics, given that (1) the discrepancies are similar for both KEMAR measurements (using the same loudspeaker) and spherical-head predictions (assuming a perfect point source) and (2) the differences are, if anything, larger for the more distant, 1-m source (where the loudspeaker directivity is less influential) than the nearby source. Thus, we conclude that generic HRTF models capture the important features of the HRTFs measured in human subjects and that the effects of the source transmission characteristics do not strongly influence the signals reaching the ears even for nearby sources, at least for the frequencies considered in the current study.

Intersubject differences in the HRTFs are large, especially for nearby sources. Of the four subjects, one subject showed consistently larger spectral gains and consistently

larger ILDs than the other subjects when the source was at 15-cm. While it is possible that some of the intersubject differences arise from inaccuracies in HRTF measurement (e.g., from hand-positioning the loudspeaker), the fact that one subject has consistently larger gains and ILDs for all nearby source locations suggests that real anatomical differences rather than measurement errors are responsible for the observed effects. It is also interesting to note that the observed intersubject differences are much smaller for the 1-m source, suggesting that intersubject differences in HRTFs are especially important when considering sources very close to the listener.

IV. BETTER-EAR AND BINAURAL CONTRIBUTIONS TO SPATIAL UNMASKING

A. Analysis

For each subject, estimates of the better-ear and binaural contributions to spatial unmasking were derived from the acoustic parameters of the HRTFs and the behavioral thresholds.

The better-ear contribution to spatial unmasking was estimated by calculating the TMR in the ERB filter centered on the target frequency at the better ear for each spatial configuration when target and masker emit the same level (and thus would yield a TMR of zero when at the same location). The resulting TMR predicts the amount by which target thresholds decrease or increase simply because of acoustic effects at the better ear (i.e., if the calculated TMR is +2 dB, it implies that at detection threshold, the intensity of the target at the better ear was 2 dB more for the given spatial configuration than if the target and masker were at the same spatial location; thus, the better-ear contribution for such a configuration is +2 dB). The subject-specific binaural contribution to spatial unmasking was estimated by subtracting the estimated better-ear contribution to spatial unmasking (derived from individually-measured HRTFs) from the individual behavioral estimates of spatial unmasking.

B. Results

1. Better-ear contributions to spatial unmasking

While intersubject differences in the better-ear contribution to spatial unmasking are large, the trends in the across-subject average data capture the important features of the individual data. For brevity, only the across-subject averages are presented in Figs. 2 and 3 for the 500- and 1000-Hz target, respectively, as dashed lines. For all spatial configurations tested, the behaviorally observed amount of spatial unmasking either equals or is larger than the predicted spatial unmasking from better-ear effects (dashed lines fall below or at measured values in all graphs). Thus, even when there are large ILDs in the signals reaching the listener, binaural performance is always better than or equal to predicted performance when listening monaurally with the acoustically better ear.

Better-ear effects account for a large portion of the observed spatial unmasking when target and masker are in the same direction and for the large influence of target and/or masker distance on spatial unmasking. Specifically, the predicted results (dashed lines) are in good agreement with the measured results when the target is at 0° in the left column, at 45° in the middle column, and at 90° in the right column. Generally, angular separation of target and masker increases the better-ear contribution to unmasking (dashed-line predictions generally increase as the target azimuth moves away from the masker azimuth). However, when the masker is at 90° (the right columns in Figs. 2 and 3), better-ear effects either decrease or are roughly the same when the target is at 45° compared to 90° (dashed-line predictions are either constant or decrease as the target azimuth moves from 90° to 45°). Better-ear contributions to unmasking change more with target azimuth when the target is at 15-cm (thick dashed lines) than at 1-m (thin dashed lines), primarily because, for nearby sources, small positional changes cause large changes in the relative distance from source to the better ear.

Finally, differences between mean subject results (solid lines) and predicted better-ear effects (dashed lines) are generally larger for the 500-Hz target (Fig. 2) than the 1000-Hz target (Fig. 3), suggesting that the better-ear contributions to unmasking are relatively more important (i.e., account for a greater portion of the observed amount of spatial unmasking) for the 1000-Hz target than the 500-Hz target. This is true both because the better-ear effects are larger in absolute terms and because the additional spatial unmasking for which better-ear effects cannot account is smaller at 1000 Hz than at 500 Hz.

2. Binaural contributions to spatial unmasking

Figures 6 and 7 show the estimated binaural contribution to spatial unmasking for the 500- and 1000-Hz target, respectively. The binaural contribution was calculated for each individual subject by subtracting the estimated better-ear contribution (the across-subject average of which is shown by dashed lines in Figs. 2 and 3) from the total amount of spatial unmasking (symbols in Figs. 2 and 3). Both Figs. 6 and 7 show results for each subject who performed that condition in a separate subplot. Each subplot is divided into six panels corresponding to the six masker locations (laid out as indicated in the legend). In each panel, symbols plot the mean binaural contribution to spatial unmasking (averaged across the repeated adaptive runs). The error bars show the *range* of thresholds obtained across the repeated adaptive runs for each condition. Results are shown for both the far target (gray) and the near target (black) as a function of target azimuth. Figures 6 and 7 also show model predictions (lines), which are derived and discussed in Sec. V.

Even though intersubject differences are large, there are a number of trends that are consistent across subjects. Not surprisingly, for both target frequencies (Figs. 6 and 7) there is no unmasking beyond the better-ear contribution when target and masker are at the same spatial location (the binaural gain is near zero when the target is at 0° in the left columns, at 45° in the middle columns, and at 90° in the right columns of Figs. 6 and 7). In fact, only the 500-Hz results for subject S1 [Fig. 6(a)] show any binaural unmasking when target and masker are at the same off-median-plane direction but at different distances. For example, looking at the top right panel of Fig. 6(a) [masker at (90°, 1 m)], the binaural gain is posi-

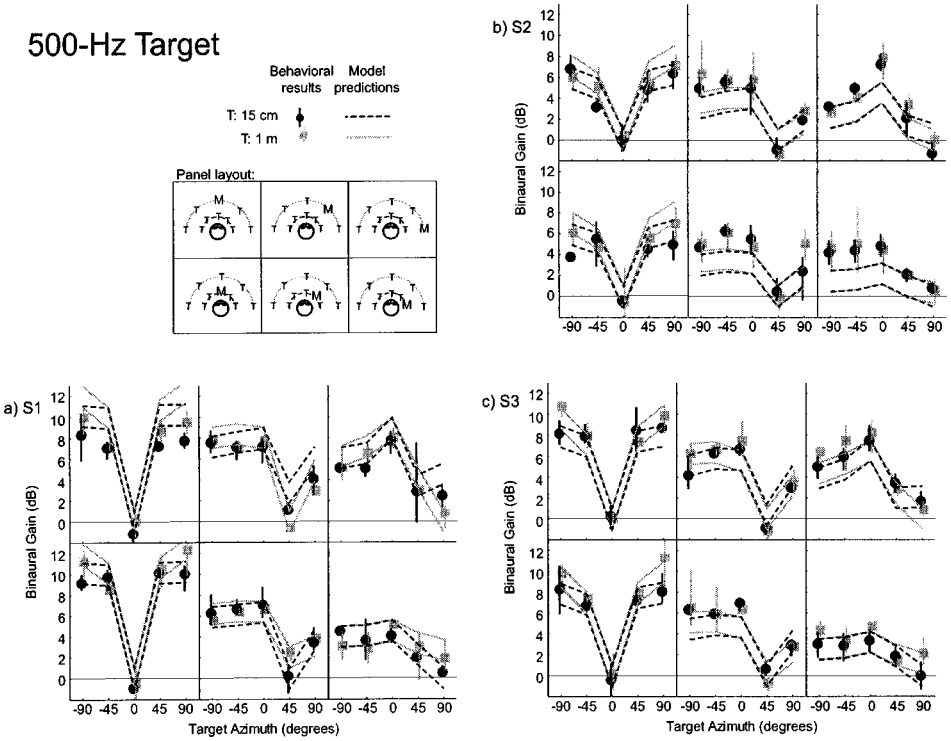


FIG. 6. Estimated binaural contribution to spatial unmasking for the 500-Hz target. Each panel plots the amount of binaural unmasking for one masker position for both the 15-cm and 1-m target. Symbols show estimates for individual subjects with error bars showing the range of results across multiple adaptive runs. Lines trace a 2-dB range around the predicted amount of binaural unmasking from the Colburn (1977a) model for the 15-cm (dashed black lines) and 1-m (solid gray lines) target. The layout of the spatial configurations of target and masker represented in each panel are shown in the legend. (a) Subject S1. (b) Subject S2. (c) Subject S3.

tive when the target is at (90°, 15 cm) (black circle); in the bottom right panel of Fig. 6(a) [masker at (90°, 15 cm)], the binaural gain is positive when the target is at (90°, 1 m) (gray square).

Overall, target distance has relatively little impact on the binaural component of the spatial release from masking (black and gray symbols are generally comparable within each panel). However, masker distance influences results for all subjects, particularly for the 500-Hz results (Fig. 6) when the masker is located at 90° (right panels). In these configurations, binaural unmasking is smaller when the masker is at 15-cm (lower right panel) than when it is at 1-m (upper right panel).

In general, the binaural contribution to spatial unmasking is larger for the 500-Hz target (Fig. 6) than the 1000-Hz target (Fig. 7). For both target frequencies, the amount of binaural unmasking tends to be largest when the masker is at 0° (left panels in each subplot) and decrease as the masker is displaced laterally (center and right panels in each subplot). Similarly, the change in binaural unmasking with target

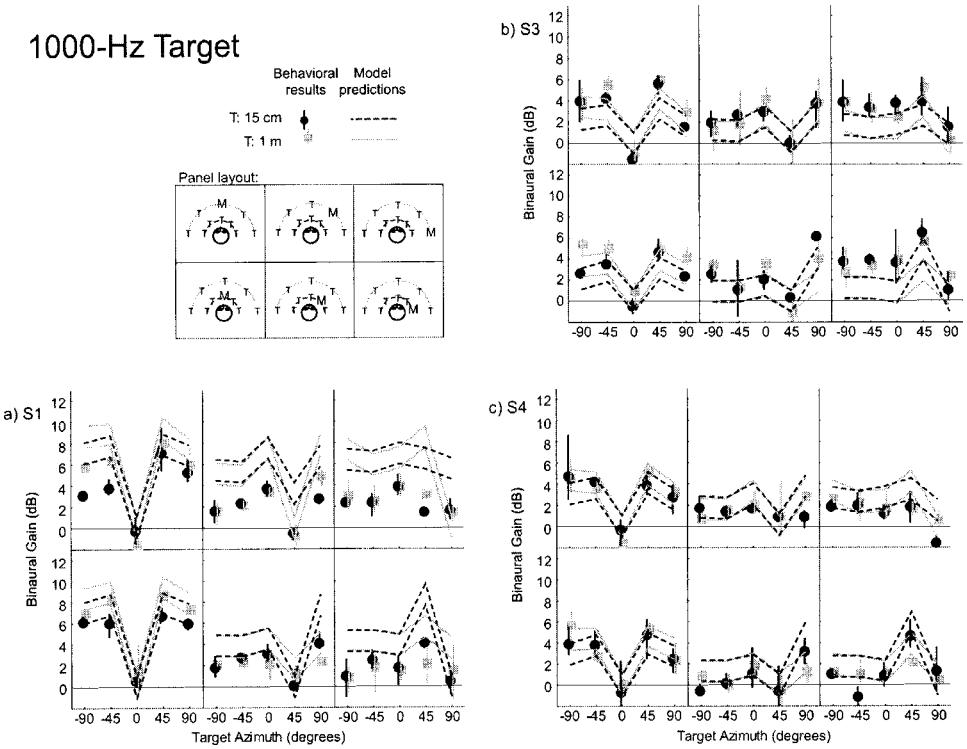


FIG. 7. Estimated binaural contribution to spatial unmasking for the 500-Hz target. See caption for Fig. 6. (a) Subject S1. (b) Subject S3. (c) Subject S4.

angle (i.e., the modulation of binaural gain with target azimuth) is smaller when the masker is laterally displaced (right panels) than when the masker is at 0° (left panels), particularly for the 1000-Hz target (Fig. 7). For instance, looking at the bottom left panel of Fig. 7(a), when the masker is at (0°, 15 cm) the binaural contributions to spatial unmasking for the 1000-Hz target for subject S1 range from 0 to 8 dB depending on the target azimuth. However, when the masker is at (90°, 15 cm) [bottom right panel in Fig. 7(a)], binaural unmasking is roughly constant, independent of target angle (roughly 0–2 dB).

The angular separation of target and masker that leads to the greatest amount of binaural unmasking depends on target frequency. For the 500-Hz target (Fig. 6), binaural unmasking tends to be greatest when target and masker angles differ by about 90° (for example, in the right columns of Fig. 6 where the masker is at 90°, the unmasking is generally greatest when the target is at 0°). However, for the 1000-Hz target (Fig. 7), binaural unmasking tends to be greatest when target and masker angles differ by roughly 45° (in the right columns of Fig. 7 where the masker is at 90°, the amount of unmasking tends to be greatest when the target is at 45°).

C. Discussion

Better-ear factors contribute significantly to spatial unmasking for all of the spatial configurations tested. Better-ear effects are larger at 1000 Hz than 500 Hz and are larger when the target is at 15-cm compared to when the target is at 1-m. The better-ear contribution to spatial unmasking does not always increase monotonically with angular separation of target and masker. In particular, when the masker is at 90°, displacing the target toward the median plane can lead to decreases in the TMR at the better ear, especially if the target and masker are at different distances. This result helps explain why angular separation of target and masker does not always improve detection performance.

Subjects show large differences in their ability to use binaural cues in detection tasks. For subject S1, binaural differences can decrease detection thresholds by as much as 12 dB at 500 Hz [see Fig. 6(a)]; for subject S2 binaural differences provide at most 7 dB of unmasking [Fig. 6(b)]. These intersubject differences in the binaural component of spatial unmasking roughly correlate with differences in BMLDs (Table I); however, intersubject differences in binaural sensitivity for one masker location do not predict results in other spatial configurations. For example, in the 500-Hz conditions when the masker is at 0°, subjects S1 and S3 [left columns in Figs. 6(a) and (c)] have larger binaural components of spatial unmasking than subject S2 [left column in Fig. 6(b)]. However, when the masker is at 90° [right columns of Figs. 6(a)–(c)], all three subjects exhibit essentially the same amount of binaural unmasking. This result suggests that intersubject differences in binaural sensitivity cannot be fully captured with a single ''binaural sensitivity'' parameter at each frequency [the degree to which intersubject differences can be predicted by Colburn's (1977b) model is considered further in Sec. V].

The magnitude of interaural level differences in the masker appears to have a large effect on the amount of binaural masking. For both target frequencies (Figs. 6 and 7), binaural unmasking is greatest when the masker is at 0° (and ITDs and ILDs in the masker are near zero; left columns in each subplot); when the masker is at 45° and 90° (center and right columns in each subplot), the amount of binaural unmasking decreases for the same angular separation of target and masker (i.e., even for roughly the same difference in target and masker ITD). When the masker is off to the side (right columns in the subplots of Figs. 6 and 7), the binaural contribution to spatial unmasking is also smaller when the masker is at 15-cm (when ILDs are very large; bottom right panels) compared to 1-m (when ILDs are smaller; top right panels). These effects are consistent with past reports showing that the BMLD decreases with masker ILD (e.g., see Durlach and Colburn, 1978, p. 433).

In general, the maximum difference in interaural phase difference (IPD) cues for target and masker arises when the ITDs for target and masker differ by one-half the period of the target frequency. For a 500-Hz target, the ITDs in target and masker need to differ by roughly 1 ms to maximize binaural unmasking. For a 1000-Hz target, the ITDs in target and masker need to differ by roughly 500 μs. This explains the dependence of maximal binaural unmasking on target and masker separation and frequency: results in Fig. 5 show that an angular separation of about 90° causes target and masker ITDs to differ by roughly 1 ms (maximizing IPD differences in target and masker for a 500-Hz target) whereas an angular separation of about 45° causes target and masker ITDs to differ by roughly 500 μs.

V. BINAURAL MODEL PREDICTIONS

A. Analysis

Subject-specific predictions of binaural unmasking were calculated using a modified version of the Colburn (1977a, 1977b) model (a description of the current implementation of the model is provided in the Appendix). Predictions depend on six parameters, evaluated at the target frequency: the ITDs and ILDs in both target and masker; the binaural sensitivity of the listener; and the spectrum level of the masker at the more intense ear relative to the absolute, monaural detection threshold in quiet.

The ITDs and ILDs used in the predictions were taken from the analysis of the cues present in the HRTFs. The ITD and ILD in masker were calculated from the values averaged over the ERB filter centered on the target frequency (see Fig. 5). The ITD and ILD in the target were taken directly from the HRTF values at the target frequency (not averaged over the ERB). Binaural sensitivity at each frequency was set to the measured BMLD for each subject and target frequency (Table I). For both the 500- and 1000-Hz targets, the monaural detection threshold (parameter K in the model) was set to 44 dB/Hz.

B. Results

Model predictions are plotted alongside behavioral estimates of the binaural contribution to spatial unmasking in Figs. 6 and 7 (for the 500- and 1000-Hz targets, respectively). In order to be somewhat conservative in identifying

conditions where the model fails to account for behavioral data, parallel lines plot a range of ±1 dB around the actual model predictions. Predictions for the nearby target are shown as dashed black lines; predictions for the far target are shown as solid gray lines.

Model predictions of binaural unmasking are nonnegative for all spatial configurations. Predictions are exactly zero whenever the target and masker are at the same spatial location and positive whenever the target and masker have differences in either their IPDs or ILDs at the target frequency. Thus, in theory, predictions of binaural unmasking are positive whenever the target and masker are at different distances but in the same direction off the median plane because of differences in ILDs in target and masker. However, in practice, predictions are near zero for all configurations when the target and masker are in the same direction for subjects S2, S3, and S4 [Figs. 6(b), 6(c), 7(b), and 7(c)]. Predictions for subject S1 [who has the largest ILDs for 15-cm sources and the largest BMLDs at both frequencies; Figs. 6(a) and 7(a)] are greater than zero for both target frequencies when the target and masker are at different distances but the same (off-median-plane) direction. For instance, in the top center and top right panels of Figs. 6(a) and 7(a) [masker at (45°, 1 m) and (90°, 1 m)], the black dotted lines (predictions for the target at 15 cm) are above zero for all target azimuths, including the target at 90°; in the bottom center and right panels of Figs. 6(a) and 7(a) [masker at (45°, 15 cm) and (90°, 15 cm)], the gray solid lines (predictions for the target at 1 m) are positive for all azimuths.

Binaural unmasking predictions are generally larger at 500 Hz (Fig. 6) than 1000 Hz (Fig. 7). At both frequencies, binaural unmasking varies with angular separation of target and masker; however, the angular separation that maximizes the predicted spatial unmasking depends on frequency. As in the behavioral results, predicted binaural unmasking is greatest when the target and masker are separated in azimuth by 90° for the 500-Hz target (Fig. 6) and 45° for the 1000-Hz target (Fig. 7), corresponding to separations that maximize the differences in target and masker IPD at the target frequency (e.g., in the left column of Fig. 6, when the 500-Hz masker is at 0°, the maximum predicted unmasking, shown by the lines, occurs for targets at +90° and −90°; however, in the left column of Fig. 7, when the 1000-Hz masker is at 0°, the maximum predicted unmasking generally occurs for targets at +45° and −45°).

Also consistent with behavioral results, the maximum predicted amount of binaural unmasking decreases with masker ILD. As a result, the predicted amount of binaural unmasking varies with masker location, systematically decreasing with increasing masker angle and decreasing when the masker is at 15-cm compared to 1-m. For instance, predicted levels of unmasking are generally largest when the masker is at 0° (left columns of Figs. 6 and 7) and decrease as the masker is laterally displaced (center and right columns). Similarly, the amount of unmasking tends to be larger for the top rows of data in Figs. 6 and 7, when the masker is at 1-m, than in the bottom rows of data, when the masker is at 15-cm.

Model predictions capture much of the variation in binaural unmasking; however, there are systematic prediction errors that are large compared to the intrasubject variability. (Note that the standard error in the mean behavioral results is less than or equal to 1 dB as a direct result of the experimental procedure. The error bars in the figure are even more conservative, showing the *range* of thresholds obtained over multiple runs.)

Predictions are first compared to behavioral results for the 500-Hz target (Fig. 6). Predictions for subject S1 agree well with behavioral results when the masker is at (0°, 15 cm) [bottom left panel of Fig. 6(a)] and reasonably well for three other masker locations [(45°, 15 cm), (90°, 15 cm), and (90°, 1 m); bottom center, bottom right, and top right panels of Fig. 6(a), respectively]. However, S1 predictions tend to overestimate binaural unmasking for two masker locations [(0°, 1 m) and (45°, 1 m); top left and top center panels of Fig. 6(a)]. For subject S2, predictions match behavioral results reasonably well when the masker is at 0° [see the top left and bottom left panels of Fig. 6(b)], independent of masker distance (although there are isolated data points for which the model overestimates binaural unmasking), but systematically underestimate binaural unmasking when the masker is at 45° and 90° for both masker distances [see center and right panels of Fig. 6(b), where symbols fall above lines]. Results for subject S3 are similar to those of subject S2: predictions are in good agreement with measurements when the masker is in the median plane [left panels of Fig. 6(c)] but underestimate binaural unmasking when the masker is laterally displaced [center and right panels of Fig. 6(c)].

Focusing on the 1000-Hz results (Fig. 7), subject S1 predictions generally overestimate binaural unmasking (in all panels in Fig. 7(a), symbols fall below lines). For subject S3, predictions generally underestimate binaural unmasking, except when the masker is at (45°, 1 m), where predictions and measurements are reasonably close [agreement between the measured data points and the prediction lines is good only for the top center panel of Fig. 7(b); for all other panels, symbols fall above lines]. Finally, predictions for subject S4 either fit reasonably well or underestimate binaural unmasking when the masker is at 0° [left panels of Fig. 7(c)] but overestimate binaural unmasking when the masker is at 45° or 90°, independent of masker distance [see center and right panels of Fig. 7(c), where symbols fall below lines].

Overall, predictions and behavioral results are in better agreement when the masker is in the median plane than when the masker is at 45° or 90° and for the 500-Hz data compared to the 1000-Hz data.

C. Discussion

The Colburn model assumes that a single value representing binaural sensitivity at a particular frequency can account for intersubject differences in binaural unmasking. This binaural sensitivity parameter was set from BMLD measures taken with a diotic masker and target that was either diotic (NoSo) or inverted at one ear to produce an interaural phase difference of π (NoSπ). These conditions are most analogous to the spatial configurations in which the masker is directly in front of the listener (and the masker is

essentially diotic). For most of the configurations with the masker at 0°, model predictions agree well with observed results. In contrast, larger discrepancies between the modeled and measured results arise when the masker is at 45° and 90° (conditions in which there are significant ILDs in the masker).

While there are some conditions in which the model predictions consistently over- or underestimate binaural unmasking [e.g., results for subject S1 at 1000 Hz in Fig. 7(a) or for subject S3 at 1000 Hz in Fig. 7(b)], there are other conditions for which changing the single subject-specific ''binaural sensitivity'' of the model cannot account for discrepancies between the model predictions and the measurements [e.g., results for subject S2 at 500 Hz in Fig. 6(b) or for subject S4 at 1000 Hz in Fig. 7(c)].

The current results suggest that subjects differ not only in their overall sensitivity to binaural differences, but also in the dependence of binaural sensitivity on the interaural parameters in masker and/or target. In particular, binaural sensitivity appears to depend on the interaural level difference in the masker differently for different subjects. As a result, individualized model prediction errors are generally larger when there are large ILDs in the masker than when the masker has near-zero ILD. While the Colburn model has been tested (and shown to predict results relatively well) in many studies in which target and masker vary in their interaural phase parameters, there are few studies that manipulate the target and masker ILD. These results suggest the need for additional behavioral and theoretical studies of the effects of ILD in binaural detection tasks.

Even though there are specific conditions for which predictions fail to account for the results for a particular subject, the model captures many of the general patterns in results, including the tendency for binaural unmasking to decrease as the ILD in the masker increases and how the amount of binaural unmasking depends on the angular separation of target and masker and the frequency of the target.

VI. GENERAL DISCUSSION

The current study is unique in measuring how tone detection thresholds are affected by target and masker location when sources are very close to the listener. Results show that for sources very close to the listener, small changes in source location can lead to large changes in detection threshold. These large changes arise from changes in both the TMR (affecting the better-ear contribution to spatial unmasking) and ILDs (affecting the binaural contribution to spatial unmasking).

The current results demonstrate how the relative importance of better-ear and binaural contributions to spatial unmasking change with target and masker location, including source distance (in contrast to previous studies that considered only angular separation of relatively distant sources). The relative importance of better-ear contributions to spatial unmasking increases as masker distance decreases, probably because of increases in the ILD in the masker, which reduce the amount of binaural unmasking. The better-ear contribution also increases as target distance decreases, primarily because the TMR changes more rapidly with target angle when the target is near the listener. The relative importance of the better-ear contribution to spatial unmasking increases with target frequency, both because the absolute magnitude of better-ear factors increases and because the binaural contribution to unmasking decreases. For a 500-Hz target, binaural and better-ear factors are roughly equally important when the masker is in the median plane. However, better-ear factors become relatively more important as the masker is displaced laterally, in part because the amount of binaural spatial unmasking decreases with masker ILD. This trend, which is predicted by the Colburn model, helps to explain large differences in the amount of spatial unmasking observed in previous studies (e.g., Ebata *et al.*, 1968; Gatehouse, 1987; Santon, 1987). Specifically, more spatial unmasking arises when the masker is positioned in front of the listener and the target location is varied (leading to near-zero ILDs in the masker) than when the target is fixed in location and the angle of masker is varied (leading to progressively larger ILDs in the masker with spatial separation of target and masker).

Binaural processing contributes up to 10 dB to spatial unmasking for the spatial configurations tested. In theory, differences in target and masker distance cause differences in target and masker ILD when the sources are off the median plane, leading to binaural unmasking. However, in the current study evidence of binaural unmasking resulting from differences in target and masker distance was observed only for Subject S1, who had both the largest BMLDs and the largest ILDs of the four subjects in the study.

Although monaural detection thresholds were not directly measured in the current study, binaural performance is always better than or equal to the performance predicted by analysis of the TMR at the better ear. Thus, the current study does not help to explain results suggesting that binaural performance sometimes falls below monaural performance using the better ear alone, particularly for configurations with large ILDs (Bronkhorst and Plomp, 1988; Shinn-Cunningham *et al.*, 2001). One important distinction between the current study and these previous reports is that the current study measured tone detection for relatively low-frequency tones, whereas both of the previously cited studies measured speech intelligibility, a suprathreshold task that emphasizes information at higher frequencies. Further studies are necessary to help determine when binaural stimulation may actually degrade performance compared to monaural, better-ear performance.

Intersubject differences in the amount of spatial unmasking are large and arise from individual differences in (1) HRTFs, (2) overall binaural sensitivity, and (3) the way in which binaural sensitivity varies with spatial configuration of target and masker. The Colburn (1977b) model of binaural processing predicts overall trends in behavioral measures of binaural unmasking, but fails to capture subject-specific variations in performance. The spatial configurations for which model predictions are least accurate are the positions for which large ILDs arise in masker and/or target, conditions that have not been extensively tested in previous studies. The current results suggest that the Colburn model must be modified so that subject differences in binaural sensitivity

vary not only in overall magnitude but as a function of the interaural differences in the masker.

While predictions from the Colburn model (taking into account differences in the stimuli presented to the individual subjects as well as individual differences in binaural sensitivity) cannot account for some small but significant intersubject differences in spatial unmasking, rough predictions of the amount of spatial unmasking capture most of the observed changes in detection threshold with spatial configuration. For instance, generic acoustic models of HRTFs (e.g., KEMAR measurements or spherical-head model predictions) combined with predictions of binaural unmasking using ''average'' model parameters should produce predictions that fall within the range of behavior observed across a population of subjects.

VII. CONCLUSIONS

(1) Acoustic cues (particularly TMR and ILD) vary dramatically with source distance and direction for nearby sources. Therefore, when source distance varies, the effect of source location on both the better-ear and binaural contributions to spatial unmasking is complex.
(2) For nearby sources, the better-ear contribution to pure-tone spatial unmasking can be very large (as much as 25 dB) compared to conditions where sources are relatively far from the listener.
(3) The binaural contribution to spatial unmasking decreases with increasing masker ILD. As a result, the binaural contribution to spatial unmasking is smaller for lateral sources very near the head than for more distant sources at the same lateral angle relative to the listener.
(4) Intersubject differences in spatial unmasking are larger for nearby sources than for far sources, in part because there are larger acoustic differences in HRTFs for nearby sources compared to more distant sources. However, there also are subject-specific differences both in binaural sensitivity and on how ILDs influence binaural sensitivity.
(5) Predictions based on Colburn's analysis (1977b) show the correct general trends in binaural detection for both near and far sources, but cannot account for small, but consistent, subject-specific differences in performance, particularly when large ILDs are present in the masker.

ACKNOWLEDGMENTS

This work was supported in part by AFOSR Grant No. F49620-98-1-0108 and the Alfred P. Sloan Foundation. Portions of this work were presented at the Spring 2000 meeting of the Acoustical Society of America. H. Steven Colburn provided valuable input throughout this work, including help in putting the results in appropriate context. Les Bernstein, Adelbert Bronkhorst, and an anonymous reviewer provided valuable criticism and comments on a previous draft of this paper.

APPENDIX: BINAURAL MODELING

A modified version of the model presented in Colburn (1977b) was used to predict the amount of binaural unmasking, defined as the difference in detection thresholds when target and masker are at the same spatial location and when they are in different locations. The predicted amount of binaural unmasking for a target at frequency f_0 is computed as

$$s(f_0,\alpha_T,\phi_T,\alpha_M,\phi_M,\mathrm{BMLD},K) = \sqrt{\max\left(1,\frac{\alpha_T^4}{\alpha_M^4}\right)+(2\cdot 10^{\mathrm{BMLD}/10}-1)R(\alpha_M,10^{K/10})\frac{F^2(\phi_M,f_0)}{16}\left(1+\frac{\alpha_T^2}{\alpha_M^2}-2\frac{\alpha_T}{\alpha_M}\cos(\phi_M-\phi_T)\right)^2}, \quad \text{(A1)}$$

where $\alpha_T = 10^{\mathrm{ILD}-T/20}$; $\alpha_M = 10^{\mathrm{ILD}-M/20}$; ILD$-T$ and ILD$-M$ are the interaural level differences in target and masker (respectively) in dB; ϕ_T and ϕ_M are the IPDs of target and masker (respectively) in radians; BMLD is the (subject-specific) binaural masking level difference in dB; K is the level of masker relative to absolute detection threshold in quiet, in dB; and the functions F^2 and R are defined below (all evaluated at the target frequency).

Function F^2 represents the extent to which phase shifts in masker cannot be compensated by internal time delays. This function is given by

$$F^2(\phi_M,f_0) = \frac{\Sigma_{k=-1000}^{1000} p(\phi_M/2\pi f_0 + k/f_0, f_0)\exp\{-G^2(f_0)[1-\gamma(\phi_M/2\pi f_0 + k/f_0)]\}}{\Sigma_{k=-1000}^{1000} p(k/f_0,f_0)\exp\{-G^2(f_0)[1-\gamma(k/f_0)]\}}, \quad \text{(A2)}$$

where $p(\tau,f)$ represents the relative number of interaural coincidence detectors (i.e., neurons in the medial superior olive) tuned to ITD τ and frequency f; $G(f)$ represents the synchrony of firings of the auditory nerve at frequency f (squared to account for the sharpening of synchrony in the cochlear nucleus); and $\gamma(\tau)$ is the envelope of the autocorrelation function of the auditory nerve fiber impulse response at autocorrelation delay τ. In the current realization of the model, function $p(\tau,f)$ was modified to allow for a frequency dependence in the distribution of interaural coincidence detectors (as suggested by Stern and Shear, 1996), using

$$p(\tau,f_0)=\begin{cases} C(e^{-2\pi k_l|0.2|}-e^{-2\pi k_h|0.2|})/0.2, & |\tau|<0.2 \text{ ms}, \\ C(e^{-2\pi k_l|0.2|}-e^{-2\pi k_h|0.2|})/|\tau|, & |\tau|\geqslant 0.2 \text{ ms}, \end{cases}$$

$$k_h=3\times 10^6,$$

$$k_l\begin{cases} 0.1(f_0 10^{-3})^{1.1}, & f_0\leqslant 1200 \text{ Hz}, \\ 0.1(1200\times 10^{-3})^{1.1}, & f_0>1200 \text{ Hz}, \end{cases} \tag{A3}$$

$$C=\begin{cases} 0.1534, & f_0=500 \text{ Hz}, \\ 0.2000, & f_0=1000 \text{ Hz}. \end{cases}$$

$G(f)$ is given by

$$G(f_0)=\begin{cases} \sqrt{10}, & f_0\leqslant 800 \text{ Hz}, \\ \sqrt{10}\dfrac{800}{f_0}, & f_0>800 \text{ Hz}. \end{cases} \tag{A4}$$

$\gamma(\tau)$ is given by

$$\gamma(\tau)=\begin{cases} 2.359\times 10^{-4}+1.5207\times 10^8\tau^4-1.764\times 10^4\tau^2 \\ \qquad +0.993, & |\tau|\leqslant 0.006, \\ -97.3236|\tau|+1.139, & |\tau|>0.006, \end{cases} \tag{A5}$$

where τ is in milliseconds.

Finally, function $R(\alpha,K)$ characterizes the decrease in the number of activated auditory nerve fibers in the ear receiving the less intense signal as a function of masker ILD. The current implementation uses a modified version of Eq. (35) from Colburn (1977b):

$$R(\alpha_n)=\begin{cases} \left(\dfrac{10\log_{10}\alpha_n^{-2}K}{40}\right)^2, & \alpha_n^{-2}K\leqslant 10^4, \\ 1, & \alpha_n^{-2}K>10^4, \end{cases} \tag{A6}$$

where K is the ratio of the spectrum level at the more intense ear to the detection threshold in quiet. This implementation of the model assumes that the auditory nerve fibers at each target frequency have thresholds uniformly distributed (on a dB scale) over a 40-dB range above the absolute detection threshold for that frequency.

[1]System identification using a MLS depends on circular convolution techniques. Theoretically, the approach requires the MLS to be concatenated with itself and presented an infinite number of times to ensure that the system is in its steady-state response prior to measuring the response (see Vanderkooy, 1994). The resulting estimated system response is a time-aliased version of the true system response. In the current measures, the MLS was presented twice and the response to the second repetition was recorded. Given the length of the MLS used, the room characteristics of and ambient noise in the environment in which we were measuring, and the noise in our measurement system, the steady-state response can be approximated with only two repetitions of the MLS and no significant time aliasing is present in our measurements.

[2]Note that this analysis assumes that detection performance depends only on the target-to-masker ratio or TMR and is independent of the overall masker level, an assumption that is not valid if the masker is near absolute threshold or at very high presentation levels. For instance, imagine two masker locations so distant from the listener that the masker is inaudible. These masker locations would produce identical signal detection thresholds if the experiment were performed with the distal stimulus intensity fixed; however, our technique might adjust the masker by different amounts for these two masker locations in order to achieve a fixed proximal stimulus level at the ear of the listener, producing two different estimates of spatial unmasking. While holding the distal masker intensity fixed may seem more natural and intuitive than holding the proximal stimulus level constant, the overall presentation level of the masker would span an extraordinarily large range in the current experiments because the masker distance varied between 15 cm and 1 m in addition to varying in direction. Therefore, we elected to fix the proximal masker intensity.

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.

Brungart, D. S., and Rabinowitz, W. M. (**1999**). "Auditory localization of nearby sources. I. Head-related transfer functions," J. Acoust. Soc. Am. **106**, 1465–1479.

Colburn, H. S. (**1977a**). "Theory of binaural interaction based on auditory-nerve data. II: Detection of tones in noise," J. Acoust. Soc. Am. **61**, 525–533.

Colburn, H. S. (**1977b**). "Theory of binaural interaction based on auditory-nerve data. II: Detection of tones in noise. Supplementary material," J. Acoust. Soc. Am. AIP document no. PAPS JASMA-61-525-98.

Doll, T. J., and Hanna, T. E. (**1995**). "Spatial and spectral release from masking in three-dimensional auditory displays," Hum. Factors **37**, 341–355.

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.

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

Ebata, M., Sone, T., and Nimura, T. (**1968**). "Improvement of hearing ability by directional information," J. Acoust. Soc. Am. **43**, 289–297.

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

Gatehouse, R. W. (**1987**). "Further research on free-field masking," J. Acoust. Soc. Am. Suppl. 1 **82**, S108.

Good, M. D., Gilkey, R. H., and Ball, J. M. (**1997**). "The relation between detection in noise and localization in noise in the free field," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. Gilkey and T. Anderson (Erlbaum, New York), pp. 349–376.

Kidd, Jr., G., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **104**, 422–431.

Levitt, H. (**1971**). "Transformed up-down methods in psychophysics," J. Acoust. Soc. Am. **49**, 467–477.

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, San Diego).

Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D. R. (**1991**). "Free-field release from masking," J. Acoust. Soc. Am. **90**, 1355–1370.

Santon, F. (**1987**). "Detection d'un son pur dans un bruit masquant suivant l'angle d'incidence du bruit. Relation avec le seuil de reception de la parole," Acustica **63**, 222–230.

Shinn-Cunningham, B. G., Santarelli, S., and Kopco, N. (**2000**). "Tori of confusion: Binaural localization cues for sources within reach of a listener," J. Acoust. Soc. Am. **107**, 1627–1636.

Shinn-Cunningham, B. G., Schickler, J., Kopco, N., and Litovsky, R. Y. (**2001**). "Spatial unmasking of nearby speech sources in a simulated anechoic environment," J. Acoust. Soc. Am. **110**, 1118–1129.

Stern, R. M., and Shear, G. D. (**1996**). "Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay," J. Acoust. Soc. Am. **100**, 2278–2288.

Vanderkooy, J. (**1994**). "Aspects of MLS measuring systems," J. Audio Eng. Soc. **42**, 219–231.

Effects of reverberation on perceptual segregation of competing voices

John F. Culling,[a] Kathryn I. Hodder, and Chaz Yee Toh
School of Psychology, Cardiff University, P.O. Box 901, Cardiff CF10 3YG, United Kingdom

(Received 19 September 2002; revised 5 August 2002; accepted 6 August 2003)

Two experiments investigated the effect of reverberation on listeners' ability to perceptually segregate two competing voices. Culling *et al.* [Speech Commun. **14**, 71–96 (1994)] found that for competing synthetic vowels, masked identification thresholds were increased by reverberation only when combined with modulation of fundamental frequency ($F0$). The present investigation extended this finding to running speech. Speech reception thresholds (SRTs) were measured for a male voice against a single interfering female voice within a virtual room with controlled reverberation. The two voices were either (1) co-located in virtual space at 0° azimuth or (2) separately located at ±60° azimuth. In experiment 1, target and interfering voices were either normally intonated or resynthesized with a fixed $F0$. In anechoic conditions, SRTs were lower for normally intonated and for spatially separated sources, while, in reverberant conditions, the SRTs were all the same. In experiment 2, additional conditions employed inverted $F0$ contours. Inverted $F0$ contours yielded higher SRTs in all conditions, regardless of reverberation. The results suggest that reverberation can seriously impair listeners' ability to exploit differences in $F0$ and spatial location between competing voices. The levels of reverberation employed had no effect on speech intelligibility in quiet. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1616922]

PACS numbers: 43.66.Pn, 43.66.Dc, 43.55.Hy [LRB] Pages: 2871–2876

I. INTRODUCTION

Most research on the perceptual effects of reverberation on speech has concentrated upon its effects on the transmission of a single voice in quiet (Houtgast and Steeneken, 1985). This work has been applied, in the form of the speech transmission index, to the particular problems of theatrical auditoria and lecture rooms where one-way verbal communication is the norm. Often these spaces are large and a degree of reverberation is desirable as a means of delivering the necessary sound level to the audience. However, too much reverberation can smear the temporal envelope of the speech, ultimately rendering it unintelligible. The speech transmission index can be used to predict the intelligibility of speech in quiet (or in simple forms of noise, such as might be produced by air conditioning) in different environments.

A relatively small amount of research has been conducted on the effects of reverberation on multi-talker communication (Plomp, 1976; Culling *et al.*, 1994; Darwin and Hukin, 2000). However, such work as exists seems to have serious implications for room design, because reverberation disrupts listeners' ability to cope with multiple overlapping voices far more easily than it does the intelligibility of a voice in quiet. Plomp used a reverberation room with varying amounts of inserted sound-absorbing material to show that thresholds for speech reception against interfering speech or noise were increased in a more reverberant enclosure. Furthermore, the beneficial effect of spatial separation of the target and interfering sources was largely abolished in the presence of reverberation. Culling *et al.* measured the masked identification thresholds for synthesized vowel sounds in a virtual-acoustic space with controlled surface absorption. Using a pink noise masker, they also found that the effect of spatial separation was easily abolished by reverberation. Using a competing vowel as a masker, the beneficial effect of differences in fundamental frequency ($F0$) between the two vowels (Scheffers, 1983; Assmann and Summerfield, 1990; Culling and Darwin, 1993a) was robust to reverberation. However, sinusoidal modulation of $F0$ (for both target and masker), which in anechoic conditions had no effect on masked thresholds, resulted in the abolition of the $F0$-difference benefit when combined with reverberation. Darwin and Hukin used a similar virtual-acoustic method to examine the effects of reverberation on listeners' ability to track a particular voice over time. They found that reverberation substantially reduced listeners' ability to use interaural time delays to attribute competing words to the correct carrier sentences. However, for this task, the benefits of continuity of $F0$ and vocal tract length were more robust in reverberation.

Current theories of segregation by $F0$ suggest that the auditory system can suppress one harmonic interfering voice, perhaps by a harmonic-cancellation process (de Cheveigné, 1997), permitting superior understanding of the remaining voice. The evidence for this scheme is largely based on experiments with simultaneous vowels. If simultaneous vowels have different $F0$s, then they can be identified more accurately than if they have the same $F0$, but two different $F0$s are not the only form of excitation of the vowels that will result in improvements in their identification. It is evidence from these alternative forms of excitation that points specifically to cancellation. If one vowel is inharmonic (Summerfield and Culling, 1992; de Cheveigné *et al.*, 1997), recognition of the competing vowel is improved compared to having

[a]Electronic mail: cullingj@cardiff.ac.uk

both on the same $F0$. Similarly, if one vowel is excited by noise, resulting in a whispered timbre, recognition of that vowel improves compared to having both vowels on the same $F0$ (Lea, 1992). In both cases, identification of the vowel that remains harmonic is unaffected. However, if both vowels are inharmonic or whispered, the advantage is lost. Thus, if one vowel has any spectral microstructure that differs from a harmonic competitor, then it can be perceptually separated from that competitor and better identified. A cancellation mechanism would be expected to produce this pattern of performance, because it will cancel the harmonic competitor, leaving inharmonic or noise-excited sounds relatively unaffected. In real listening situations, both voices are harmonic, but (most of the time) differ in $F0$. It seems likely that the cancelled voice is the dominant and/or more intense one, because ability to match the pitch of each vowel correlates with identification accuracy (Assmann and Paschall, 1998) and identification of the $F0$ is presumably a prerequisite for cancellation.

The human voice varies rapidly in $F0$ over a full octave during normally intonated speech. The question therefore arises of how the cancellation mechanism deals with this moving target. Further experiments with simultaneous vowels have modulated $F0$ sinusoidally, creating an effect similar to operatic vibrato. Using these stimuli, it has been found that the ability to exploit differences in $F0$ seems to correlate with the mean instantaneous difference in $F0$ across the stimulus (as opposed to the difference between the long-term mean $F0$'s). Thus, vowels modulated out of phase around the same mean $F0$ are better identified than if they are modulated in the same phase (Darwin and Culling, 1990).

Harmonic cancellation of the dominant voice will provide the listener with better identification of individual speech sounds, but the reconstruction of separated sentences also requires the linkage of separated speech elements across time. In addition to this cancellationlike process, therefore, it is possible that listeners use $F0$ in a number of other ways. First, the mean $F0$ of a person's voice may be used in order to focus attention on that voice in the presence of a competing interferer with a different mean $F0$ (Cherry, 1953); this would enable a listener to acquire or reacquire the appropriate stream of information and to avoid confusing it with the interfering stream. Second, the attention on the correct stream can also be maintained if the $F0$ of the target voice is tracked continuously (Parsons, 1976). Continuous tracking of the $F0$ may enable a listener to deal with two voices with the same mean $F0$, although the tracking process is susceptible to confusion when the two voices' $F0$'s intersect (Culling and Darwin, 1993b). Darwin and Hukin's (2000) experiments with reverberation indicate that use of the $F0$ contour to track a target voice is also affected by reverberation, but that it is more robust to reverberation than benefits due to differences in spatial location.

It is not entirely clear how the combination of $F0$ modulation and reverberation disrupts these $F0$-segregation mechanisms. However, it seems likely that, when the $F0$ varies over time, wavefronts that have been delayed by their passage around the walls of the room have a different $F0$ from direct sound that arrives simultaneously at the receiver; direct and reflected waves were emitted from the modulating source at different times. In this way, the $F0$ of the interfering source is smeared in the sense that the harmonic series is less clearly defined in the stimulus. This smearing may make the interfering voice more difficult to cancel. Darwin and Hukin (2000) showed that reverberation can also upset listeners ability to use $F0$ in order to link successive words from the same voice. It is less clear how the smearing would affect that process.

The present investigation is a follow up to that of Culling *et al.* (1994), using running speech. The stimuli used in their study were highly artificial, but contained key features found in everyday listening situations. Some degree of reverberation is common to practically all listening environments and modulation of $F0$, while not normally sinusoidal, is unavoidable in natural speech. Indeed, normally intonated speech involves modulation of $F0$ that is both rapid (up to 5 oct/s) and typically varies over a full octave (O'Shaunessy and Allen, 1983). This modulation is both faster and more extreme than the $\pm 0.7\%-12\%$, 5-Hz sinusoidal modulation used by Culling *et al.* It is noteworthy that the combination of such subtle modulation of $F0$ and reverberation resulted in a collapse in listeners' ability to use differences in $F0$, one of the best-established cues to perceptual separation of competing voices.

II. EXPERIMENT 1

Because Culling *et al.* (1994) found that the effect of differences in $F0$ was robust to reverberation when $F0$ was not modulated, but not when it was modulated, experiment 1 tested whether the same happens with running speech. In order to do this, the speech was resynthesized with either the original or a monotonized $F0$ contour. This method has previously been used in order to control differences in $F0$ for concurrent speech (Brokx and Nooteboom, 1982). Then, speech reception thresholds (SRTs) were measured for target and interfering voices that had these different contours. The results of Culling *et al.* suggest that SRTs might be lower using the monotonized speech than using normally intonated speech when reverberation is present, because the $F0$-segregation mechanism is only impaired in the intonated case.

A. Stimuli

The corpus of sentences was from the Harvard Sentence List (Rothauser *et al.*, 1969). The recordings of voice DA, made at M.I.T. and digitized at 20 kHz with 16-bit quantization, were used as the basis of all stimuli. The sentences have low predictability and each has five designated keywords (given here in capitals). For instance, one sentence used in the current experiment was ''the STEMS of the TALL GLASSES CRACKED and BROKE.'' These sentences were manipulated using the Praat PSOLA speech analysis and resynthesis package. For monotonized speech, the mean $F0$ for each sentence was calculated and the sentence was resynthesized with this $F0$ throughout.

Interfering sentences were generated by feminizing the voice of DA. His voice was increased in $F0$ by a factor of 1.8 and, using the resynthesis and resampling[1] method of

Darwin and Hukin (2000), the spectral envelope was shifted up in frequency by 15%, to simulate a shorter vocal tract. The factor of 1.8 reduced the number of target/interferer pairs for which the monotonous versions were an exact octave apart; the resulting mean absolute deviation from an octave relationship was just over 2 semitones, but 15% were still within half a semitone of an octave difference. Eight interfering sentences were created in this way.

Reverberation was added using the image (ray-tracing) method (Allen and Berkley, 1979; Peterson, 1986) as implemented in the |WAVE signal processing package (Culling, 1996). The virtual room and source/receiver configuration was identical to that of Culling *et al.* (1994). The room had dimensions 5 m long×3.2 m wide×2.5 m high and virtual sources were 2 m from the receivers. The two receivers, separated by 20 cm, were placed along an axis at 30° to the 5-m wall on either side of a center point located 1.2 m from the 5-m wall and 1.9 m from the 3.2-m wall. The receivers were modeled as microphones suspended in space with no head between them. Absorption coefficients for each internal surface of the room were 0.3 for the reverberant room, giving a direct-to-reverberant ratio of −10 dB and a reverberation time of approximately 400 ms. For the anechoic room the coefficients were set to 1, giving an infinite direct-to-reverberant ratio. Binaural stimuli were produced by generating the impulse responses for the two receivers in virtual space and convolving the speech samples with these two impulse responses.

Stimuli were created for eight different conditions. These conditions covered two levels of reverberation (anechoic versus reverberant), two forms of intonation (original versus monotonized) and two spatial configurations (0°/0° vs +60°/−60°) in all possible combinations. Ten target sentences were created for each condition. Target and interfering sounds shared the same reverberation and form of intonation.

B. Procedure

Sixteen listeners each attended a single 90-min session. The session began with two practice runs using monaurally presented and unprocessed speech, in order to familiarize the listeners with the task. The following eight runs measured SRTs in each of the eight different conditions. The order of the conditions was rotated for successive listeners, while the sentence materials remained in the same order. Each of the 80 target sentences was thus presented to every listener in the same order and contributed equally to each condition. This procedure also ensured that each condition was presented in each serial position within the experimental session, counterbalancing order effects.

SRTs were measured using a 1-up/1-down adaptive threshold method (Plomp and Mimpen, 1979; Plomp, 1986; Culling and Colburn, 2000). For an individual SRT measurement, the ten male-voice target sentences were presented one after another, each one against the same ''female-voice'' interfering sentence. The listeners were instructed to listen to the male voice. The target-to-interferer ratio was initially very low. In the initial phase, listeners had the opportunity to listen to the first sentence a number of times, each time with an increased target-to-interferer ratio. Listeners were instructed to attempt a transcript of the first sentence using a computer terminal when they believed that they could hear more than half the words of the male voice. Once the first transcript was entered, the correct transcript was displayed on the computer terminal, with the five key words in capitals. The listener self-marked how many of the key words were correct. Subsequent target sentences were presented only once and self-marked in a similar manner; the level of the target speech was decreased by 2 dB if the listener correctly identified three or more of the five key words in the previous sentence, and otherwise increased by 2 dB. SRTs for a given condition/run were taken as the mean signal level derived in this way on the last eight trials. Each SRT measurement used a different interfering sentence.

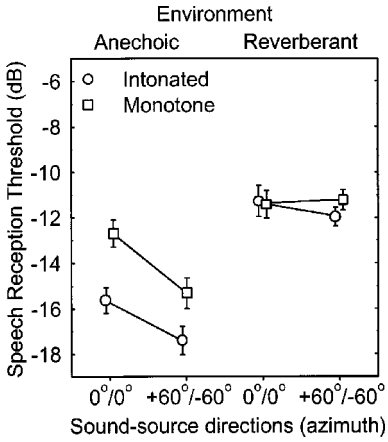


FIG. 1. Mean speech-reception thresholds in anechoic and reverberant conditions and for intonated (circles) and monotonous (squares) speech. Target and interfering sources were either both in front (0°/0°) or on either side (+60°/−60°). Lower thresholds imply greater intelligibility and/or perceptual separation. Error bars are one standard error of the mean.

Signals were digitally mixed, D/A converted, and amplified using a Tucker-Davis System II psychoacoustic rig (AP2, DD1, PA4, HB6) and presented to listeners over Sennheisser HD414 headphones in a single-walled IAC sound-attenuating booth within a sound-treated room. A computer terminal screen was visible outside the booth window; its keyboard was inside.

C. Results

In the anechoic conditions, Fig. 1 shows that mean SRTs were lower for intonated speech, indicating that listeners found the intonated speech intrinsically more intelligible than the monotonized speech. However, in the reverberant conditions there was no such effect. A three-factor analysis of variance (environment×$F0$ contour×spatial separation) reflected this pattern with a significant main effect of $F0$ contour $[F(1,15)=10.4, p<0.01]$ and an interaction between environment and $F0$ contour $[F(1,15)=20.0, p<0.001]$. Similarly, SRTs were lower for spatially separated voices in anechoic conditions, but not in reverberant conditions, producing a significant main effect of spatial separation $[F(1,15)=14.7, p<0.005]$ and an interaction between spatial separation and environment $[F(1,15)=5.4, p<0.05]$. Thus listeners could exploit the differences in spa-

tial location between the two competing voices, but only in the anechoic case. As a result of both these environment-specific effects, SRTs were also significantly lower overall in anechoic conditions [$F(1,15)=112.2$, $p<0.001$].

D. Discussion

Consistent with the results of Plomp (1976) and Culling *et al.* (1994), reverberation abolished listeners' ability to exploit differences in spatial location. The effect was more complete in the present study (and in that of Culling *et al.*) than in Plomp's experiments, probably because the sound sources in the present study were placed at a greater distance (2 m vs 1 m), within a smaller (virtual) room of (40 m^3 vs 63 m^3); both of these factors would have the effect of reducing the direct to reverberant ratio. This result is also consistent with Hukin and Darwin's work (using a similar virtual room, but with slightly different listener position) on the roles of $F0$ and ITD regarding the specific task of linking words from the same utterance. They found that reverberation disrupted both cues, but that the usefulness of different ITDs was more easily disrupted by reverberation than the effects of different $F0$s.

However, contrary to expectations based on the results of Culling *et al.*, monotonous speech was no more intelligible than intonated speech under reverberant conditions. In fact, intonated speech gave lower SRTs than monotonous speech in anechoic conditions and the two were approximately equal in reverberant conditions. It is possible that the monotonous condition was impaired to some degree by the occasional pair of target and interfering sentences that were close to an octave relationship. However, the advantage of intonated speech in anechoic conditions can probably be related more to exploitation of prosodic information. Prosodic information is provided by variations in the $F0$, amplitude, and rhythm of speech, so monotonization removes one of these three sources of information. The information contributes to intelligibility at multiple levels (Cutler *et al.*, 1997) and the removal of the $F0$-modulation element produces a cost in intelligibility equal to a 2.5-dB change in SRT (based on the difference in SRT for anechoic monotone and intonated $F0$ contours).

Since intonated speech is intrinsically more intelligible in anechoic conditions, one interpretation of the pattern of results is that reverberation destroys listeners' ability to exploit prosodic information conveyed by the intonation contour to assist speech intelligibility. However, given the results from Culling *et al.*'s experiments with concurrent synthetic vowel sounds, there is a more likely interpretation. It may be that intonated speech is intrinsically more intelligible than monotonous speech for all conditions, but that it is difficult to use $F0$ differences to perceptually separate two intonated voices in a reverberant setting; the monotonous speech may be perceptually separated from the (monotonous) interfering voice relatively well in the reverberation, but since it is less intelligible than the intonated speech, the SRT is no better. These two effects may be offsetting each other and yielding similar SRTs in all the reverberant conditions. Experiment 2 was designed to differentiate between these two possibilities.

III. EXPERIMENT 2

Experiment 2 discriminated between the different possible interpretations of the results from experiment 1 by adding conditions that used inverted intonation contours. These contours provide equal modulation of $F0$ (to disrupt segregation by $F0$ under reverberation), but were not expected to contribute to intrinsic speech intelligibility. Speech with an inverted $F0$ contour has a vague, questioning tone; the fall in $F0$ characteristic of the end of a statement is replaced with the rising $F0$ contour of a question and the stress sounds odd, because stressed syllables have an unnatural combination of low pitch and high intensity. Otherwise, the inverted-$F0$ speech sounded clearly articulated and natural.

A. Stimuli

The stimuli were largely similar to those for experiment 1, but using different target sentences from the same voice. In addition, the larger number of conditions required some additional interfering sentences; the choice of all 12 interfering sentences was reviewed to ensure that they were longer than all target sentences.

The eight conditions from experiment 1 were replicated. Four additional conditions were added that had inverted $F0$ contours. Inversion of the $F0$ contour was applied to both target and interferer. For inverted speech, the new $F0$, $F0'$, was derived for each analysis frame using the following equation:

$$F0' = \frac{\text{mean}\, F0^2}{F0}. \tag{1}$$

Here, $F0$ is the fundamental frequency of the frame and mean $F0$ is the mean fundamental frequency calculated over the duration of the sentence.

B. Procedure

Thirty-six new listeners each attended a single 2-h session. They completed the same two practice runs as in experiment 1 and 12 experimental runs, covering the 12 different conditions. As in experiment 1 the conditions were rotated from one listener to the next, while the sentence materials remained in the same order. The equipment was identical save for the use of Sennheiser HD590 headphones.

C. Results

Figure 2 shows mean SRTs for 36 listeners in experiment 2. SRTs for the eight conditions replicated from experiment 1 were similar in pattern to those from that experiment, although on average several dB higher. The effect of spatial location was, again, abolished by reverberation, and intonated speech again gave lower thresholds than monotonous speech in anechoic conditions only. SRTs for the four additional conditions with inverted $F0$ contours were substantially higher than the other conditions across all conditions of reverberation and spatial separation.

The results were analyzed with a three-way analysis of variance (environment$\times F0$ contour$\times$spatial separation). SRTs were, again, significantly lower for spatially separated sources [$F(1,35)=26.9$, $p<0.001$] and under anechoic con-

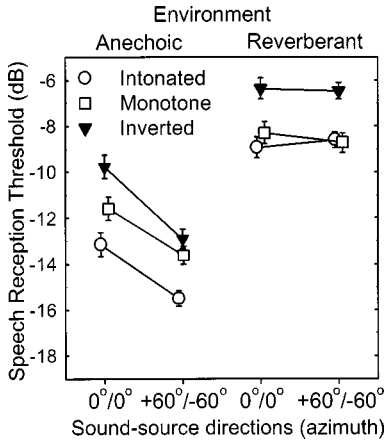


FIG. 2. Mean speech-reception thresholds in anechoic and reverberant conditions and for intonated (open circles), monotonous (open squares), and inverted-$F0$ (filled inverted triangles) speech. Target and interfering sources were either both in front (0°/0°) or on either side (+60°/−60°). Lower thresholds imply greater intelligibility and/or perceptual separation. Error bars are one standard error of the mean.

ditions [$F(1,35)=407.5$, $p<0.001$]. In addition, the three types of $F0$ contour differed significantly [$F(2,70)=60.3$, $p<0.001$]. The effects of spatial separation and $F0$ contour interacted with the presence of reverberation. First, the effect of spatial separation was abolished in reverberation, producing an interaction between environment and spatial separation [$F(1,35)=51.1$, $p<0.001$]. Simple main effects showed the effect of spatial separation to be significant only in anechoic conditions [$F(1,1)=50.9$, $p<0.001$]. Second, the convergence of SRTs from the monotonous and normally intonated conditions under reverberation produced an interaction between $F0$ contour and environment [$F(2,70)=4.3$, $p<0.02$]. However, SRTs from the new inverted-$F0$ conditions did not converge with the other conditions in reverberation. Tukey pairwise comparisons confirmed that all three $F0$ contours differed from each other in anechoic conditions (normally intonated versus monotonous, $q=7.01$, $p<0.001$; normally intonated versus inverted, $q=12.16$, $p<0.001$; monotonous versus inverted, $q=5.15$, $p<0.01$), but that in reverberant conditions the inverted condition produced higher SRTs (intonated versus inverted, $q=9.56$, $p<0.001$; monotonous versus inverted, $q=8.51$, $p<0.001$), while the monotonous and normally intonated conditions were indistinguishable. It is worth noting that the difference between the normally intonated and the inverted conditions contracted only marginally from 2.99 dB in anechoic conditions to 2.35 dB in the reverberant conditions. Thus, the $F0$-contour×environment interaction was produced by a change in the SRTs for the monotonous conditions *relative* to the other two when the environment is changed from anechoic to reverberant.

D. Discussion

Surprisingly, the inverted-$F0$ speech was *less* intelligible than the monotonized speech, despite the fact that it sounded considerably more acceptable, and less artificial, than the monotonized speech. The best explanation we can offer for this outcome is that when the $F0$ contour is monotonized there is a simple loss of prosodic information. Prosodic information usually comes from three sources, the $F0$ contour, the intensity contour, and the rhythm of the utterance. A monotonous $F0$ contour provides no prosodic information; listeners either disregard it, or simply perform worse due to the loss of information. In the inverted-$F0$ condition, on the other hand, the intonation contour is plausible, and listeners clearly attempt to exploit it. Since it is not the correct contour, it does not provide the correct information. Indeed, it probably disturbs listeners' normal processing by providing information that conflicts with that from the rhythmic and intensive aspects of the speech. The listeners' comprehension of the sentences is thus actively mislead.

It has previously been demonstrated that distortions of normal prosodic information can affect speech processing. For instance, Cutler and Clifton (1984) made recordings of isolated two-syllable words using a trained speaker who deliberately misplaced the primary lexical stress. Reaction times in a semantic processing task showed that listeners could process correctly intonated words more rapidly than incorrectly intonated ones. However, Cutler and Clifton's experiment and similar experiments by other authors were confounded to some extent by phonetic differences between stressed and unstressed syllables (Cutler *et al.*, 1997). Cutler *et al.* conclude that the role of lexical stress in lexical access is probably quite limited for English, because few words are distinguished by prosody alone. Using a cross-splicing technique, Cutler and Darwin (1981) showed that preceding prosodic context had a strong influence on the speed and accuracy of processing of subsequent words. By independently modulating the amplitude, timing, and $F0$ cues, Cutler (1987) showed that each cue made its own contribution to this effect, although, when intensity and $F0$ cues were inconsistent (as in experiment 2), reaction times were particularly long. In addition to these effects, it is possible that distortions of vowel intrinsic pitch are making some contribution to the deleterious effect of inverted $F0$ contours.

Regarding the original purpose of the experiment, the large difference between the intonated and inverted-$F0$ conditions shows that listeners *were* able to exploit information conveyed by the $F0$ contour in the presence of reverberation. Since this difference in thresholds is of a similar magnitude in both anechoic and reverberant conditions, it seems likely that the inverted $F0$ contour continues to actively mislead listeners in the reverberant case. This outcome clarifies the interpretation of experiment 1.; the idea that reverberation destroys listeners' ability to make use of the prosodic information in the $F0$ contours must be abandoned. In both experiments, the differences between normally intonated and monotonized speech were abolished in reverberant conditions. Since reverberation does not affect prosodic processing, then this effect must be attributed to better perceptual separation of the monotonized speech, compared to the normally intonated speech under reverberation. The more robust perceptual segregation of monotonized speech in reverberant conditions can be seen from the fact that it has a lower SRT compared to the intonated and inverted conditions in the reverberant case than it does in the anechoic case.

Finally, overall differences in mean SRT between ex-

periments 1 and 2 can be mainly attributed to the change in the set of target sentences. The differences observed here are consistent with unpublished measurements by Zurek (1996) using the same recordings. These show that lists 1–12 from the Harvard corpus of sentences (used in experiment 1) tend to yield SRTs 2–3 dB lower than lists 40–73. Experiment 2 used lists 40–51, inclusive. More careful selection of interfering sentences in experiment 2 (so that they were always longer than the targets) may also have contributed to the higher SRTs observed in that experiment.

IV. CONCLUSIONS

The hypothesis that speaking in a monotone at reverberant cocktail parties would aid communication is not supported, because monotonous speech is intrinsically less intelligible than normally intonated speech. Nevertheless, we have shown that reverberation has a detrimental effect on listeners' ability to perceptually separate voices with normally intonated $F0$ contours. Reverberation also disrupts listeners' ability to exploit differences in the spatial location of competing voices/sounds. These two effects both degrade social communication in reverberant rooms, and should be considered when designing spaces intended for social interaction.

ACKNOWLEDGMENTS

This work was supported by Hanse-Wissenschaftskolleg (HWK) Lehmkuhlenbusch 4, 27753 Delmenhorst, Germany. Chris McGowan collected additional data for experiment 2. We are indebted to Chris Darwin, one anonymous reviewer, and Les Bernstein as editor for their thorough and thoughtful reviews. We are also indebted to Paul Boersma and David Weenink for the use of their Praat software.

[1]A 15% vocal-tract shortening is achieved by resampling to a sampling rate 15% lower, and then playing back using the original sampling rate. This operation also increases the articulation rate and the $F0$ by 15%, so the speech is first resynthesized with a time-warp so that its articulation rate is reduced by 15%. At this point, the $F0$ can also be transformed. A reduction by 15% will compensate for the resampling. However, in the present application the $F0$ was increased overall by a factor of 1.8 in order to bring it into the female range. Thus, the $F0$ transformation applied was 1.8/1.15.

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950.

Assmann, P. F., and Paschall, D. (**1998**). "Pitches of concurrent vowels," J. Acoust. Soc. Am. **103**, 1150–1160.

Assmann, P. F., and Summerfield, Q. (**1990**). "Modelling the perception of concurrent vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech with one and two ears," J. Acoust. Soc. Am. **25**, 957–959.

Culling, J. F. (**1996**). "Signal processing software for teaching and research in psychoacoustics under UNIX and X-windows," Behav. Res. Methods Instrum. Comput. **28**, 376–382.

Culling, J. F., and Colburn, H. S. (**2000**). "Binaural sluggishness in the perception of tone sequences and speech in noise," J. Acoust. Soc. Am. **107**, 517–527.

Culling, J. F., and Darwin, C. J. (**1993a**). "Perceptual separation of simultaneous vowels: within and across-format grouping by $F0$," J. Acoust. Soc. Am. **93**, 3454–3467.

Culling, J. F., and Darwin, C. J. (**1993b**). "The role of timbre in the segregation of simultaneous voices with intersecting $F0$ contours," Percept. Psychophys. **54**, 303–309.

Culling, J. F., Summerfield, Q., and Marshall, D. H. (**1994**). "Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for separating concurrent vowels," Speech Commun. **14**, 71–96.

Cutler, A. (**1987**). "Components of prosodic effects in speech recognition," in *Proceedings of the Eleventh International Congress of Phonetic Sciences*, Tallinn, Estonia, Vol. 1, pp. 84–87.

Cutler, A., and Clifton, C. (**1984**). "The use of prosodic information in word recognition," in *Attention and Performance X: Control of Language Processes*, edited by H. Bouma and D. G. Bouwhuis (Erlbaum, Hillsdale, NJ).

Cutler, A., and Darwin, C. J. (**1981**). "Phoneme-monitoring reaction time and preceding prosody: effects of stop closure duration and of fundamental frequency," Percept. Psychophys. **29**, 217–224.

Cutler, A., Dahan, D., and van Donselaar, W. (**1997**). "Prosody in the comprehension of spoken language: A literature review," Lang. Speech **40**, 141–201.

Darwin, C. J., and Culling, J. F. (**1990**). "Speech perception seen through the ear," Speech Commun. **9**, 469–476.

Darwin, C. J., and Hukin, R. W. (**2000**). "Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention," J. Acoust. Soc. Am. **108**, 335–342.

de Cheveigné, A. (**1997**). "Concurrent vowel identification. III. A neural model of harmonic interference cancellation," J. Acoust. Soc. Am. **101**, 2857–2865.

de Cheveigné, A., McAdams, S., and Marin, C. (**1997**). "Concurrent vowel identification. II. Effects of phase, harmonicity, and task," J. Acoust. Soc. Am. **101**, 2848–2856.

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.

Lea, A. (**1992**). "Auditory models of vowel perception," Ph.D. thesis, Nottingham (unpublished).

O'Shaunessy, D., and Allen, J. (**1983**). "Linguistic modality effects on fundamental frequency in speech," J. Acoust. Soc. Am. **74**, 1155–1171.

Parsons, T. W. (**1976**). "Separation of speech from interfering speech by means of harmonic selection," J. Acoust. Soc. Am. **60**, 911–918.

Peterson, P. M. (**1986**). "Simulating the response of multiple microphones to single acoustic source in a reverberant room," J. Acoust. Soc. Am. **80**, 1527–1529.

Plomp, R. (**1976**). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," Acustica **34**, 200–211.

Plomp, R. (**1986**). "A signal-to-noise ratio model for the speech reception threshold of the hearing impaired," J. Speech Hear. Res. **29**, 146–154.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.

Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). "I.E.E.E. recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.

Scheffers, M. T. M. (**1983**). "Sifting vowels: Auditory pitch analysis and sound segregation," Ph.D. thesis, Gronigen.

Summerfield, Q., and Culling, J. F. (**1992**). "Periodicity of maskers not targets determines ease of perceptual segregation using differences in fundamental frequency," J. Acoust. Soc. Am. **92**, 2317 (A).

Zurek, P. (**1996**). Personal communication.

Blind estimation of reverberation time

Rama Ratnam,[a] Douglas L. Jones, Bruce C. Wheeler, William D. O'Brien, Jr., Charissa R. Lansing, and Albert S. Feng

Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

(Received 14 March 2003; revised 2 August 2003; accepted 18 August 2003)

The reverberation time (RT) is an important parameter for characterizing the quality of an auditory space. Sounds in reverberant environments are subject to coloration. This affects speech intelligibility and sound localization. Many state-of-the-art audio signal processing algorithms, for example in hearing-aids and telephony, are expected to have the ability to characterize the listening environment, and turn on an appropriate processing strategy accordingly. Thus, a method for characterization of room RT based on passively received microphone signals represents an important enabling technology. Current RT estimators, such as Schroeder's method, depend on a controlled sound source, and thus cannot produce an online, blind RT estimate. Here, a method for estimating RT without prior knowledge of sound sources or room geometry is presented. The diffusive tail of reverberation was modeled as an exponentially damped Gaussian white noise process. The time-constant of the decay, which provided a measure of the RT, was estimated using a maximum-likelihood procedure. The estimates were obtained continuously, and an order-statistics filter was used to extract the most likely RT from the accumulated estimates. The procedure was illustrated for connected speech. Results obtained for simulated and real room data are in good agreement with the real RT values.
[DOI: 10.1121/1.1616578]

PACS numbers: 43.66.Yw [MK] Pages: 2877–2892

I. INTRODUCTION

The estimation of room reverberation time (RT) has been of interest to engineers and acousticians for nearly a century (Sabine, 1922; Kuttruff, 1991). The RT of a room specifies the duration for which a sound persists after it has been switched off. The persistence of sound is due to the multiple reflections of sound from the various surfaces within the room. Historically, the RT has been referred to as the T_{60} time, which is the time taken for the sound to decay to 60 dB below its value at cessation.

Reverberation results in temporal and spectral smearing of the sound pattern, thus distorting both the envelope and fine structure of the received sound. Consequently, the RT of a room provides a measure of the listening quality of the room. This is of particular importance in speech perception where it has been noted that speech intelligibility reduces as the RT increases, due to masking within and across phonemes (Knudsen, 1929; Bolt and MacDonald, 1949; Nabelek and Pickett, 1974; Nabelek and Robinson, 1982; Nabelek *et al.*, 1989). The effect of reverberation is most noticeable when speech recorded by microphones is played back via headphones. Previously unnoticed distortions in the sound pattern are now clearly discerned even by normal listeners [see Hartmann (1997) for a discussion], highlighting the remarkable echo suppression and dereverberation capabilities of the normal auditory system when the ears receive sounds directly. For hearing-impaired listeners, the reception of reverberant signals via the microphone of a hearing aid exacerbates the problem of listening in challenging environments. While derverberation is an active area of investigation, state-of-the-art hearing aids, or other audio processing instruments, implement signal processing strategies tailored to specific listening environments. These instruments are expected to have the ability to evaluate the characteristics of the environment, and accordingly turn on the most appropriate signal processing strategy. Thus, a method that can characterize the RT of a room from passively received microphone signals represents an important enabling technology.

In the early 20th century, Sabine (1922) provided an empirical formula for the explicit determination of RT based solely on the geometry of the environment (volume and surface area) and the absorptive characteristics of its surfaces. Since then, Sabine's reverberation-time equation has been extensively modified and its accuracy improved [see Kuttruff (1991) for a historical review of the modifications], so that, today, it finds use in a number of commercial software packages for the acoustic design of interiors. Formulas for calculation of RT are used in anechoic chamber measurements, design of concert halls, classrooms, and other acoustic spaces where the quality of the received sound is of greatest importance, and the extent of reverberation must be controlled. However, to determine the RT of existing environments, both the geometry and the absorptive characteristics have to be first determined. When these cannot be determined easily, it is necessary to search for other methods, such as those based purely on the controlled recordings of excitation sounds radiated into the test enclosure.

Methods that employ an excitation signal for measuring RT are based on sound decay curves. In the Interrupted Noise Method (ISO 3382, 1997), a burst of broad- or

[a]Electronic mail: ratnam@uiuc.edu

narrow-band noise is radiated into the test enclosure. When the sound field attains steady state, the noise source is switched off and the decay curve is recorded. RT is estimated from the slope of the decay curve. However, because of fluctuations in the excitation noise signal, the decay curve will differ from trial to trial, and so RTs from a large number of decay curves must be averaged to obtain a reliable estimate. To overcome this drawback Schroeder (1965a, 1966) developed the Integrated Impulse Response Method where the excitation signal is a brief pulse, either broad- or narrow-band. For a brief pulse the enclosure output is simply the impulse response of the enclosure in the specified frequency band. Schroeder showed that the impulse response of the enclosure is related via a certain integral to the ensemble average of the decay curve obtained using the interrupted noise method, and so repeated trials were unnecessary. Both methods, while theoretically and practically important, require careful control of the experiment. Specifically, a suitable excitation signal must be available, and it must have sufficient power to provide at least a 35-dB decay range before the noise floor is encountered [see ISO 3382 (1997) for specifications of the experiment]. Under these conditions, both methods provide reliable RT estimates, with Schroeder's method being superior because it is the average of an infinite number of interrupted noise measurements.

While Schroeder's method continues to have immense practical utility, and has been improved over the years (see Chu, 1978; Xiang, 1995, for example), there is at present no ''blind'' method that can estimate room RT from passively received microphone signals. The objective of this work is to establish a method for determining RT when the room geometry and absorptive characteristics are unknown, or when a controlled test sound cannot be employed. A blind method that works with speech sounds would be particularly important for incorporating in hearing-aids or hands-free telephony devices. Partially blind methods have been developed in which the characteristics of the room are ''learned'' using neural network approaches (Tahara and Miyajima, 1998; Nannariello and Fricke, 1999; Cox *et al.*, 2001), or some form of segmentation procedure is used for detecting gaps in sounds to allow the sound decay curve to be tracked (Lebart *et al.*, 2001). The only other method that can be described as truly blind is ''blind dereverberation,'' where the aim is to recover a sound source by deconvolving the room output with the unknown room impulse response. When deconvolution is successful, a useful by-product is the room impulse response from which the RT can be estimated (using, say, Schroeder's method). However, deconvolution is difficult to perform because it requires the room impulse response to be minimum phase, a condition that is not met in most real environments (Neely and Allen, 1979; Miyoshi and Kaneda, 1988). It should be noted that RT can always be determined if the room impulse response is known, whether it is minimum or nonminimum phase. The minimum phase condition is only necessary for determining the impulse response via a deconvolution. This limits the applicability of the method.

Here we develop a technique for blind estimation of reverberation time based solely on passively recorded sounds. The estimator is based on a simplified noise decay curve model describing the reverberation characteristics of the enclosure. Sounds in the test enclosure (speech, music, or other pre-existing sounds) are continuously processed and a running estimate of the reverberation time is produced by the system using a maximum-likelihood parameter estimation procedure. A decision-making step then collects estimates of RT over a period of time and arrives at the most likely RT using an order-statistics filter. The method complements existing methods of RT estimation, being useful in situations where only passively received microphone signals are available.

II. THEORY

A model for blind estimation of reverberation time is presented. This is followed by an algorithm for implementation, and a decision-making strategy for selecting the estimate that best represents the reverberation time of listening rooms.

A widely used measure of the reverberation time is the T_{60} time first defined by Sabine (1922) and which is now a part of the ISO reverberation measurement procedure (ISO 3382, 1997). The T_{60} time measures the time taken for the sound level to drop 60 dB below the level at sound cessation. In practice, a decaying sound in a real environment reaches the ambient noise floor, thus limiting the dynamic range of the measured sound to values less than 60 dB, and so it is usually not possible to directly measure T_{60}. Instead, the decay rate is estimated by a ''linear least-squares regression of the measured decay curve from a level 5 dB below the initial level to 35 dB.'' [definition adopted from ISO 3382 (1997), p. 2]. If a 30-dB decay range cannot be measured, then a 20-dB range can be used. The T_{60} is simply the time taken to decrease by 60 dB from the initial level at the same decay rate given by the above measurements.

Before describing the model, we motivate the work with an example. The recorded response of a room to an impulsive sound source (a hand-clap) is shown in Fig. 1(a). As can be expected, there are strong early reflections followed by a decaying reverberant tail. If the early reflections are ignored, the decay rate of the tail can be estimated from the envelope. Figure 1(c) shows the measurement of T_{60} using the decay rate estimated from the -5- to -25-dB decay region. The procedure that was followed was that developed by Schroeder (1965a) described below.

We begin with a model for the diffusive or reverberant tail of sounds in a room. This refers to the dense reflections that follow the early reflections. All that can be said about them is that they are the result of multiple reflections, and appear in random order, with successive reflections being damped to a greater degree if they occur later in time. The assumption of randomness is crucial to the development of a statistical model. When a burst of white noise is radiated into a test enclosure, the phase and amplitudes of the normal modes are random in the instant preceding the cessation of the sound. Consequently, the decaying output of the enclosure following sound cessation will also be random, even if repeated trials were conducted with the same source and receiver geometry. Traditionally, and dating back to Sabine, the late decay envelope has been modeled as an exponential with

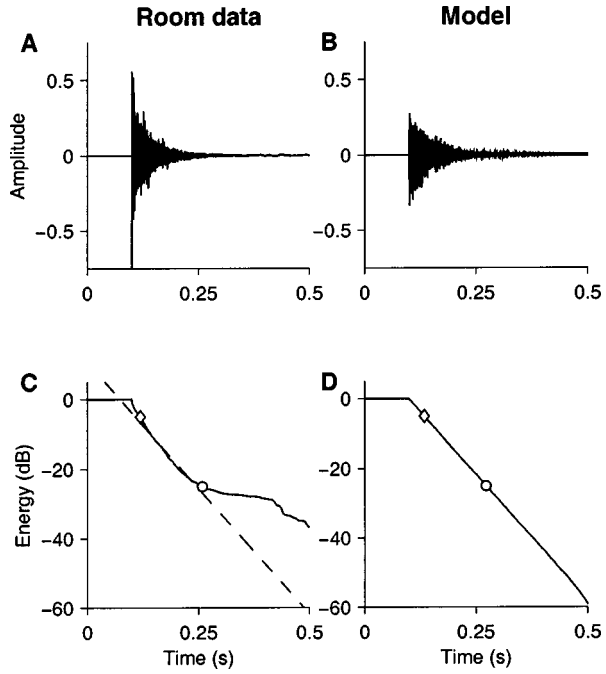


FIG. 1. Temporal decay of a hand-clap at $t=0.1$ s as recorded by a microphone (left column) and the model matching the reverberation (right column). (a) The recorded sound shows strong early reflections followed by a reverberant tail. Direct sound is excluded from the trace. (b) Model matching the reverberant tail shown in (a). Direct and early reflections are excluded. The model is a Gaussian white noise process damped by a decaying exponential, parametrized by the noise power σ and decay rate τ. (c) Decay rate estimated from Schroeder's backward integration method (Schroeder, 1965a) between -5 dB ($\diamond$) and -25 dB ($\circ$). Slope of linear fit (dashed line) yields $\tau=59$ ms ($T_{60}=0.4$ s). (d) Decay curve for model has identical slope everywhere following sound offset, and captures the most significant part of decay (-5 to -25 dB).

a single (deterministic) time-constant (hereafter referred to as decay rate). But, because the dense reflections are assumed to be uncorrelated, a convenient though highly simplified model is to consider the reverberant tail to be an exponentially damped uncorrelated noise sequence with Gaussian characteristics. The model does not include the direct sound or early reflections. The goal is to estimate the decay rate of the envelope.

A. Model of sound decay

We assume that the reverberant tail of a decaying sound y is the product of a fine structure x that is random process, and an envelope a that is deterministic. A central assumption is that x is a wideband process subject to rapid fluctuations, whereas the variations in a are over much longer time scales. Here, we will provide a statistical description of the reverberant tail with the goal of estimating the decay rate of the envelope.

Let the fine structure of the reverberant tail be denoted by a random sequence $x(n)$, $n\geqslant 0$, of independent and identically random variables drawn from the normal distribution $\mathcal{N}(0,\sigma)$. Further, for each n we define a deterministic sequence $a(n)>0$. The model for room decay then suggests that the observations y are specified by the sequence $y(n)=a(n)x(n)$. Due to the time-varying term $a(n)$, the $y(n)$ are independent but not identically distributed, and their probability density function is $\mathcal{N}(0,\sigma a(n))$. That is, the sequence $a(n)$ modulates the instantaneous power of the fine structure. For purposes of estimating the decay rate, we consider a finite sequence of observations, $n=0,\ldots,N-1$, where N will be referred to as the estimation interval, or estimation window length. For notational simplicity, denote the N-dimensional vectors of y and a by $\mathbf{y}$ and $\mathbf{a}$, respectively. Then the likelihood function of $\mathbf{y}$ (the joint probability density), parametrized by $\mathbf{a}$ and σ, is

$$L(\mathbf{y};\mathbf{a},\sigma)=\frac{1}{a(0)\cdots a(N-1)}\left(\frac{1}{2\pi\sigma^2}\right)^{N/2}\times\exp\left(-\frac{\Sigma_{n=0}^{N-1}(y(n)/a(n))^2}{2\sigma^2}\right), \tag{1}$$

where $\mathbf{a}$ and σ are the $(N+1)$ unknown parameters to be estimated from the observation $\mathbf{y}$. The likelihood function given by Eq. (1) is somewhat general, and, while it is possible to develop a procedure for estimating all $(N+1)$ parameters, suitable simplifications can be made when modeling sound decay in a room. Let a single decay rate τ describe the damping of the sound envelope during free decay. Then the sequence $a(n)$ is uniquely determined by

$$a(n)=\exp(-n/\tau). \tag{2}$$

Thus, the N-dimensional parameter a can be replaced by a scalar parameter a that is expressible in terms of τ and a single parameter $a=\exp(-1/\tau)$, so that

$$a(n)=a^n. \tag{3}$$

Introducing Eq. (3) into Eq. (1) yields

$$L(\mathbf{y};a,\sigma)=\left(\frac{1}{2\pi a^{(N-1)}\sigma^2}\right)^{N/2}\times\exp\left(-\frac{\Sigma_{n=0}^{N-1}a^{-2n}y(n)^2}{2\sigma^2}\right). \tag{4}$$

For a fixed observation window N and a sequence of observations $y(n)$, the likelihood function is parametrized solely by the decay rate a and the diffusive power σ.

The model is shown in Fig. 1(b) with parameters a and σ matched to the experimental hand-clap data shown in Fig. 1(a). Note that the model does not include the early reflections shown in panel (a). The Schroeder decay curve for the model is shown in Fig. 1(d) with a T_{60} time of 0.4 s in agreement with the measured T_{60}. The agreement between model and real T_{60} time motivates the search for an algorithm that can optimally estimate the two parameters.

B. Maximum-likelihood estimation

Given the likelihood function, the parameters a and σ can be estimated using a maximum-likelihood approach (Poor, 1994). First, we take the logarithm of Eq. (4) to obtain the log-likelihood function

$$\ln L(\mathbf{y};a,\sigma) = -\frac{N(N-1)}{2}\ln(a) - \frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1} a^{-2n}y(n)^2. \quad (5)$$

To find the maximum of $\ln(L)$, we differentiate the log-likelihood function Eq. (5) with respect to a to obtain the score function s_a (Poor, 1994):

$$s_a(a;\mathbf{y},\sigma) = \frac{\partial \ln L(\mathbf{y};a,\sigma)}{\partial a} \quad (6)$$

$$= -\frac{N(N-1)}{2a} + \frac{1}{a\sigma^2}\sum_{n=0}^{N-1} n a^{-2n}y(n)^2. \quad (7)$$

The log-likelihood function achieves an extremum when $\partial \ln L(\mathbf{y};a,\sigma)/\partial a = 0$; that is, when

$$-\frac{N(N-1)}{2a} + \frac{1}{a\sigma^2}\sum_{n=0}^{N-1} n a^{-2n}y(n)^2 = 0. \quad (8)$$

The zero of the score function provides a best estimate in the sense that $\mathbf{E}[s_a] = 0$.

Denote the zero of the score function s_a, and satisfying Eq. (8), by a^*. It can be shown that the second derivative $\partial^2 \ln L(\mathbf{y};a,\sigma)/\partial a^2|_{a=a^*} < 0$, i.e., the estimate a^* maximizes the log-likelihood function.

The diffusive power of the reverberant tail, or variance σ^2, can be estimated in a similar manner. Differentiating the log-likelihood function Eq. (5) with respect to σ, we have

$$s_\sigma(\sigma;\mathbf{y},a) = \frac{\partial \ln L(\mathbf{y};a,\sigma)}{\partial \sigma} \quad (9)$$

$$= -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_{n=0}^{N-1} a^{-2n}y(n)^2, \quad (10)$$

which achieves an extremum when

$$\sigma^2 = \frac{1}{N}\sum_{n=0}^{N-1} a^{-2n}y(n)^2. \quad (11)$$

As before, it can be shown that the $\mathbf{E}[s_\sigma] = 0$. Denote the zero of the score function s_σ, and satisfying Eq. (11), by σ^*. It can be shown that the second derivative $\partial^2 \ln L(\mathbf{y};a,\sigma)/\partial \sigma^2|_{\sigma=\sigma^*} < 0$, i.e., the estimate σ^* maximizes the log-likelihood function. Note that the maximum-likelihood equation given by Eq. (8) is a transcendental equation and cannot be inverted to solve directly for a^*, whereas the solution of Eq. (11) for σ^* is direct. Bounds on the variance of the estimates are presented in the Appendix.

C. Algorithm for estimating decay rate

Given an estimation window length and the sequence of observations $y(n)$ in the window, the zero of the score function Eq. (8) provides an estimate of a. The function is a transcendental equation that must be solved numerically using an iterative procedure. However, the estimate of σ can be obtained directly from Eq. (11). A two-step procedure was followed: (1) an approximate solution for a^* from Eq. (8) was obtained, and (2) the value of σ^* was updated from Eq. (11). The procedure was repeated, providing successively better approximations to a^* and σ^*, and so converging on the root of Eq. (8).

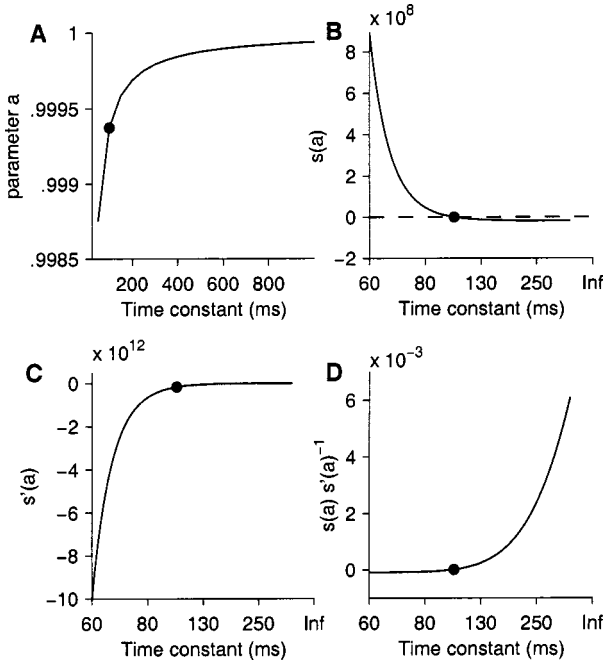


FIG. 2. Maximum-likelihood estimation (MLE) of room decay rate. (a) The decay rate of the exponential decay (τ, abscissa) is mapped to a parameter $a = \exp(-1/\tau)$ (ordinate) where τ is given in sampling periods. The function is monotone but highly compressive and maps $\tau \in [0,\infty)$ onto $a \in [0,1)$. Filled circle shows $\tau = 100$ ms ($a = 0.9994$). (b) Score function (derivative of log likelihood function) $s_a(a)$ (ordinate), decreases rapidly as a function of a [abscissa, marked in time constants using the map in (a)]. MLE of a is given by the root of $s(a)$ (filled circle). (c) The derivative $s_a'(a)$ as a function of a. At the root of s_a (filled circle), the derivative is negative. Note the nearly 8–12 orders of magnitude change in s_a and s_a' for commonly encountered values of τ. (d) The ratio $s_a(a)/s_a'(a)$ (ordinate) as a function of a is the incremental step size of the Newton–Raphson procedure for finding the root of Eq. (8). It provides an estimate of the convergence properties of the root-finding algorithm. Sampling frequency was 16 kHz, and the log-likelihood function was calculated assuming a 400-ms window.

Here we address the strategy for extracting the root in the smallest number of iterative steps. To gain an understanding of the root-solving procedure, we consider the example shown in Fig. 2. The function $a = \exp(-1/\tau)$ maps the room decay rate τ one-to-one and onto a as shown in Fig. 2(a). For instance, consider a room decay rate of 0.1 s and a sampling rate of 16 kHz. Then the decay rate is 1600 samples, and so $a = 0.9994$ (filled circle). The significance of the number becomes clear if we consider that when $\tau = 0.03$ s, then $a = 0.9979$, whereas for $\tau = \infty$, $a = 1$. Hence the geometric ratio is highly compressive and values of a for real environments are likely to be close to 1. Thus, the advantage of estimating a rather than τ is due to the bounded nature of a. The score function s_a from Eq. (7), on the other hand, has a wide range [about eight orders of magnitude, see Fig. 2(b)] and is zero at the room decay rate (filled circle). The gradient of the score function ds_a/da shown in Fig. 2(c) also demonstrates a wide range, but takes a negative value at the zero of s_a.

Thus, if we start with an initial value of $a_0^* < a$, the root-solving strategy must descend the gradient sufficiently

rapidly. The standard method for solving this kind of nonlinear equation, where an explicit form for the gradient is available, is the Newton–Raphson method which offers second-order convergence (Press *et al.*, 1992). The order of convergence can be assessed from $s_a(ds_a/da)^{-1}$ which is the incremental step size Δa in the iterative procedure [Fig. 2(d)]. For example, with true value of $\tau=100$ ms, Δa at intermediate values in the iteration can be as small as 10^{-6} when $a=0.9993$ ($\tau=90$ ms) or $a=0.9995$ ($\tau=120$ ms). This corresponds to an incremental improvement of about 0.01 ms for every iteration, thus providing slow convergence if the initial value is far from the zero. On the other hand, the bisection method (Press *et al.*, 1992) guarantees rapid gradient descent but works poorly in regions where the gradient changes relatively slowly (such as near the true value of a). Furthermore, it guarantees only first-order convergence.

However, the specific structure of the root-solving problem can be exploited because the behavior of s_a is known. Here, both methods were used to obtain rapid convergence to the root. First, the root was bisected until the zero was bracketed, after which the Newton–Raphson method was applied to polish the root. For the example shown, the root bracketing was accomplished in about eight steps and the root polishing in two to four steps. In contrast, with the same initial conditions, the Newton–Raphson method took about 500 steps to converge. Taken together, the analysis presented here suggests that the estimation procedure is feasible and does not lead to significant errors although values of a for real rooms are close to 1, and the score function and its derivative vary over many orders of magnitude. While other root-solving procedures are possible, such as iterative gradient optimization, these are not dealt with here.

D. Strategy for assigning the correct decay rate from the estimates

The theory presented in the preceding section provides one estimate of a and σ in a given time frame of N samples. By advancing the frame as the signal evolves in time, a series of estimates a_k^* will be obtained, where k is the time frame. Some of these estimates will be obtained during a free decay following the offset of a sound segment (correct estimations), whereas some will be obtained when the sound is ongoing (incorrect estimations due to model failure). Thus, a strategy is required for selecting only those estimates that correctly represent regions of free decay and hence the real room decay rate. This requires a decision-making strategy that examines the distribution of the estimates after a sufficient number of frames have been processed, and makes a decision regarding the true value of the room decay rate.

In a blind estimation procedure the input is unknown, and so the model will fail when (1) an estimate is obtained in a frame that is not occurring during a free decay. This includes regions where there is sound onset or sound is ongoing. In these periods, the MLE scheme can provide widely fluctuating or implausible estimates due to model failure. (2) The model will also fail during a region of free decay initiated by a sound with a gradual rather than rapid offset. In this case, the offset decay of the sound will be convolved with the room response, prolonging the sound even further and, so, the estimated decay rate will be larger than the real room decay rate. Gradual offsets occur in many natural sounds, such as terminating vowels in speech. We address both issues here and provide a strategy for selecting the correct room time constant.

In the first case where the estimation frames do not fall within a region of free decay, many of the time frames will provide estimates of a close to unity (i.e., infinite τ) or implausible values. On the other hand, the estimates will accurately track the true value when a free decay occurs. Intuitively, a strategy for selecting a from the sequence a_k^* is guided by the following observation: the damping of sound in a room cannot occur at a rate *faster* than the free decay, and thus all estimates a^* must attain the true value of a as a lower bound. The bound is achieved only when a sound terminates abruptly, upon which the model conditions will be satisfied, and the estimator will track the true value of the decay rate.

Although it seems intuitive to set $a=\min\{a_k^*\}$, it should be recognized that even during a free decay the estimate is inherently variable (due to the underlying stochastic process), and so selecting the minimum is likely to underestimate a.

A robust strategy would be to select a threshold value of a^* such that the left tail of the probability density function of a^*, $p(a^*)$, occupies a prespecified percentile value γ. This can be implemented using an order statistics filter specified by

$$a=\arg\left\{P(x)=\gamma:P(x)=\int_0^x p(a^*)da^*\right\}. \tag{12}$$

For a unimodal symmetric distribution with $\gamma=0.5$ the filter will track the peak value, i.e., the median. Order-statistics filters play an important role in robust estimation, especially when data is contaminated with outliers (Pitas and Venetsanopoulos, 1992), as is the case here. It should be noted that for γ values approaching 0, the filter Eq. (12) performs like the minimum filter $a=\min\{a_k^*\}$ suggested above.

In the second case described above, where the sound offset is gradual, $p(a^*)$ is likely to be multimodal because sound offsets (such as terminating phonemes in speech) will have varying rates of decay, and their presence will give rise to multiple peaks. The strategy then is to select the first dominant peak in $p(a^*)$ when a^* is increasing from zero (i.e., left most peak), that is,

$$a=\min\arg\{dp(a^*)/da^*=0\}, \tag{13}$$

where the minimum is taken over all zeros of the equation. If the histogram is unimodal but asymmetric, the filter tracks the mode and resembles the order-statistics filter.

In connected speech, where peaks cannot be clearly discriminated or the distribution is multi-modal, Eq. (12) can be employed by choosing a value of γ based on the statistics of gap durations. For instance, if gaps constitute approximately 10% of total duration, then $\gamma=0.1$ would be a reasonable choice. A judicious choice of γ can result in the filter per-

forming like an edge detector, because it captures the transition from larger to smaller values of the time-evolving sequence a_k^*.

The decision strategies, as depicted in Eqs. (12) and (13), were used to validate the model in simulated and real environments (see Sec. IV).

III. EXPERIMENTAL METHODS

In addition to simulations, the MLE approach was validated with real room data. The experimental methods and data analysis procedures are described in the following sections.

A. Sound recordings

To validate the MLE method, sound recordings were made in several rooms, building corridors and an auditorium, with the aim of determining their reverberation times. Sound stimuli that were used included 18-tap maximum length sequences (period length of $2^{18}-1$), clicks (100 μs), handclaps, word utterances (International Phonetic Assoc., 1999), and connected speech from the Connected Speech Test (CST) corpus (Cox *et al.*, 1987). Recordings were made using a Sennheiser MK-II omni-directional microphone (frequency response 100–20 000 Hz). Microphone cables (Sennheiser KA 100 S-60) were connected to the XLR input of a portable PC-based sound recording device (Sound Devices USBPre 1.5). The recorder transmitted data sampled at 44.1 kHz to a laptop computer (Compaq Presario 1700, running Microsoft Windows XP) via a USB link. The sound stimuli, stored as single-channel presampled (44.1 kHz) WAV files, were played through the headphone output of the laptop, amplified by a power amplifier (ADCOM GFA-535II) and presented through a loudspeaker (Analog and Digital Systems Inc., ADS L200e). Data acquisition and test material playback were controlled by a custom-written script in MATLAB (The MathWorks, Inc.) using the Sound PC Toolbox (Torsten Marquardt).

B. Measurement of T_{60} time using Schroeder's method

To validate the estimation procedure, experimentally recorded data from real listening environments were processed using the MLE procedure and compared to results obtained from a widely used method of Schroeder (1965a). Experimentally, RT is determined from decay curves obtained by radiating sound into the test enclosure. The sound source is switched on, and when the received sound level reaches a steady state, it is switched off. The decay curve is the received signal following the cessation of the sound source, according to the Interrupted Noise Method (ISO 3382, 1997). When the excitation signal is a noise source, the decay curve will be different from trial to trial due to random fluctuations in the signal, even when the experimental conditions are unchanged. This is in part due to the random phase and amplitudes of the normal modes at the moment the excitation signal is turned off. Prior to Schroeder's method, fluctuations in RT estimates were minimized by averaging the RTs obtained from many decay curves. Schroeder (1965a) developed an alternate method that, in a single measurement, yields the average decay curve of infinitely many interrupted noise experiments. Thus, Schroeder's method eliminates the averaging procedure.

Following Schroeder (1965a), let $n(t)$ be a stationary white noise source with power σ^2 per unit frequency, and $r(t)$ be the impulse response of the system consisting of the receiver, transmitter, and the enclosure. Then a single realization of the decay curve $s(t)$ from the interrupted noise experiment is given by

$$s(t)=\int_{-\infty}^{0} n(\tau)r(t-\tau)d\tau, \tag{14}$$

where the noise is assumed to be switched off at $t=0$, and the lower limit is meant to signify that sufficient time elapsed for the sound level to reach a steady state in the enclosure before it was switched off. The reverberation time is obtained from the decay curve $s(t)$ (see below).

In practice, the experiment was repeated to obtain a large number of decay curves, and RTs from these curves were averaged. Schroeder used Eq. (14) to establish a relationship between the mean squared average of the decay curve $s(t)$ and the impulse response of the enclosure $r(t)$, namely,

$$\mathbf{E}[s^2(t)]=\sigma^2\int_{t}^{\infty} r^2(\tau)d\tau. \tag{15}$$

While the ensemble average on the left-hand side requires averaging over many trials, the right-hand side requires only a single measurement, as it is the impulse response of the enclosure plus receiver and transmitter.

Schroeder's method, called the Integrated Impulse Response Method (or sometimes, Backward Integration Method), can be applied to a single broadband channel (say an impulsive sound covering a broad range of frequencies) or to a narrow-band channel consisting of a filtered impulse (such as a pistol shot). The only requirement is that the power spectrum of the excitation signal [in Schroeder's method, right-side of Eq. (15)] should be identical to the power spectrum of the noise burst [in the noise decay method, left-side of Eq. (15)].[1]

The recorded data were filtered offline in ISO one-third octave bands (21 bands with center frequencies ranging from 100 to 10 000 Hz) using a fourth-order Type II Chebyshev band-pass filter with stopband ripple 20-dB down. The output from each channel was processed by the MLE procedure and Schroeder's method using Eq. (15). For broadband estimation, the microphone output was processed directly using the two methods.

Due to the limited dynamic range of sounds in real environments, Schroeder's method requires the specification of a decay range. The decay ranges normally used are from -5 to -25 dB (20-dB range), and from -5 to -35 dB (30-dB range). The decay curves in each range were fitted to a regression line using a nonlinear least squares fitting function (function nonlinsq provided by MATLAB). The fitted function was of the form Aa_d^n, where A is a constant, n is the sample number within the decay window, and a_d is the geometric

ratio related to the decay rate τ_d of the integrated impulse response curve by $a_d = \exp(-1/\tau_d)$. This is in contrast to the model depicted in Eq. (2) which assumes an exponentially decaying envelope with decay rate τ, whereas Schroeder's decay curve is obtained by squaring the signal. Hence, $\tau_d = \tau/2$. Two estimates of the decay rate were obtained from decay curves fitted to the −5- to −25-dB and −5- to −35-dB drop-offs. For each fit, the line was extrapolated to obtain T_{60} time (in seconds) using the expression

$$T_{60} = \frac{6}{\log_{10}(e^{-1})\log_e(a_d)} = \frac{-6\tau_d}{\log_{10}(e^{-1})} = 13.82\tau_d. \quad (16)$$

The same procedure was followed for determining the decay rate from broadband signals. It should be noted that the MLE procedure does not require the specification of a decay range, but only the specification of the estimation window length; thus, only one estimate per band is obtained.

C. Verification of MLE procedure with ideal stimuli

Microphone data were processed using the MLE procedure to obtain a running estimate of the decay rate. For model verification, estimation was performed on (1) the segment following the cessation of a maximum-length sequence or a hand-clap, and (2) the entire run of a string of isolated word utterances. These were considered ideal stimuli, because they fulfilled the model assumptions of free decay or possessed long gaps between sound segments. The estimates were binned for each run and a histogram was produced. The histogram was examined for peaks, and the decay rate was selected using the order-statistics filter Eq. (13) if there were multiple peaks, or Eq. (12) if the histogram was unimodal. The estimate $\hat{a}$ so obtained was used to calculate T_{60} (in seconds) using the formula

$$T_{60} = \frac{3}{\log_{10}(e^{-1})\log_e(\hat{a})} = \frac{-3\tau}{\log_{10}(e^{-1})} = 6.91\tau. \quad (17)$$

In theory, the T_{60} expressions given by Eqs. (16) and (17) are identical due to the relationship between τ and τ_d. However, the calculated values may differ, and this can be ascribed to either model inadequacies or discrepancies in measurement and analysis.

D. Verification of MLE procedure for speech

The performance of the MLE was also verified using connected speech played back in a circular building foyer (6-m diameter). Test materials were connected sentences from the CST corpus. Estimates from nonoverlapping 1-s intervals were binned to yield a histogram, and the first dominant peak from the left of the histogram was selected to determine the room decay rate. The procedure for calculating T_{60} time followed Eq. (17).

IV. RESULTS

The estimation procedure was applied to a variety of data sets, including simulated data and real room responses. To illustrate the methods and identify the strengths and deficiencies of the estimation procedure, we first consider simulated data sets. Subsequently we will provide results for real data that validate the room decay rate estimates, and compare these to results from Schroeder's method.

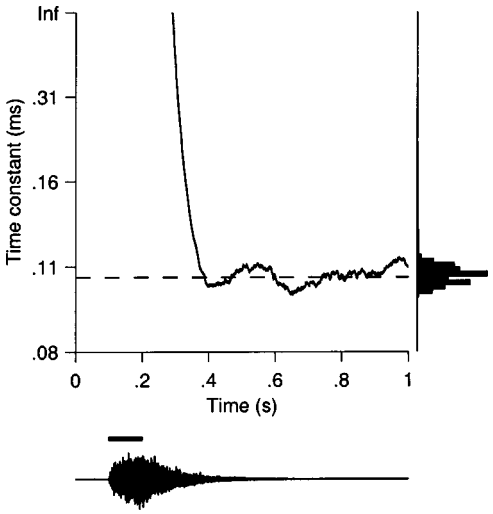


FIG. 3. Illustration of procedure for continuous estimation of decay rate. A burst of white noise was applied at time $t = 0.1$ s (black bar, bottom trace, 100-ms duration). Simulated room output (bottom trace) shows the buildup and decay of sound in the room. A running estimate of the parameter a in 200-ms windows is shown in the graph (ordinate, a converted to decay rate in seconds). The true value of decay rate (100 ms) is shown as horizontal dashed line. The estimation window was advanced by one sample to obtain the trace, with each point at time t being the estimate in the window $(t - 0.2, t]$. During the buildup and ongoing phase of the sound, estimated a sometimes exceeded 1 (i.e., negative values of τ). These were discarded and are not shown. As the window moved into the region of sound decay ($t > 0.3$ s), the estimates converged to the correct value. A histogram of the estimated decay rate is shown to the right of the trace. An order-statistics filter, such as the mode of the histogram, can be used to extract the room decay rate. Sampling rate was 16 kHz.

A. Broadband white noise bursts in simulated rooms

A 100-ms burst of broadband white noise (8-kHz bandwidth) was convolved with the impulse response of a simulated room having a decay rate $\tau = 100$ ms (Fig. 3). Room output (bottom trace of Fig. 3) shows the characteristic rise and decay of sound following onset and offset of noise burst (horizontal bar). The graph shows the running estimate of decay rate obtained in a 200-ms time window by advancing every sample. Time frames up to about $t = 0.3$ s are not regions of free decay, and so the estimator tended to produce values of $a > 1$. When this was observed in the root-bracketing step of the estimate, the root-solving procedure was aborted. Thus all estimates of a were bounded above by 1. It can be seen that when the window crosses into the region of free decay, the estimator output stabilizes at the true value (horizontal dashed line). A histogram of the decay rate estimates (right axis) was input to the order statistics filter Eq. (12) with $\gamma = 0.5$. The reported decay rate from the filter was $\tau = 101$ ms.

For comparison, the procedure was repeated with the simulated noise burst input (i.e., before it was convolved with the room impulse response) to mimic anechoic conditions. The histogram of a^* demonstrated a strong peak at $a = 1$ $(\tau = \infty)$ (not shown). This showed that in the absence of

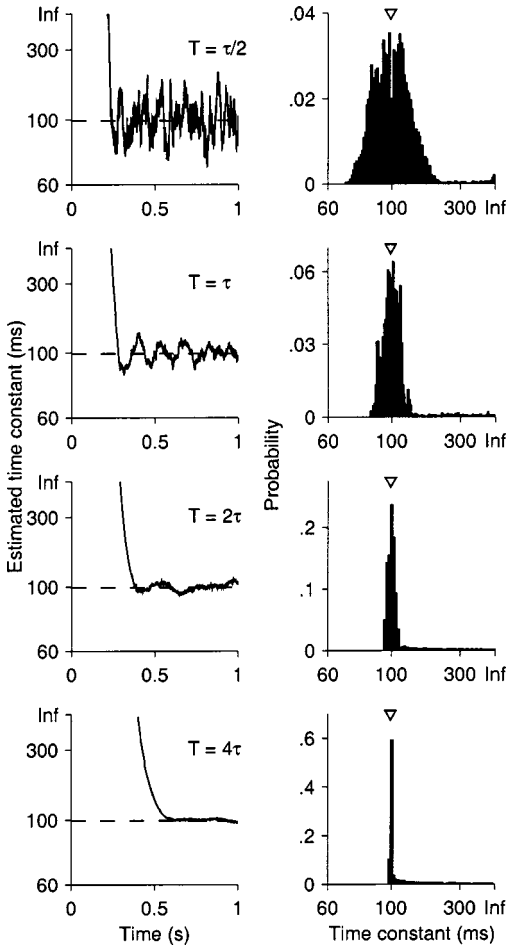


FIG. 4. Effect of estimation window length on the variance of the estimate. The simulation shown in Fig. 3 was repeated for windows of duration 0.5τ, τ, 2τ, and 4τ (top to bottom), where $\tau=100$ ms is the true value of the room decay rate. The left column shows the running estimate of parameter a (ordinate, shown as decay rate in ms) as a function of time (abscissa). The right column shows the histogram of the estimates. The variance of the estimate decreases with increasing window length (arrowheads mark true value of τ).

reverberation, as in an anechoic environment or open space, histograms showing strong peaks at $a=1$ are to be expected.

B. Effect of window length on estimation

A parameter that is critical for estimation performance is the window length N specified in Eq. (8). Small window lengths are expected to increase the variance of the estimate, as also indicated by the Crámer–Rao lower bound [Eq. (21)]. To test the effect of window length a burst of white noise (100-ms duration) was convolved with a simulated room impulse response ($\tau=100$ ms), and the estimator tracked the decay curve using four different window lengths. The results are shown in Fig. 4. As window length increased from 0.5τ to 4τ, the MLE procedure gave improved estimates. Further, for all four window lengths, there was no bias in the estimates of the peak position. We concluded that increasing window length reduced the variability in the estimates, and did not introduce significant bias.

Although it is desirable to have long window lengths, in practice this is limited by the duration and occurrence of gaps between sound segments. Ideally the filter length should be of the order of τ or longer, but if the gaps are short, then increasing the filter length beyond the mean gap will produce undesirable effects where the next sound segment creeps into the window. Thus, the window length should not be less than one-half or one-third of τ, but the upper limit is dictated by the mean duration of gaps.

C. Speech sounds in simulated room

The examples considered above illustrated the performance of the algorithm when the input was broadband white noise. To be applicable in realistic conditions, the algorithm must perform in a variety of conditions and with different signal types. Speech represents an example where the algorithm is expected to perform poorly, because it is nonstationary and non-Gaussian. Further, the offset transients in speech sounds (including plosives) have a natural decay rate (not to be confused with the room decay rate) that can vary from 5–40 ms.[2] Thus, estimation of room decay rate with speech presents a challenge to the algorithm. We took a sequence of 15 distinct and isolated American-English words recorded in an anechoic environment at a sampling rate of 20 kHz (International Phonetic Assoc., 1999). These included 11 consonant–vowel–consonant words (/p,b,g/V/d/, e.g., ''bed''), and four consonant–vowel words (/b/V/, e.g., ''bay'') separated by a mean interval of 200 ms. These were convolved with a simulated room impulse response having decay rate $\tau=100$ ms. The task of the estimator was to track the decays for the entire duration of the sequence (approximately 11.4 s). The control condition was the clean input (i.e., anechoic). The results are shown in Fig. 5. Four different filter lengths were used as in Fig. 4. For the control condition (left column) no reliable estimates were produced for the smallest three windows (top three panels) because the histogram peaked at values of τ approaching ∞. For the simulated room response (right column), the peak shifted towards the true value of τ, with the best estimates being obtained for the largest window size of 4τ (right column, bottom row). In all the histograms the peak was located at about 115 ms (arrow). This estimate deviated from the real decay rate of 100 ms due to the lack of sharp transients in the clean speech. A gradual sound offset tends to prolong the reverberated sound even further. This can be seen in the ''anechoic'' control condition where a small peak is noticeable when window size is 4τ (bottom panel, left column). The peak occurs around 60 ms, and corresponds to the gradual offsets of speech sounds. Thus, this introduces a bias in the estimates under reverberant conditions.

The results of the preceding sections demonstrate the importance of selection of a suitable estimation window length. The choice of window length determines the variability of the estimates, and is critical to obtaining a histogram with a clearly resolved peak at the true value of the room decay rate. However, the effect of variability on the order-statistics filter is difficult to determine as the filtering operation is nonlinear. Further, bimodal or multimodal histograms may be obtained if there is fluctuating background noise or if the sound segments have an intrinsic offset decay rate (as shown in Fig. 5).

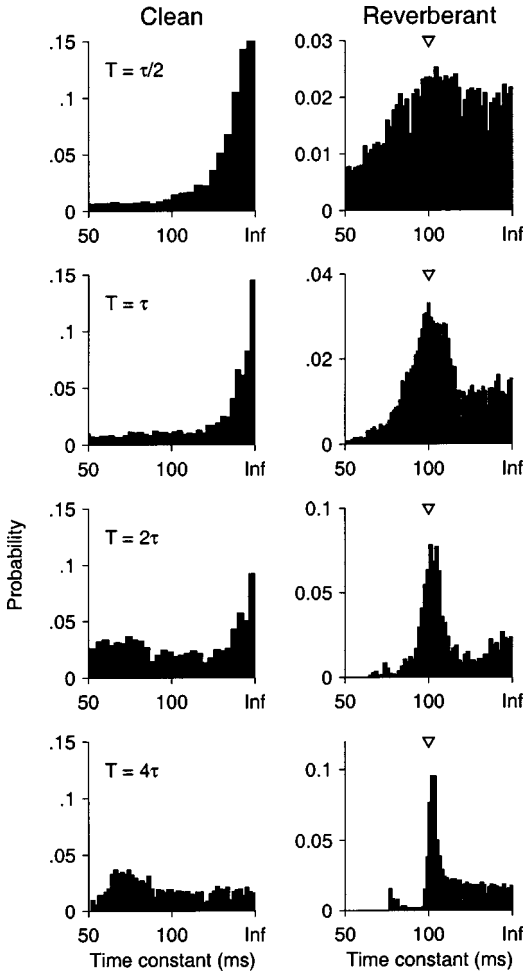


FIG. 5. Estimation of room decay rate from speech. Fifteen words recorded in an anechoic (clean) environment (200-ms interword spacing) were convolved with a room model (τ=100 ms). Histograms of decay rates were estimated from clean (left column) and simulated reverberant responses (right column), and are shown for window durations 0.5τ, τ, 2τ, and 4τ (top to bottom). The histogram for clean speech served as a control. Description follows Fig. 4. Estimation from reverberant speech produces a clearly defined peak, especially for the longer window lengths, albeit with a small bias (right column, 2τ and 4τ). The bias can be attributed to the gradually decaying offsets inherent in speech so that the resultant decay is a convolution of speech offset and the room response. For the control condition (left column), the offset decay is visible only in the bottom two rows where the histogram exhibits a broad bump between 50 and 100 ms. The 15 words included 11 /p,b,g/V/d/ and 4 /b/V/ sampled at 20 kHz.

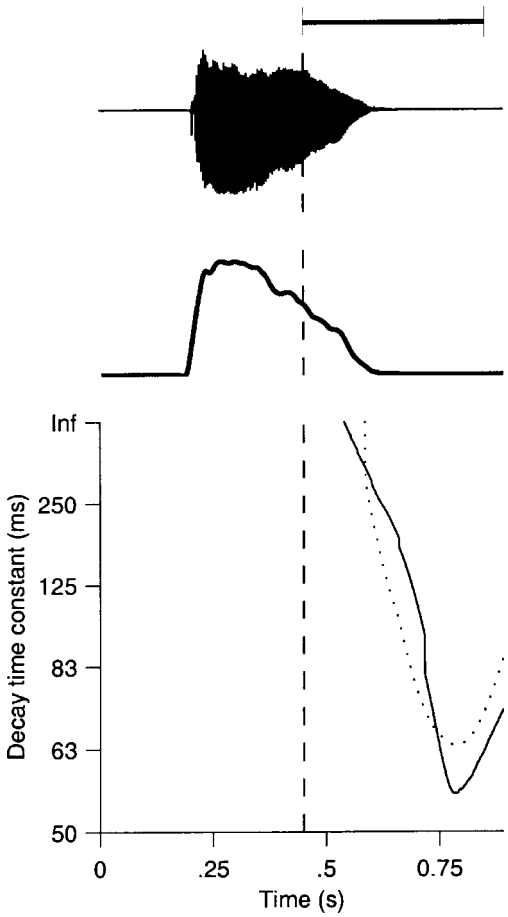


FIG. 6. Illustration of decay rate estimation when a terminating phoneme is encountered. The word ''bough'' recorded under anechoic (clean) conditions (top row) has a gradually decaying offset. The envelope was extracted by filtering the absolute value of the analytic signal (second row, heavy outline), and its decay rate was estimated for the segment following the dashed line using two methods (bottom row). Overlapping segments (duration given by bar, with step size indicated by the thickness of the vertical end) were converted to a decibel scale and the decay rate obtained by a least squares fit to a straight line (dotted trace). The same segments were analyzed using the MLE algorithm to obtain a second estimate of the decay rate (solid trace). While the estimators provide somewhat different results, they are in qualitative agreement. Both methods suggest that the fastest decay rate is in the range of 50–70 ms (see also Fig. 5).

D. The effect of gradual offsets in speech sounds on decay rate estimation

The preceding section introduced the problem of estimating the room decay rate when the input signal exhibited varying offset decay time courses. Here we examine in greater detail the performance of the estimator with input comprising a single word (/b/V/, ''bough''). The word was recorded under anechoic conditions and presented to the estimator without modification so that the effect of the vowel offset could be determined. The results are shown in Fig. 6. The terminating vowel has a gradually decaying offset (top panel). Estimation of the offset decay was performed from $t=0.45$ s (vertical dashed line) using two procedures. First, the envelope was extracted from the analytic signal via a Hilbert transform, windowed, and filtered to eliminate frequency components above 100 Hz. The envelope is shown in the middle panel (heavy outline). The envelope was then squared and transformed to a decibel scale, and the decay rate was estimated in windows of duration 0.4 s (horizontal bar), using a least squares fit to a straight line. Successive estimates were obtained by sliding the window forward in steps of one sample. Note that the time at which an estimate is reported for any given window is the end point of the window. The estimate for the window indicated by the horizontal bar, for instance, is plotted at time $t=0.85$ s. A curve of the estimated decay rates was thus obtained (dotted curve, bottom panel). The envelope-based method employed here is similar to the method of Lebart *et al.* (2001), except that they performed a one-time RT estimation over the entire decay period using linear regression. The MLE procedure was applied to the same segments and produced an independent estimate of the decay rate (solid line, bottom panel). While the estimates differ somewhat, they are in qualitative agree-

ment. Both procedures indicate that the terminating vowel had a time-dependent decay rate, and the greatest rate was between 50 and 70 ms.

The results confirm the presence of the peak in Fig. 5 (left column, bottom panel), although the histogram shown in Fig. 5 was obtained for a sequence of 15 words. The analysis shown in Fig. 6 also indicates the reason for estimation bias under reverberant conditions using speech samples. The offset decays present in clean speech segments will be convolved with the room impulse response, and the estimated decay rates will consequently be greater than the room decay rate. Taken together, the results from Figs. 4–6 suggest that the factors responsible for estimation performance are the presence of adequate numbers of gaps, sharp offset transients, and estimation window length.

E. Validation of method

The above results demonstrate that estimation of decay rate is possible for a variety of sounds including impulses, noise bursts, and speech. Although we have shown that a reasonable agreement exists with a nonlinear least squares fit to the data (Fig. 6), a more careful evaluation is necessary to determine the conditions under which the MLE procedure is likely to provide accurate estimates. Here we establish that the estimated decay rates are comparable to decay rates obtained from the method by Schroeder (1965a). Furthermore, any data collected must be under sufficiently realistic conditions where there is background noise and where the testing sound is not subject to experimental control. A comparison of MLE performance with the standard method in real environments will therefore establish the utility of the method.

We compared the estimates using the method by Schroeder (1965a) in both single-channel (i.e., the broadband signal), and multi-channel frameworks (i.e., narrow-band signals occupying ISO one-third octave bands). Schroeder's method requires a fitting procedure to estimate the decay rate in a preselected decay range (either 20 or 30 dB below a reference level of −5 dB, see Sec. III). The MLE procedure does not require the specification of such a range.

To determine whether the two methods provide the same RT value, estimations were performed on a simulated room decay curve with RT=0.5 s (Fig. 7). Broadband and one-third octave band estimates were obtained using the MLE method (circle) and Schroeder's method (20 dB: lozenge, 30 dB: square). Figure 7(a) shows the mean value of RT as a function of center frequency of the one-third octave bands (open symbols) and the broadband estimate (filled symbols near *y* axis). range) averaged over 100 trials. The broadband estimates were 0.504 s (MLE) and 0.5 s (Schroeder's method) for both 20- and 30-dB decay ranges. While the MLE estimate was significantly different from Schroeder's method ($p<0.0001$, Wilcoxon rank sum test), the discrepancy was less than 1%. The one-third band MLE estimates in most cases were somewhat higher than the Schroeder estimates by about 0.5% (mean RT over all bands were, MLE: 0.505, Schroeder's method: 0.502 s for 20 dB and 0.501 s for 30 dB). However, the estimates were not significantly different, except for one estimate obtained from the 30-dB decay curve in the band centered at 8 kHz. The most noticeable

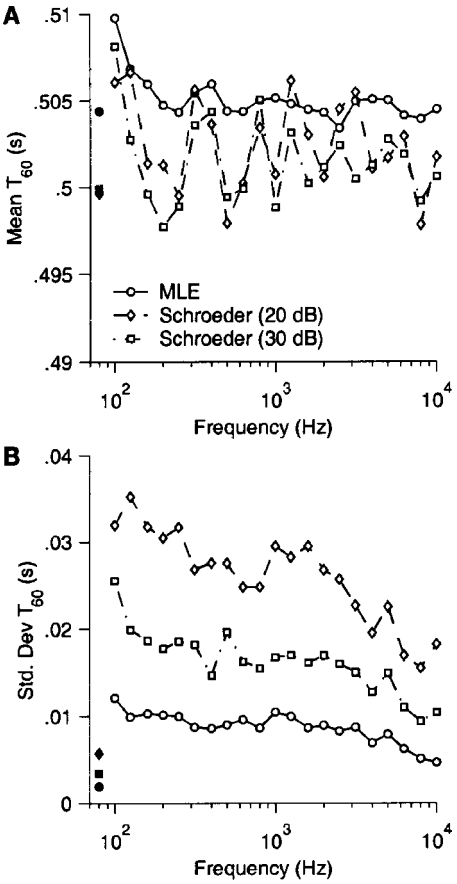


FIG. 7. Comparison of RT estimates obtained from MLE method and Schroeder's method. (a) Mean RT (ordinate) in one-third octave bands (abscissa) averaged over 100 independent trials of a simulated decay curve (RT=0.5 s). RT estimates were obtained using the MLE procedure (circles), and Schroeder's method in 20-dB decay range (lozenge), and 30-dB decay range (square). The filled symbols are broad-band estimates. (b) Standard deviation of the RT for broadband and one-third octave bands over 100 trials. Symbols follow (a).

difference between the two methods was in the variability of the estimates as measured by the standard deviation over the trials [Fig. 7(b)]. The MLE method demonstrated lower SD across trials than Schroeder's method, by factors of 2 (for the 20-dB curve) and 3 (for the 30-dB curve). Further, MLE estimates were similar across one-third octave bands at frequencies above 200 Hz [Fig. 7(a)], whereas estimates from Schroeder's method exhibited greater variability. The results establish that the MLE method and Schroeder's method are in good agreement when tested on model data. While the MLE method may overestimate the RT when using broadband signals (although this is no more than 1%), the narrow-band estimates are comparable to those obtained from Schroeder's method, are consistent over a wide range of frequencies, and subject to less variability.

We first report on the comparison between the methods using a hand-clap in a small office (8×3×3 m). Subsequently we will summarize results obtained in rooms of different sizes. Figures 8(a) and (b) depict a hand-clap event and its spectrogram, respectively. The data in panel (a) is the same as shown in Fig. 1(a), except that Fig. 8(a) also includes the direct sound. The rms noise level in the room was 50 dBA SPL, and the peak sound pressure level resulting

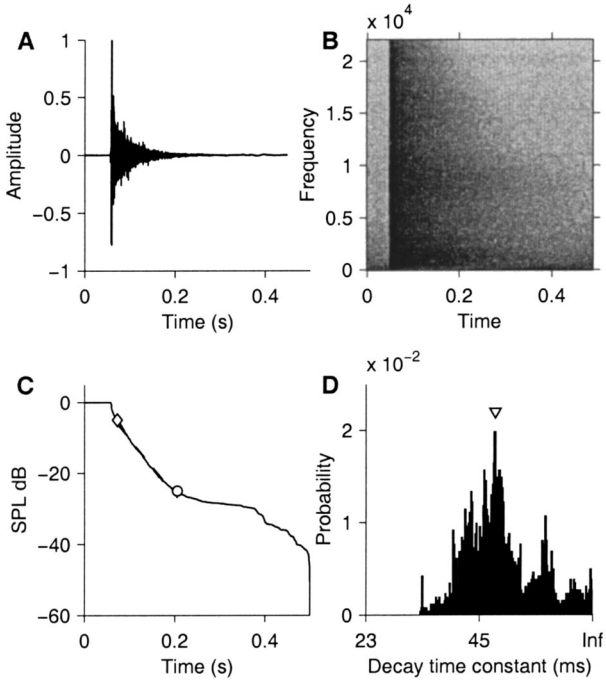


FIG. 8. Estimation of decay rate from real room data. (a) The room response to a hand-clap [same as Fig. 1(a) but includes the direct sound]. (b) Spectrogram of the hand-clap demonstrates a sharp broadband onset transient and the decay as a function of frequency. (c) The decay rate was estimated using Schroeder's backward impulse integration method in the −5-dB (lozenge) to −25-dB (circle) range, followed by a least-squares fit to a straight line to obtain the decay rate (τ=56 ms, T_{60}=0.39 s). (d) Histogram of decay rate obtained from signal shown in (a) using MLE. The median value of the histogram (arrow) is τ=53 ms, T_{60}=0.37.

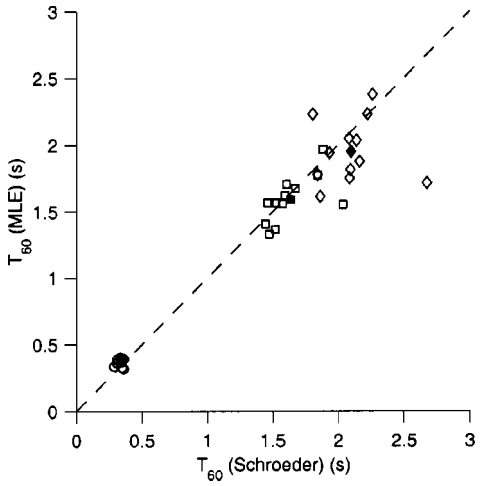


FIG. 9. Comparison of Schroeder's method and the MLE procedure for T_{60} times obtained in one-third octave bands. Three environments were tested: a moderately reverberant environment (circles; the environment is the same as shown in Fig. 8), a highly reverberant circular foyer (squares), and a highly reverberant enclosed cafeteria (diamonds). In each environment, a single hand-clap was filtered using a bank of ISO one-third octave band-pass filters with center frequencies exceeding 1 kHz. The ordinate shows the best estimates obtained from the MLE procedure for each band, and the abscissa shows the T_{60} times obtained from Schroeder's method. Averages over all bands for each environment are shown as filled symbols. The diagonal dashed line (with unity slope) is shown for reference, and points lying close to this line suggest good agreement between the two procedures.

from the hand-clap was 85 dBA SPL. The decay curve obtained using Schroeder's method is shown in Fig. 8(c), normalized so that the peak SPL was 0 dB. This is the broadband curve obtained by integrating the recorded microphone signal. A straight-line fit to the 20-dB drop-off point (circle) from a reference level of −5 dB (lozenge) yielded τ=56 ms (T_{60}=0.39 s). The discrepancy between this value and that presented in Fig. 1 (τ=59 ms) was due to the inclusion of the direct sound in Fig. 8. The windows over which the 20-dB drop-off was computed were not identical for the two cases. The data were run through the MLE procedure and a histogram of estimates was obtained, and the decay rate was calculated from the peak of the histogram using Eq. (12). This gave an estimate τ=53 ms (T_{60}=0.37 s), which is in good agreement with the estimate obtained using Schroeder's method. Note that the estimates reported in this work are based on a single trial. The normal practice is to average over large numbers of trials. However, our goal is to develop an online estimation procedure, and so we felt that it would be more realistic to use a single trial.

To test a range of room RTs, ISO one-third octave band analysis (exceeding 1 kHz center frequency) was performed in three environments. These were (1) the moderately reverberant room described above (Fig. 8), (2) a highly reverberant circular foyer, and (3) a highly reverberant enclosed cafeteria. In all cases, the signal was a hand-clap generated at a distance of 2 m from the recording microphone (peak value 90 dB SPL). Output from the band-pass filters were analyzed using the MLE procedure, and the τ value for each band was obtained from the histogram by selecting the dominant peak. For Schroeder's method, a 20-dB decay range was used. Figure 9 shows the T_{60} estimates from Schroeder's method (abscissa) versus the MLE estimates (ordinate) for each ISO one-third band (open symbols), and the average over these bands (closed symbols).

Figure 9 shows that the variability of estimates for highly reverberant environments increases with increasing mean RT for both methods. However, the two methods are in good agreement, especially in the high-frequency bands (the single outlier falling below the diagonal in Fig. 9 is the lowest center frequency used in the analysis, namely 1 kHz). The agreement between the methods is best when the T_{60} values are averaged over all bands (filled symbols), as is usually reported in the literature.

A more extensive test to determine the variability in estimates across different environments, and between bands, was performed in 12 environments, including small office rooms, an auditorium, large conference rooms, corridors, and building foyers. The data were analyzed as in Fig. 9 and are shown in Fig. 10(a). In comparison with Schroeder's method, the MLE procedure consistently overestimated T_{60} in low to moderately reverberant environments ($T_{60}<0.3$ s) whereas it underestimated the reverberation time for more reverberant environments ($T_{60}>1.3$ s). There was a good agreement between the two methods for intermediate ranges. The average T_{60} over all bands (filled squares) were, however, in good agreement. Broadband estimates were made using the same procedures but without band-pass filtering of recorded signals. These are shown in Fig. 10(b). The trend in the estimates was similar to that observed with narrow-band

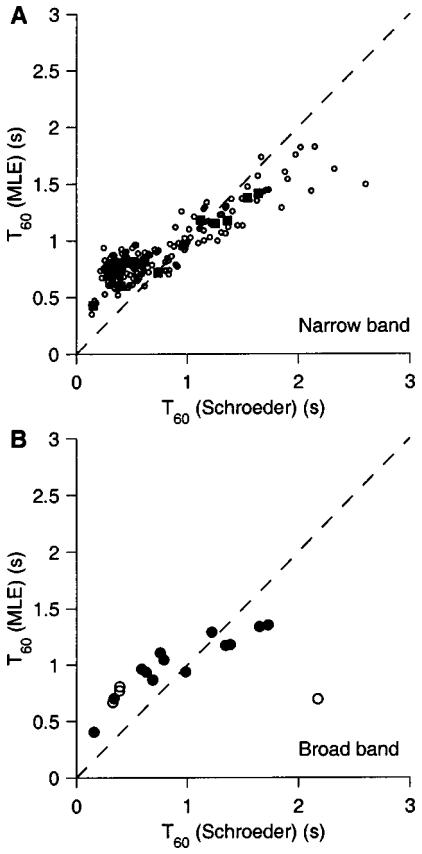


FIG. 10. Reverberation-time estimates from real environments. Seventeen tests in 12 environments were conducted using noise bursts. Decay rates were estimated using the MLE algorithm (ordinate) and the extrapolated T_{60} times were compared with estimates from Schroeder's method (abscissa). (a) Estimates of T_{60} in one-third octave bands with center frequencies exceeding 1 kHz (open circle) and their average (filled square). (b) Broad-band estimates of T_{60} from the recorded room response. Data shown by open circles are from different locations in an auditorium. Results are from one presentation of noise burst in each test.

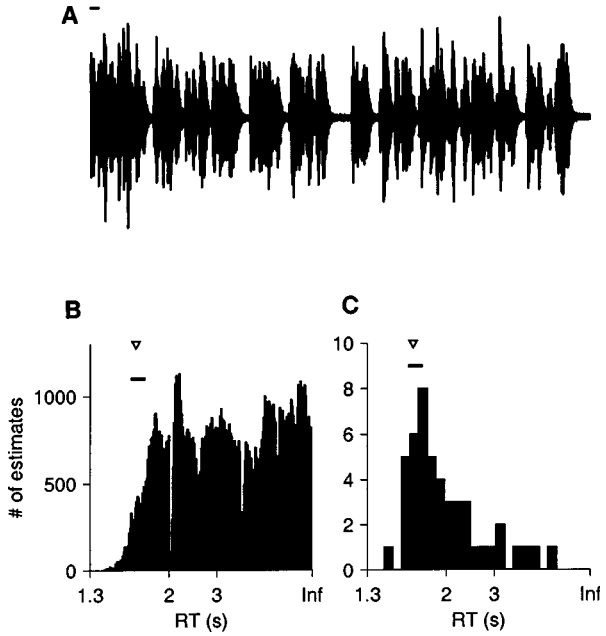


FIG. 11. Evaluation of room reverberation time (RT) from connected speech played back in a partially open circular foyer. The RT for this environment as measured from hand-claps was 1.66±0.07 s (Schroeder's method) and 1.62 s (from MLE procedure). (a) Trace of CST passage (duration 50 s) recorded in the environment. Bar indicates 1 s. (b) The histogram of MLE estimates over the duration of recording. The first peak in the aggregate histogram is the best RT estimate from connected speech (1.83 s). The horizontal bar is the range of RT estimates obtained from Schroeder's method, and the triangle indicates the MLE estimate. (c) Peak values from histogram of estimates were obtained every 1 s, and the 50 peak values were used to produce the histogram shown. The best estimate of RT from this histogram is at the dominant peak (1.7 s).

signals, except for one outlier. The outlier along with three other data points (open circles) were obtained in a large auditorium. The outlying data point was obtained at a source-to-microphone distance of 4 m, whereas the three remaining data points were obtained at a source-to-microphone distance of 1.5 m (at three different locations in the auditorium). The sound levels were not adjusted to compensate for the distance, and hence the experiment corresponding to the outlier was at a lower SPL, resulting in reduced dynamic range (from peak SPL to noise floor). For the four experiments in the auditorium, the Schroeder estimates of T_{60} (in seconds) were 2.18 (outlier), 0.39, 0.39, and 0.33, respectively. The MLE estimates, on the other hand, were 0.69 (outlier), 0.77, 0.80, and 0.67, respectively. Schroeder's method appears to be sensitive to the peak-SPL to noise-floor range, because the remaining three locations provided RT values that were in good agreement. On the other hand, the MLE estimates, while larger than the Schroeder estimates, were consistent and relatively robust irrespective of the source-to-microphone distance. That is, a reduction in dynamic range appears to affect estimates from the MLE to a lesser extent than estimates from Schroeder's method. A more detailed study is required to quantify the effect of dynamic range on the various estimation methods, and has not been attempted here.

These results raise the issue of estimation in narrow bands. It appears, although it is by no means conclusive, that the upper one-third octave bands (over 1 kHz) may provide more accurate estimates than the lower bands. Frequency decomposition is a standard part of most audio signal processing algorithms, and so it may be useful to track estimates in the higher frequency bands, or in select bands where the energy is greatest. Tracking high-energy bands is likely to provide more temporal range in tracking decays before encountering the noise floor, and thus sharpen the peak in the histogram of estimates. Alternatively, averaging over all high-frequency bands can provide estimates that are in closer agreement with T_{60} times obtained from Schroeder's method.

The above findings suggest that there is good correlation between the estimates obtained with the MLE procedure and those obtained with Schroeder's method. While Schroeder's method provides the most accurate estimates, in situations where the peak to noise-floor range is limited, the MLE method can provide robust estimation.

F. Estimation of RT from connected speech in real listening environments

The results presented in the preceding sections indicate that the MLE output is in good agreement with actual or simulated room RTs. In particular, the estimator can be ap-

plied to isolated word utterances, even though the naturally decaying offsets of terminating phonemes may lead to an overestimation of RT (see Fig. 6). Here, we test the performance of the procedure explicitly in a challenging estimation task, namely estimating room RT from connected speech.

A segment of speech (about 50 s in duration) from the Connected Speech Test (CST) corpus was played back in a partially open, circular foyer (one-third octave band analysis shown in Fig. 9, square symbols). The RT for this environment was first estimated with hand-claps using Schroeder's method (1.66 ± 0.07 s) and independently confirmed with the MLE procedure (estimated RT from histogram was 1.62 s). The MLE procedure was then applied to the recorded speech data [Fig. 11(a)]. A histogram of room decay rates for the duration of the recorded data was constructed [Fig. 11(b)]. The order-statistics filter was used to select the first dominant peak in the histogram (RT=1.83 s). This is the best RT estimate based on the aggregate data. It is possible to refine the procedure for arriving at the best estimate by applying the order-statistics filter at much shorter time intervals. Towards this end, a histogram was constructed at intervals of 1 s, and the best RT estimate for this interval was obtained. The resulting best estimates from all 1-s durations (50 in all) were binned to produce the histogram shown in Fig. 11(c). It can be seen that the number of estimates peaks at RT=1.7 s, which agrees with the mean value of 1.66 s from Schroeder's method (using hand-claps), and is well within its standard deviation [0.07 s; the one-sigma interval is indicated by the horizontal bar in Figs. 11(b) and (c)].

Given that terminal phonemes have a natural decay rate (see Fig. 6), it is not surprising that the MLE procedure produces estimates somewhat larger than the real room RT. Further, the discrepancy between the actual RT and those estimated from connected speech arise from the absence of adequate numbers, and the limited duration, of gaps [see Fig. 11(a)]. Thus, regions of free decay where estimation is accurate are limited. Not withstanding these constraints, the procedure works well, in part due to the decision-making capability built into the order-statistics filter. By selecting the first dominant peak (from the left) in the histogram, the filter in effect rejects spurious estimates, thereby reducing the error in the estimation procedure. The mean value of the histogram or its median, for instance, would result in significantly higher estimates of RT. The performance of the order-statistics filter can be further improved if one were to obtain a statistical characterization of gap duration from a large corpus of connected speech or other sounds. Such a characterization can provide a robust percentile cutoff value [see Eq. (12)] which could then be used to select the best RT value for the room (results not shown).

In conclusion, the MLE procedure, in combination with order-statistics filtering, provides a robust means for blind estimation of room RT. The procedure has been validated against Schroeder's method, and with real room data such as hand-claps, isolated word utterances, and connected speech.

V. DISCUSSION

The estimation of reverberation time is a widely investigated problem. Traditionally, two approaches have been taken. The RT is computed analytically using formulas that incorporate the geometry and absorptive characteristics of the reflecting surfaces, or empirically using a test sound with known properties that is radiated into the environment, and for which the RT is estimated from the received sounds. The former approach is embodied in the Sabine-type formulas (Sabine, 1922; Eyring, 1930; Millington, 1932; Sette, 1933; see Young, 1959; Kuttruff, 1991, for reviews), while the latter is based on the analysis of decay curves, such as using Schroeder's method (Schroeder, 1965a; Chu, 1978; Xiang, 1995). These methods have wide applicability and have been used extensively since they were developed. The current work complements these methods, and provides a technique for evaluating RT from passively received microphone signals.

In the Interrupted Noise Method the excitation signal is broadband or narrow-band filtered noise that is abruptly switched off at a known time, and is followed by a sufficiently long pause to track the decay. The reverberation experiment thus requires careful control of a known sound source (the excitation signal), and is repeated many times to arrive at an average RT estimate. In contrast, Schroeder's method eliminates multiple trials, and can be carried out with an impulsive sound, such as a pistol shot, to obtain a reliable RT estimate on a single trial. For narrow-band estimates, a filtered impulse can be used (see also footnote 1). For the interrupted noise method, the RT is determined by selecting a decay range, beginning 5 dB below the initial level at sound offset and going down a further 20 or 30 dB, taking care to remain above the noise floor. For this method, sound offset time should be known. In Schroeder's method, precise knowledge of the impulse occurrence time is not necessary, except that the decay range should be above the noise floor [see ISO 3382 (1997) for a discussion of this point]. When the impulsive sound is well-isolated, i.e., preceded and followed by a sufficiently long period of silence, and the background noise level is well outside the measured decay range, Schroeder's method will provide the best estimates of RT.

The motivation for developing the MLE procedure was to extend the utility of the decay curve method to situations where excitation signals are not available to conduct a reverberation experiment. Instead one has to rely on passively received microphone signals consisting of unknown sound segments with randomly occurring gaps. In such a blind situation, it is expected that the method will be less reliable than Schroeder's method, and so the goal was to combine a theoretically proven procedure (the maximum-likelihood approach) followed by a decision strategy that reduces the estimation error. It was hoped that such an approach would allow the estimator performance to approach that of Schroeder's method.

The MLE procedure is similar to the noise decay curve and Schroeder's methods. It differs from these methods in that it is parametric, and is based on a widely accepted model of the reverberant tail, namely the exponential decay model [see Young (1959) for a discussion on how the Sabine type formulas are related to a linear decay of the sound pressure level after the source is turned off]. The model assumes that

the amplitude of successive reflections are damped exponentially, while the fine structure is a random uncorrelated process. The random fine structure is a reasonable assumption because the excitation signal is random, and so the room output is also random. Schroeder makes this assumption explicit when developing his method (Schroeder, 1965a), arguing that the phase and amplitudes of the normal modes at the time of sound offset are unknown, and so the decaying normal modes (of different frequencies) constitute a random process even if the room response is deterministic. For most diffusive environments, this approximates the reverberant tail fairly well (see Fig. 1) and forms the central assumption of the work reported here.

The success of the MLE approach derives largely from the analytically tractable nature of the maximum-likelihood formulation, reducing the problem to the estimation of a single parameter that can be determined computationally. We also showed that for ongoing and onset segments of the sound, the estimates will assume implausible values as the model is not valid in these regions. However, an order-statistics filter downstream to the maximum-likelihood estimator can reject these estimates and extract the room RT with improved confidence. This is based on the intuitive idea that sounds cannot decay faster than the rate prescribed by the room decay rate, and thus selecting the earliest peak improves the confidence of the estimates. To our best knowledge, this MLE approach to blind decay rate estimation in enclosures has not been reported in the major acoustical literature.

The two encouraging results of this study are the validation of the estimates using Schroeder's method, and the RT estimates obtained from speech sounds. Under ideal conditions (impulsive hand-claps), the MLE method produced results that were comparable to Schroeder's method (Figs. 1 and 7), and provided motivation to carry out further tests using speech sounds. Speech sounds present particular problems to most estimation algorithms because they violate the two most commonly held assumptions, namely stationarity and Gaussian statistics. Further, even abruptly terminating phonemes such as stop consonants demonstrate a gradual decay, with a rate that may be in the range of 5–40 ms. Thus gradual offsets can increase the overall decay rate estimated in reverberant environments. However, except for the increase in estimated decay rate (a variation up to about 15% for sounds terminating in /d/), the tracking and histogram procedure works rather well, indicating that the method is relatively robust to model uncertainties.

Partially blind approaches to RT estimation have previously been described. (1) A neural network can be trained to learn the characteristics of room reverberation (Nannariello and Fricke, 1999; Cox *et al.*, 2001). Here, it is necessary to train the network whenever the environment changes. (2) The signal is explicitly segmented to identify gaps wherein decays can be tracked (Lebart *et al.*, 2001). It should be noted that the order-statistics filter developed in this work performs an implicit segmentation of the signal by rejecting estimates that are implausible. (3) A blind dereverberation procedure can be used to obtain the room impulse response. However, the room impulse response must be minimum phase, a condition that most listening environments fail to satisfy (Neely and Allen, 1979; Miyoshi and Kaneda, 1988).

The MLE procedure presented here is just one method for estimating room RT. Other methods are also possible. For instance the envelope of the sound can be extracted in the estimation interval, converted to sound pressure level, and a regression line could be fitted to obtain the T_{60} time. This is a blind version of the RT estimation procedure followed by Lebart *et al.* (2001). The order-statistics filter can be applied to the histogram of estimates as with the MLE procedure. The method is nonparametric and so is not subject to model uncertainties. This approach was used to estimate the decay rate of isolated word utterances (Fig. 6). While a detailed comparison of the methods is beyond the scope of this work, we note that the MLE procedure is a theoretically principled way of extracting the decay rate from the sound envelope.

The MLE procedure is model-based and is expected to perform reasonably well in diffuse sound fields (i.e., uniform with respect to directional distribution) and where a single decay rate describes the reverberant tail. For most sound fields this is a reasonable approximation [see Kuttruff (1991) for a discussion on this point]. The estimates of T_{60} are in good agreement with Schroeder's method in most of the listening environments tested, including challenging situations where the source or recording microphone was close to a wall, or there was moderate background noise (see Fig. 8). While the MLE procedure produces best results when there are isolated impulsive sounds or abruptly terminating white noise bursts, the results of tests with isolated word utterances and connected speech are in good agreement with the actual T_{60}. Thus, the procedure is expected to work under most listening conditions.

A result that was particularly interesting was the apparent robustness of the MLE method to reduced dynamic range of sounds (i.e., situations where the peak to noise floor range of the decay curve was small). The MLE method provided consistent estimates even when the dynamic range of sound decay was reduced. This is illustrated in Fig. 10(b) which shows the effect of source-to-microphone distance on RT estimation. The four open circles were broadband estimates obtained in an auditorium, of which one experiment (corresponding to the outlier) was at a larger source-to-microphone distance (4 m) than the remaining three (1.5 m) (see Sec. III for details). Broadly speaking, at constant sound level output from the source, the MLE method provided comparable estimates of RT including when the source-to-microphone distance ranged from 1.5 to 4 m, with reduced dynamic range of the sound decay curve. In contrast, Schroeder's method is dependent on the peak to noise floor range of the decay curve, and reducing the range can result in overestimation of the RT.[3] Consequently, increasing the source-to-microphone distance affected estimates for Schroeder's method more than those for the MLE method. The ISO recommendations for measurement using Schroeder's method specify that ''the level of the noise floor shall be at least 10 dB below the lower value of the evaluation range'' (ISO 3382, 1997). For example, if a 20-dB range is to be used, then the recommended peak to noise floor range must be at least 35 dB (including the initial 5-dB response from peak). This finding

should be interpreted with caution because the MLE method was not tested in high levels of background noise or when the dynamic range was drastically reduced. The method appears to be more robust than Schroeder's method only for the conditions tested here. To properly evaluate this effect two lines of inquiry need to be pursued: (1) quantify the effect of source-to-microphone distance on the RT estimates from the two methods, and (2) explicitly incorporate additive background noise in the MLE procedure. Incorporating background noise would require the estimation of an additional parameter, the power of the background noise. This would help to determine more precisely the relative merits of the different methods, and, in particular, to indentify situations where the MLE method can provide improvements over Schroeder's method.

The method proposed here can be expected to perform poorly when there are room resonances and the sound pressure level decays nonlinearly with time. This can be a result of the room geometry, or positioning the recording microphone in a region of the sound field that is nondiffusive, or in acoustically coupled spaces with widely differing RTs. In addition to model failure, the performance of the estimator may be poor when there are insufficient numbers of gaps, or there is fluctuating background noise. Good performance results when there are about 10% gaps and the peak sound level (at the time of offset) is about 25 dB SPL over the noise floor. Performance may also be compromised when background noise is modulated (such as with background music or babble) as the procedure will attempt to track any modulation present in the environment, and hence produce multimodal histograms with peaks that may not be easily discriminated.

The blind estimation procedure suggested here can be applied in a number of situations. Because only passive sounds are used, any audio processor that has access to microphone input can estimate the room reverberation time, either in single-channel (broadband) or multi-channel (narrow-band) mode. Further, while the method presented here is for a single microphone, it can be applied with no modifications to an array of microphones, providing several independent estimates of the RT. One of the most interesting applications is in the selection of signal processing strategies tailored to specific listening environments. These include hearing aids and hands-free telephony. Programmable hearing aids often have the ability to switch between several processing schemes depending on the listening environment (Allegro *et al.*, 2001). For instance, in highly diffusive environments, where the source-to-listener distance exceeds the critical distance, adaptive beamformers are ineffective (Greenberg and Zurek, 2001). In such situations, it would be convenient to switch off the adaptive algorithm and revert to the relatively simple (fixed) delay-and-sum beamformer. Alternatively, in highly confined listening environments such as automobile interiors, where a reflecting surface is located in close proximity to the ear, it may be convenient to switch-off the proximal ear microphone, and use the input from the microphone located in the better (more distal) ear. Such decisions can be made if there is a passive method for determining reverberation characteristics. Other potential applications could include hands-free telephony, and room acoustics evaluation in sound-level meters. A limitation of the method is its relatively poor performance with narrow-band signals whose center frequencies are below 1 kHz. However, the performance is good for broadband signals, and narrow-band signals whose center frequencies exceed 1 kHz.

The computational costs of implementing the procedure are largely due to the iterative solution of the maximum-likelihood equation. We have developed fast algorithms for reducing the computational cost so that the procedure can be implemented in real-time (Ratnam *et al.*, 2003). Thus, the method can be implemented in passive listening devices to determine the reverberation characteristics of the environment.

ACKNOWLEDGMENTS

We would like to thank the members of the Intelligent Hearing Aid Project at the Beckman Institute, University of Illinois, for their criticisms and comments at various stages of the work. The constructive comments and suggestions of the anonymous referee helped greatly in improving the clarity of the manuscript, and in providing the necessary perspective for evaluating the work with respect to Schroeder's method. The work was supported by grants from the National Institutes of Health (R21DC04840), Phonak AG, Charles M. Goodenberger Foundation, and the Beckman Institute.

APPENDIX: CRÁMER–RAO BOUNDS FOR DECAY RATE ESTIMATION

Bounds on the estimate of a and σ are obtained from the variance of the score function, also called the Fisher information J. This is more conveniently expressed in terms of the derivatives of the score functions (Poor, 1994). Given the parameter $\boldsymbol{\theta}^{\mathrm{T}}=[a\,\sigma]$ and the score function $s_{\theta}^{\mathrm{T}}(\mathbf{y};\boldsymbol{\theta})=[s_a(\mathbf{y};a,\sigma)\,s_{\sigma}(\mathbf{y};a,\sigma)]$, we have

$$J(\boldsymbol{\theta})=-\mathbf{E}\left[\frac{\partial s_{\boldsymbol{\theta}}^{\mathrm{T}}(y;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]. \tag{A1}$$

From Eqs. (7), (9), and (A1), we have

$$J(\boldsymbol{\theta})=\begin{pmatrix}\dfrac{N(N-1)(2N-1)}{3a^2} & \dfrac{N(N-1)}{a\sigma}\\[2ex] \dfrac{N(N-1)}{a\sigma} & \dfrac{2N}{\sigma^2}\end{pmatrix}. \tag{A2}$$

By the Crámer–Rao theorem (Poor, 1994), a lower bound on the variance of any unbiased estimator is simply $J^{-1}(\boldsymbol{\theta})$, which is

$$J^{-1}(\boldsymbol{\theta})=\begin{pmatrix}\dfrac{6a^2}{N(N^2-1)} & -\dfrac{3a\sigma}{N(N+1)}\\[2ex] -\dfrac{3a\sigma}{N(N+1)} & \dfrac{\sigma^2(2N-1)}{N(N+1)}\end{pmatrix}. \tag{A3}$$

From the asymptotic properties of maximum-likelihood estimators (Poor, 1994), we know that the estimates of a and

σ are asymptotically unbiased and their variances achieve the Crámer–Rao lower bound (i.e., they are efficient estimates). Thus, if a^* and σ^* are the estimates obtained from the solutions of Eqs. (8) and (11), the variance of the estimates are

$$\mathbf{E}[(a^*-a)^2] \geqslant \frac{6a^2}{N(N^2-1)}, \tag{A4}$$

$$\mathbf{E}[(\sigma^*-\sigma)^2] \geqslant \frac{\sigma^2(2N-1)}{N(N+1)}, \tag{A5}$$

with equality being achieved in the limit of large N. As the variance of a and σ are $O(N^{-3})$ and $O(N^{-1})$, the estimation error can be made arbitrarily small if observation windows are made sufficiently large.

[1]An interesting discussion on the use of excitation signals can be found in an exchange of letters between Smith (1965) and Schroeder (1965b). Schroeder uses the term ''tone-burst'' to denote a filtered impulse having a narrow-band, and this appears to have caused some confusion.

[2]The offset time courses for speech that are reported here are the results of analyzing isolated word utterances. These are the authors' unpublished observations based on a preliminary study. The offset time courses of speech segments in connected speech and connected discourse require further study.

[3]The noise floor of the integrated impulse response curve (Schroeder's method) manifests itself as a relatively flat line. When the decay range for measuring RT includes a portion of the noise floor, the estimated RT will be greater than the true value [see ISO 3382 (1997) for a discussion on this point].

Allegro, S., Buechler, M., and Launer, S. (**2001**). ''Automatic sound classification inspired by auditory scene analysis,'' Proc. European Conf. Sig. Proc., EURASIP.

Bolt, R. K., and MacDonald, A. D. (**1949**). ''Theory of speech masking by reverberation,'' J. Acoust. Soc. Am. **21**, 577–580.

Chu, W. T. (**1978**). ''Comparison of reverberation measurements using Schroeder's impulse method and decay curve averaging method,'' J. Acoust. Soc. Am. **63**, 1444–1450.

Cox, R. M., Alexander, G. C., and Gilmore, C. (**1987**). ''Development of the connected speech test (CST),'' Ear Hear. **8**, 119–126.

Cox, T. J., Li, F., and Darlington, P. (**2001**). ''Extracting room reverberation time from speech using artificial neural networks,'' J. Audio Eng. Soc. **49**, 219–230.

Eyring, C. F. (**1930**). ''Reverberation time in 'dead' rooms,'' J. Acoust. Soc. Am. **1**, 217–241.

Greenberg, J. E., and Zurek, P. M. (**2001**). ''Microphone-array hearing aids,'' in *Microphone Arrays: Signal Processing Techniques and Applications*, edited by M. Brandstein and D. Ward (Springer-Verlag, Berlin), pp. 229–253.

Hartmann, W. M. (**1997**). ''Listening in a room and the precedence effect,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, New York), pp. 191–210.

International Phonetic Association (**1999**). *Handbook of the International Phonetic Association* (Cambridge U.P., Cambridge). The American-English sound files are available on-line at ⟨http://uk.cambridge.org/linguistics/resources/ipahandbook/american-English.zip⟩.

ISO 3382 (**1997**). *Acoustics—Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters*, 2nd ed. (International Organization for Standardization, Géneve).

Knudsen, V. O. (**1929**). ''The hearing of speech in auditoriums,'' J. Acoust. Soc. Am. **1**, 56–82.

Kuttruff, H. (**1991**). *Room Acoustics*, 3rd ed. (Elsevier Science Publishers Ltd., Lindin).

Lebart, K., Boucher, J. M., and Denbigh, P. N. (**2001**). ''A new method based on spectral subtraction for speech dereverberation,'' Acustica **87**, 359–366.

Millington, G. (**1932**). ''A modified formula for reverberation,'' J. Acoust. Soc. Am. **4**, 69–82.

Miyoshi, M., and Kaneda, Y. (**1988**). ''Inverse filtering of room impulse response,'' IEEE Trans. Acoust., Speech, Signal Process. **36**, 145–152.

Nabalek, A. K., and Pickett, J. M. (**1974**). ''Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners,'' J. Speech Hear. Res. **17**, 724–739.

Nabalek, A. K., and Robinson, P. K. (**1982**). ''Monaural and binaural speech perception in reverberation for listeners of various ages,'' J. Acoust. Soc. Am. **71**, 1242–1248.

Nabalek, A. K., Letowski, T. R., and Tucker, F. M. (**1989**). ''Reverberant overlap- and self-masking in consonant identification,'' J. Acoust. Soc. Am. **86**, 1259–1265.

Nannariello, J., and Fricke, F. (**1999**). ''The prediction of reverberation time using neural network analysis,'' Appl. Acoust. **58**, 305–325.

Neely, S. T., and Allen, J. B. (**1979**). ''Invertibility of room impulse response,'' J. Acoust. Soc. Am. **66**, 165–169.

Pitas, I., and Venetsanopoulos, A. N. (**1992**). ''Order statistics in digital image processing,'' Proc. IEEE **80**, 1893–1921.

Poor, V. (**1994**). *An Introduction to Signal Detection and Estimation* (Springer-Verlag, New York).

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (**1992**). *Numerical Recipes in C* (Cambridge U.P., Cambridge).

Ratnam, R., Jones, D. L., and O'Brien, Jr., W. D. (**2003**). ''Fast algorithms for the blind estimation of reverberation time,'' IEEE Signal Process Lett. (in press).

Sabine, W. C. (**1922**). *Collected Papers on Acoustics* (Harvard U.P., Cambridge).

Schroeder, M. R. (**1965a**). ''New method for measuring reverberation time,'' J. Acoust. Soc. Am. **37**, 409–412.

Schroeder, M. R. (**1965b**). ''Response to 'Comments on New method of measuring reverberation time,' '' J. Acoust. Soc. Am. **38**, 359–361(L).

Schroeder, M. R. (**1966**). ''Complementarity of sound buildup and decay,'' J. Acoust. Soc. Am. **40**, 549–551.

Sette, W. J. (**1933**). ''A new reverberation time formula,'' J. Acoust. Soc. Am. **4**, 193–210.

Smith, Jr., P. W. (**1965**). ''Comment on 'New method of measuring reverberation time,''' J. Acoust. Soc. Am. **38**, 359(L).

Tahara, Y., and Miyajima, T. (**1998**). ''A new approach to optimum reverberation time characteristics,'' Appl. Acoust. **54**, 113–129.

Xiang, N. (**1995**). ''Evaluation of reverberation times using a nonlinear regression approach,'' J. Acoust. Soc. Am. **98**, 2112–2121.

Young, R. W. (**1959**). ''Sabine reverberation equation and sound power calculations,'' J. Acoust. Soc. Am. **31**, 912–921.

A contribution to simulating a three-dimensional larynx model using the finite element method

Marcelo de Oliveira Rosa[a] and José Carlos Pereira[b]
Depto. Engenharia Elétrica, Escola de Engenharia de São Carlos, Universidade de São Paulo, Av. Trabalhador sancarlense, 400, 13566-590, São Carlos, São Paulo, Brazil

Marcos Grellet
Depto. de Oftalmologia e Otorrinolaringologia, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes, 3900, 14049-900, Ribeirão Preto, São Paulo, Brazil

Abeer Alwan
Speech Processing and Auditory Perception Laboratory, Department of Electrical Engineering, 66-147E Engr. IV, 405 Hilgard Avenue, Box 951594, University of California, Los Angeles, California

(Received 3 January 2003; revised 14 August 2003; accepted 25 August 2003)

A three-dimensional model is presented to simulate the larynx during vocalization. The finite element method is used to calculate the airflow velocity and pressure along the larynx as well as tissue displacement. It is assumed that the larynx tissue is transversely isotropic and divided into three tissues: cover, ligament, and body. A contact-impact algorithm is incorporated to deal with the physics of the collision between both true vocal folds. The results show that the simulated larynx can reproduce the vertical and horizontal phase difference in the tissue movements and that the false vocal folds affect the pressure distribution over the larynx surfaces. The effects of exciting the larynx with different pressure drops are also investigated. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1619981]

PACS numbers: 43.70.Aj, 43.70.Bk, 43.70.Ep [AL] Pages: 2893–2905

I. INTRODUCTION

Larynx modeling is critical for providing insight into the dynamic behavior of human phonation. Using mathematical techniques and experimental data, models have been proposed in order to better understand how the larynx vibrates, what are its vibration conditions, and the influence of the tissue mechanics on the glottal and speech waveform. A better model of the larynx may also help in addressing two important clinical issues: Is it possible to predict a certain voice pathology based on a mathematical model of the larynx? Could a mathematical model simulate, *a priori*, the effects of a surgical intervention in a damaged larynx?

Several models have been proposed to deal with voice synthesis and larynx simulation. Ishizaka and Flanagan (1972), for example, present a very simple and efficient two-mass model where several aspects of normal phonation were investigated. Alipour *et al.* (2000), on the other hand, present a sophisticated larynx model where complex and rigorous mechanical and geometric characteristics of the tissues are applied in order to obtain a more realistic representation of the vocal fold dynamics. The advantage of continuum models is their ability to directly relate the geometric, viscoelastic, and aerodynamic characteristics of the larynx with the synthesized glottal waveform. In this paper, a three-dimensional finite element model of the larynx is proposed. The model builds and expands on that proposed by Alipour *et al.* (2000).

Three-dimensional larynx models are challenging in terms of computational efforts to solve the physical laws that control human phonation. The airflow pressure and velocities are described by Navier–Stokes equations while viscoelastic models deal with the larynx tissue movements. In addition, the collision of the vocal folds has an important impact on modeling because it affects the glottal waveform shape. The proposed model uses well-established algorithms to account for these three phenomena.

Berry *et al.* (1994) analyze the vibration pattern resulting from the simulation of a two-dimensional model of the vocal folds. They show that a nonchaotic pattern is found when the viscoelastic and aerodynamic conditions in the larynx are properly set. Berry and Titze (1996) state that tissue incompressibility is an important factor in the vibration modes. Both studies present sets of viscoelastic parameters of the larynx tissues for normal vibration.

De Vries *et al.* (1999) describes a three-dimensional model similar to Alipour's one. The only difference between both models is that De Vries *et al.* (1999) assume that the larynx tissues consist of body and cover tissues while Alipour *et al.* (2000) consider the body, ligament, and cover in their tissue formulation.

Alipour *et al.* (2000) present a three-dimensional model of a vocal fold (assuming larynx symmetry) where the longitudinal movements of the tissue model are restricted by assuming that these movements are negligible (the shear stress produced by the movements of successive transversal planes was incorporated in their formulation by algebraic manipulations). Additionally, the collision between the vocal folds are treated in their larynx model by bounding the tissue

[a]This work is part of the first author's Ph.D. thesis.
[b]Author to whom correspondence should be addressed. Electronic mail: pereira@sel.eesc.sc.usp.br

displacements: when their surface nodes cross the symmetry plane, they simply zero the nodal displacements in order to avoid the vocal folds penetrating into each other.

In terms of aerodynamic simulation, Guo and Scherer (1993) use a finite element model to determine the air velocity and pressure in a two-dimensional representation of the larynx channel (considering only the true vocal folds). They demonstrate that different laryngeal profiles significantly change the aerodynamic forces applied over the vocal fold surfaces.

The present model considers the whole larynx (including false vocal folds and laryngeal ventricles) in three dimensions and not just the true vocal folds. Hence, important vibration properties like the horizontal phase-difference can be studied. Following Alipour *et al.* (2000), the tissue model was linearly transversal isotropic because of the strain–stress differences in the longitudinal direction and transversal planes. However, the model assumes that the larynx moves longitudinally.

The aerodynamic influence of the false vocal folds is analyzed in order to demonstrate their influence in whispered voices as mentioned by Miller *et al.* (1988). In addition, the nonlinear pressure distribution calculated along the larynx surfaces allows us to confirm the myoelastic-aerodynamic theory. By solving the Navier–Stokes equation in three dimensions, the present model extends the Guo and Scherer one.

Finally, a physically based method to deal with the collision phenomenon is presented. It calculates the collision forces that avoid the larynx tissue interpenetration during phonation (Newton's second law). It is different from the two-mass model approach where the strain–stress relationship of the larynx tissue is changed during the collision, and from the Alipour *et al.* (2000) technique of direct manipulation of the surface tissue displacements.

The fundamental mathematical technique used in this work is the finite-element method which is a procedure of discretizing partial differential equations. Consequently, all the physical systems (fluid, tissue, and collision problem) involved in phonation are addressed by this technique. Descriptions of the finite-element algorithms are well-documented elsewhere (Bathe, 1996).

II. MATERIALS AND METHODS

The simulation of the larynx in phonatory conditions requires the solution of four different engineering problems: fluid flow, tissue displacement, mesh deforming, and contact problem. Because well-established algorithms are used to solve these problems, only ''how-to-use'' procedures will be addressed.

A. Mesh generator and deforming

Both airflow and tissue meshes are constructed from geometric specifications of the larynx. It is considered that the larynx is sliced into several cross sections along the neck. Therefore several small ''hyper-ellipses'' are specified to define both the internal and external larynx tissue surfaces—the internal tissue surface is the external flow channel. Figure 1 shows a geometric description of a larynx from its hyper-ellipses where the larynx surfaces are shown.

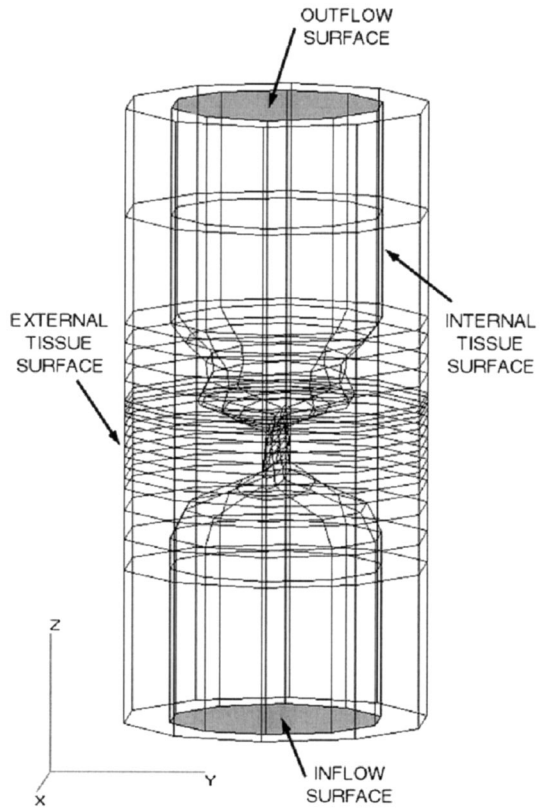


FIG. 1. Geometric description of the larynx surfaces from its hyper-elliptical sections.

The equations used to define all hyper-ellipses are

$$\begin{aligned} &x = R_1 \cos(\theta), \quad y = R_1 \sin(\theta), \quad 0 \leqslant \theta \leqslant \pi/2, \\ &x = R_2 \cos(\theta), \quad y = R_2 \sin(\theta), \quad \pi/2 \leqslant \theta \leqslant \pi, \\ &x = R_3 \cos(\theta), \quad y = R_3 \sin(\theta), \quad \pi \leqslant \theta \leqslant 3\pi/2, \\ &x = R_4 \cos(\theta), \quad y = R_4 \sin(\theta), \quad 3\pi/2 \leqslant \theta \leqslant 2\pi. \end{aligned} \tag{1}$$

To complete the tissue specification, the thickness of the larynx cover and ligament are defined—the remaining space between both internal and external tissue surfaces is considered as being part of the body layer. Later the geometric description of the laryngeal model will be presented.

Once these geometric characteristics are defined, an automatic mesh algorithm creates the tetrahedrons (polygonal elements used by FEM). At the airflow–muscle interface, both the flow and tissue elements are EXACTLY connected to each other by their triangular surfaces. It means that the airflow and tissues ''share'' the same nodes at their interfaces. In all simulations, the number of nodes for airflow and tissue meshes are 4151 and 5140, respectively. The numbers of tetrahedrons for airflow and tissue meshes are 2600 and 3000, respectively. The node and tetrahedron distribution in both meshes is determined by the spatial position of the hyper-ellipses: denser nodal distributions are obtained where successive hyper-ellipses are closer. The larynx movements demand that all of the larynx mesh moves. The position of all tissue nodes are naturally updated using the computed dis-

placements of the tissue FEM model. However, only the displacements of the external flow nodes are known because of their connection to nodes of the tissue surfaces.

The mesh moving technique of Johnson and Tezduyar (1994) is used to update the internal flow nodes. They considered that the flow mesh is a rigid material subject to specific displacements of all external nodes. Therefore the same FEM procedure used to model the larynx tissues is employed here. Specifically, tetrahedrons with linear polynomials for interpolation and shape functions are used. Because the ''material'' is isotropic, Johnson and Tezduyar (1994) defined the factor λ/μ_L (where λ and μ_L are Lamé elastic constants) as the only ''physical'' parameter for the mesh moving method. They proposed $\lambda/\mu_L=2$ in order to control the deformation that is used in the mesh moving algorithm.

The boundary conditions of this algorithm are Dirichlet conditions and no body force is applied. Therefore only flow nodes in contact with the tissue ones have the same displacement. The displacement of inflow and outflow nodes are set to null.

This strategy provides a fast and reliable deformation of the flow mesh in several situations. At each simulation time step, the displacements of the external tissue nodes are collected and applied as boundary conditions to the mesh moving algorithm. It adequately places all internal flow nodes inside the larynx channel.

However, there are situations where this algorithm could not deal with the boundary conditions requirements. We noticed this phenomenon by examining the resulting mesh: some tetrahedrons had null or negative volumes. When this happens, the mesh generator is triggered to reconstruct all flow tetrahedrons from the deformed internal tissue surfaces.

B. Solving the tissue displacement

The concepts of continuum mechanics presented by Alipour *et al.* (2000) to describe tissue displacement during phonation are adopted. Thus, the simulated larynxes are divided into finite geometric elements and the displacement vectors and pressures are distributed throughout all the elements that compose the meshes.

It is assumed that the larynx tissues have transversal isotropic properties along the vocal fold length (longitudinal direction) attempting to match the fibrous directions of the laryngeal tissues. Mathematically the larynx tissues follow the linear elasticity theory, demanding therefore small time integration intervals as indicated by Alipour *et al.* (2000). The complete derivation of the element matrices involved in FEM will not be presented here because it can be found in Alipour *et al.* (2000), Huebner and Thornton (1982), and Bathe (1996).

However, there are some differences between our studies and Alipour *et al.* (2000) modeling. First, their model did not account for the longitudinal movements of the vocal folds in order to significantly reduce the number of equations whereas we analyze the magnitude of the movements in that direction. It is interesting to note that Alipour *et al.* (2000) coupled all the transversal planes through the use of shear forces. Here such forces are implicitly considered in the conventional and truly three-dimensional finite element modeling of the larynx tissues as in Gunter (2003).

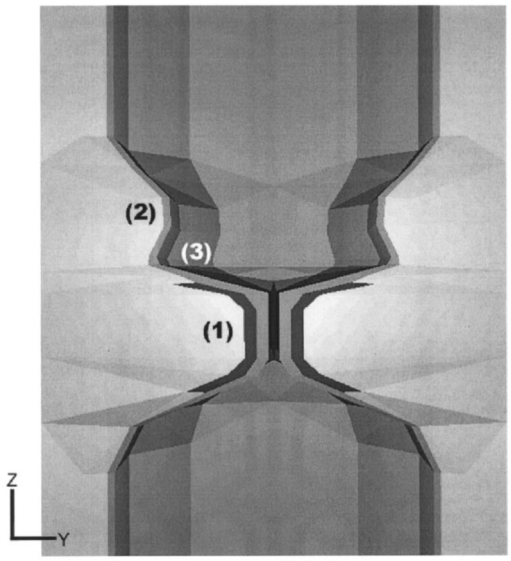


FIG. 2. Transversal view of a simulated normal larynx. The different shades refer to the three different tissues in the laryngeal tissue. (1) True vocal folds. (2) False vocal folds. (3) Laryngeal ventricles.

Second, their model—including Gunter (2003)—considered only the true vocal folds and it assumed symmetry of the right and left vocal folds. This allowed for the suppression of one of the folds in order to also reduce the number of equations to be solved. Figure 2 shows that the simulated larynx in this paper included both true and false vocal folds.

In terms of FEM, tetrahedrons are used to describe the tissue mesh. Different from Alipour *et al.* (2000), quadratic polynomials are employed as interpolation functions for displacement vectors and as shape functions for describing the geometry of the tetrahedrons. It is important to mention that straight-edged tetrahedrons are used to simplify the integral functions required to compute the element matrices in FEM. Although more accurate solutions are obtained with higher-order polynomials, the decision of using quadratic instead of linear elements was made to simplify numerical algorithms for both tissue and airflow modeling. Further explanations in the following subsection will justify the use of such interpolation and shape functions.

The boundary conditions that excite the tissue movements are air pressure at the internal surface of the larynx model (contact interface between air and tissue) and null displacement at all other external surfaces of the model. It means that the larynx is ''stuck'' at the neck during the entire phonation process. Specifically, the air pressures at the tissue nodes in contact with the airflow are linearly distributed along the external triangular surfaces. Air shear forces at the larynx walls are nulls because of the boundary conditions applied to solve the airflow model. Titze and Talkin (1979) pointed that these shear forces are very small when compared to the pressure forces, which justifies our assumption.

While Alipour *et al.* (2000) use a central difference approximation method to time-integrate the tissue equations, the Newmark method (Bathe, 1996) is used because of its

''unconditional'' stability. The time step (Δt) for all simulations is 300 μs. Its mathematical structure is implicit and produces a global linear system of equations to be solved. Therefore a direct sparse solver based on a supernodal approach (Demmel *et al.*, 1999) is employed.

The laryngeal tissue is divided into three kinds of tissues (cover, ligament, and body) following Alipour *et al.* (2000) and different from De Vries *et al.* (1999). These three kind of tissues differ from each other in terms of elastic constant, damping ratios, and geometric characteristics.

C. Solving the velocity-pressure flow

A laminar, incompressible, isothermal, steady Navier–Stokes equation (NS), as shown in Eq. (2), describes the airflow modeled by FEM. Although the larynx airflow is time dependent, the approach provided by Guo and Scherer (1993) is used in order to directly compare the results. In addition, it simplifies the mathematical description of the problems since the larynx movements should be incorporated into time-dependent airflow equations.

$$\nabla \cdot \mathbf{U} = 0,$$
$$\rho_f(\mathbf{U}\cdot\nabla)\mathbf{U} = \mu_f(\nabla^2\mathbf{U}) - \nabla\mathbf{P}, \tag{2}$$

where $\mathbf{U}$ is the velocity vector, $\mathbf{P}$ is the scalar pressure, and ρ_f and μ_f are the density and viscosity air, respectively.

The airflow mesh is divided into tetrahedrons (as in the tissue model) with quadratic and linear polynomials to respectively interpolate the velocity and pressure along the element (known as mixed formulation). This procedure is necessary to obtain a well-posed problem (Gresho and Sani, 1998). While Guo and Scherer (1993) use a penalty formulation, a mixed formulation of NS (using Lagrange multipliers) is adopted here.

To calculate the velocity and pressure along the larynx for several profiles, several strategies are applied in order to deal with the convergence difficulties for high-Reynolds number flows. These strategies are described in the next paragraphs.

One of the sources of difficulty is the convection term of NS. High-speed flows tend to produce ''wiggles'' that can destabilize the solution. One way to control this problem is to make the mesh finer. However, the drawback is the significant increase in the number of equations to be solved, specially in three dimensions.

A hybrid approach (FIDAP, 1999) is followed to obtain a robust FEM formulation of NS in high-Reynolds laminar flows by applying artificial numerical diffusion in the solution. It means that low- and high-order schemes are linearly combined to stabilize the convection term. Low-order schemes are first-order accurate and unconditionally stable but they introduce excessive amounts of numerical diffusion. On the other hand, high-order schemes (in our case, second-order accurate) significantly reduce numerical diffusion but are inherently unstable.

The first- and second-order schemes (the last is also called Streamline-Upwind/Petrov-Garlekin or SUPG) are respectively described by Tabata and Fujima (1991) and Sampaio (1991). Tabata and Fujima (1991) use triangles (with quadratic polynomials) while this work employs tetrahedrons (the adaptation between both geometric elements is direct). Sampaio (1991) also uses triangles but with linear polynomials for velocity. Deblois (1996) provides additional information to construct the element matrices associated to SUPG for triangles with quadratic polynomials for velocity (easily extended to tetrahedrons).

The blending procedure to mix both first- and second-order schemes follows FIDAP (1999) where the velocity gradient is locally evaluated by using a simple Laplacian filter applied over the faces of the analyzed tetrahedron. It produces a numerical estimation of the flow changes along the mesh and allows for quantifying the amount of both schemes (first- and second-order, respectively) used to construct all FEM element matrices. This hybrid approach is equivalent to finite-difference methods like MUSCL and TVD (Hirsh, 1988) in combining robustness (first-order schemes) and accuracy (high-order schemes).

To complete the problem specification, the boundary conditions are set to null velocity for external flow nodes in contact with tissue nodes. It means that the air velocities through the larynx wall are zero. Although these wall velocities should be not zero because of the tissue movements, they are considered zero in order to simplify the algorithms, specially when there is a collision between both true vocal folds. It is important to mention that tissue movements are significantly slower (approximately 100 times in the glottal channel) than the average airflow velocity, making the algorithm-based decision suitable for all simulations.

Pressures are set constant for the inflow and outflow larynx surfaces (respectively ''lungs'' and ''supraglottal'' pressures)—null velocity gradients are assumed at both surfaces. Therefore an aerodynamic condition is created to excite flow through the larynx using specific pressure drops between the inflow and outflow surfaces. Mathematically the ''do-nothing'' concept of Sani and Gresho (1994) is followed to keep the problem well-posed.

After the definition of the FEM approach to discretize NS for all finite elements (producing local nonlinear equation systems), a global nonlinear equation system to be solved is constructed. As Guo and Scherer (1993) the successive substitution method is used, i.e.,

$$K(\mathbf{U_{n-1}})\mathbf{U}^* = \mathbf{R},$$
$$\mathbf{U_n} = 0.5\mathbf{U_{n-1}} + 0.5\mathbf{U}^*, \tag{3}$$

where $\mathbf{U}^*$, $\mathbf{U_{n-1}}$, and $\mathbf{U_n}$ are global solution vectors (velocity+pressure) for temporary, $(n-1)$th and nth iteration steps. The solution is achieved when Eq. (4) is reached. A direct sparse solver (Demmel *et al.*, 1999) is used to solve the linear system in Eq. (3):

$$\frac{|\mathbf{U_n} - \mathbf{U_{n-1}}|}{|\mathbf{U_n}|} < 0.001. \tag{4}$$

The reduced basis method (Peterson, 1989) is used to speed up and control the convergence. The simulation is pressure driven and therefore large pressure drops (around 8 kdyne/cm^2) excite the flow in the larynx. Instead of fully applying this amount of pressure difference, a set of N sub-

problems at low pressure drops using Eq. (3) is solved. Once all these subsolutions (also called basis) have been determined, an approximate solution is finally calculated in order to form the first guess of the final iterative process to reach the correct solution, using again Eqs. (3) and (4). The following steps describe the reduced basis method:

(1) Estimate first base **U0th_base**
(2) Calculate additional basis **U1st_base...U4th_base**
(3) Calculate the final approximated solution **Ureduced_base**
(4) Converge up to the solution from **Usolved.**

$N=5$ was used to speed up convergence by quickly getting a good first guess (**Ureduced_base**). The choice for this value is a tradeoff between speed and computational efforts.

The final aspect in our method is the choice of the pressure drops for each subproblem. A simple approach is used: Starting from a specific pressure drop $\Delta P4$, $\Delta P0=\Delta P4/5$, $\Delta P1=2*\Delta P4/5$, $\Delta P2=3*\Delta P4/5$ and $\Delta P3=4*\Delta P4/5$.

Although we tried to mathematically define an association between $\Delta P4$ and parameters of the flow (Reynolds number, residue, $|\mathbf{U_n}-\mathbf{U_{n-1}}|$, etc.), it was finally decided to pick the parameter based on the convergence rate: low and high pressure drops result in fast and low convergence, respectively (sometimes, high flows even oscillate without reaching a stationary solution). However, when high pressure drops are employed, the approximate vector **Ureduced_base** is very close to the final solution and the number of iterations in step 4 of the reduced basis method is very low. It means that we have to choose $\Delta P4$ thinking of the tradeoff between convergence and closeness to the solution in Eq. (4). In the case of the simulated larynx presented in this work, $\Delta P4=5$ Pa.

D. Coupling air and tissue systems

The links between both physical systems (air and tissue) are the airflow pressure and the moving mesh. As mentioned before, surface nodes of the airflow mesh are directly attached to surface nodes of the tissue nodes. It means that the air pressures are collected at the larynx walls and transferred to the tissue system in order to calculate the nodal displacement. Note that this direct node matching overcome problems of pressure estimation along the larynx walls because the pressures are exactly calculated over the tissue nodes.

The mesh updating also couples both systems because the change of airflow mesh generates a new velocity and pressure profiles along the laryngeal cavity. Again, moving together the airflow and tissue nodes located at the physical interface allows an easy information transfer.

E. Contact problem

The contact problem refers to the physical phenomena occurred when two or more bodies touch each other. In phonation, both true vocal folds collide according to the myoelastic-aerodynamic theory. Ishizaka and Flanagan (1972) changed the stress-displacement curve during the collision instant in order to avoid the penetration among the bodies. They also assumed that such phenomena always took place at the middle larynx section due to the laryngeal symmetry.

Alipour *et al.* (2000) considered a different approach: when their true vocal fold reached the middle section, they simply removed one degree of freedom of the collided node. Different from Ishizaka and Flanagan (1972), they did not apply the required force to avoid the body penetration. Apparently Gunter (2003) used a similar procedure to the method that will be described in the next paragraphs.

This simulation uses an algorithm (Bathe, 1996) to deal with the contact problem: all collided nodes (only the ones located at the tissue surface) are collected in order to calculate the required forces to avoid body interpenetration. The fundamental aspect is that such forces are calculated considering how deep one vocal fold can enter into each other.

Mathematically the following restrictions are kept for all surface tissue nodes (using contact problem notation):

$$G=(\mathbf{X}-\mathbf{Y})^T*\mathbf{N}\geqq 0$$
$$L=\mathbf{F}*\mathbf{N}\geqq 0 \qquad (5)$$
$$G*L=0.$$

It means that the gap (G) between two nodes ($\mathbf{X}$ and $\mathbf{Y}$ are the spatial position of master and slave nodes at different bodies) should be non-negative and that the collision force $\mathbf{F}$ should always take the bodies apart. The last restriction in Eq. (5) indicates that both the inequalities are exclusive: when one condition is reached, the other one is automatically equal to zero. Further theoretical details are found in Bathe (1996).

To add Eq. (5) into the linear system of equations produced by FEM to the tissue, the mathematical procedure described by Narayanaswamy (1985) is followed where the coupling of both FEM equations and collision restrictions are not necessary. In this procedure the restriction forces associated with the collided nodes are computed and then the tissue displacements under collision restrictions are calculated.

Therefore the algorithm to compute the tissue displacement in a specific time step is as follows:

1. Calculate the displacements without collision restrictions
2. While there are some body penetration:
2.1. Verify possible nodal collisions
2.2. Calculate the force collisions and nodal displacements according to Narayanaswamy (1985) under Eq. (5).
2.3. End while

The procedure is iterative because it is not known, *a priori*, which nodes will collide with some larynx surface, resulting in a nonlinear behavior to the contact problem. It can be realized that the contact problem is part of the tissue model. Therefore, the time step is only increased after the calculation of the nodal tissue displacements under collision restrictions.

As shown, the larynx symmetry is not used to simplify or reduce computational efforts. It means that the concept of master-slave bodies is not directly employed because there is

TABLE I. Geometric description of a simulated normal larynx.

Section	Center (cm)	Radius R1, R2, R3, R4 (cm)
S0	0, 0, −2	0.7, 0.7, 0.7, 0.7
S1	0, 0, −1	0.7, 0.7, 0.7, 0.7
S2	0, 0, −0.8	0.7, 0.7, 0.7, 0.7
S3	0, 0, −0.61	0.7, 0.6, 0.7, 0.6
S4	0, 0, −0.47	0.7, 0.25, 0.7, 0.25
S5	0, 0, −0.40	0.7, 0.075, 0.7, 0.075
S6	0, 0, −0.33	0.7, 0.025, 0.7, 0.025
S7	0, 0, −0.26	0.7, 0.025, 0.7, 0.025
S8	0, 0, −0.19	0.7, 0.025, 0.7, 0.025
S9	0, 0, −0.12	0.7, 0.025, 0.7, 0.025
S10	0, 0, −0.05	0.7, 0.025, 0.7, 0.025
S11	0, 0, 0	0.7, 0.235, 0.7, 0.235
S12	0, 0, 0.03	0.7, 0.35, 0.7, 0.35
S13	0, 0, 0.07	0.7, 0.5, 0.7, 0.5
S14	0, 0, 0.10	0.7, 0.515, 0.7, 0.515
S15	0, 0, 0.23	0.7, 0.45, 0.7, 0.45
S16	0, 0, 0.35	0.7, 0.46, 0.7, 0.46
S17	0, 0, 0.48	0.7, 0.6, 0.7, 0.6
S18	0, 0, 0.60	0.7, 0.7, 0.7, 0.7
S19	0, 0, 1.30	0.7, 0.7, 0.7, 0.7
S20	0, 0, 2	0.7, 0.7, 0.7, 0.7

Layer	Thickness (mm)	Sections
Cover	0.5	S5–S12
Ligament	0.5	S5–S12

no preferential body. All the surface nodes are tested against the tissue body to see if there is a collision and then the collided surface, which provides the normal **N** and position **Y** vectors in Eq. (5), is backtracked. This method is called two-way collision detection (ANSYS, 2000).

F. Final iterative procedure

After separately describing the mathematical algorithms for both physical systems, they are put together in an iterative loop in order to simulate the larynx as follow:

(1) Collect air pressures.
(2) Calculate tissue displacements verifying vocal fold collisions.
(3) Update tissue mesh.
(4) Update airflow mesh.
(5) Calculate airflow velocities and pressures.

Note that in the first iterative step, the collected air pressures are null.

G. Description of larynx parameters

Here the parameters used to simulate the larynx will be presented. The geometric description of the simulated larynx is shown in Table I. As mentioned before, the larynx is ''sliced'' into sections using ellipses for geometric definition.

The airflow parameters for simulation are $\mu_f\,(\text{viscosity}) = 1.79e\text{-}5$ Pa.s and $\rho_f\,(\text{density}) = 1.23\,\text{kg/m}^3$. The inflow and outflow air pressure are respectively set to 0.8 and 0 (zero) kPa. As in Alipour *et al.* (2000), the tissue layers have transversal isotropic elastic behavior. Therefore six viscoelastic parameters are needed: transverse and longitudinal Poisson's ratios (ν and ν'), transversal and longitudinal Young's modulus (E and E') and longitudinal shear modulus (μ'_m) and tissue viscosity η. Transversal shear modulus is not necessary because it is directly related to E and ν by

$$\mu_m = E/(2*(1+\nu)). \tag{6}$$

Alipour *et al.* (2000) defined $\nu' = 0$ because they eliminated any longitudinal movement. Here non-null longitudinal Poisson's ratios are allowed because all tissue nodes are free to move (except when they are under collision or boundary condition restrictions). This additional degree of freedom in the tissue specifications requires care in order to control the almost incompressible behavior of the laryngeal tissues. In order to simplify the model, all analyses are conducted by assuming $\nu = \nu'$. The final value is then chosen in order to reach almost incompressible tissues. Titze and Strong (1975) and Berry and Titze (1996) provide more details about this mechanical parameter.

TABLE II. Description of the viscoelastic tissue parameters of a normal larynx.

Tissue	E (kPa)	ν	E' (kPa)	μ'_m (kPa)	ν'	η (Pa.s)
Cover	2	0.76	20	20	0.76	1
Ligament	4	0.68	20	40	0.68	1
Body	20	0.45	20	30	0.45	15

Table II shows the values of tissue parameters for a specific simulation. Different from Alipour *et al.* (2000) the tissue viscosity of the body is set high according to observations of Titze and Strong (1975) where the authors discussed the similarities in the organic composition of the larynx tissues and other tissue tissues in the human body. Mathematically it establishes a critically damped system for the body tissues to quickly reach their stationary position. Lowering such a parameter brings the tissue to an under-damped condition and therefore an initial transitory oscillation is observed.

The transversal Young's modulus (E) is defined according to Tran *et al.* (1993) who measured it *in vivo*. Because their estimation resulted from the whole vocal fold reaction to an applied force, a fixed value is just extended for all the tissue layers. The other parameters are described in Berry and Titze (1994), Titze and Strong (1975), and Titze and Talkin (1979).

III. RESULTS AND DISCUSSIONS

Two analyses are done in order to demonstrate the effectiveness of the proposed model. First, the larynx geometries are fixed in three different profiles and the pressure distribution are calculated and compared with other models. Second, dynamic analyses are conducted to verify the effects of changing viscoelastic parameters and subglottal pressure on the synthesized glottal signal.

A. Airflow in fixed laryngeal geometries

During phonation, the human larynx roughly assumes three different shapes: divergent, convergent, and parallel. In each condition, there are variations according to the attack

TABLE III. Geometric description of larynxes in divergent, convergent, and parallel profiles. Only the differences from Table I are presented.

Profile	Section	Center (cm)	Radius R1, R2, R3, R4 (cm)
Divergent	S6	0, 0, −0.33	0.7 0.025 0.7 0.025
	S7	0, 0, −0.26	0.7 0.035 0.7 0.035
	S8	0, 0, −0.19	0.7 0.045 0.7 0.045
	S9	0, 0, −0.12	0.7 0.055 0.7 0.055
	S10	0, 0, −0.05	0.7 0.065 0.7 0.065
Convergent	S6	0, 0, −0.33	0.7 0.065 0.7 0.065
	S7	0, 0, −0.26	0.7 0.055 0.7 0.055
	S8	0, 0, −0.19	0.7 0.045 0.7 0.045
	S9	0, 0, −0.12	0.7 0.035 0.7 0.035
	S10	0, 0, −0.05	0.7 0.025 0.7 0.025
Parallel	S6	0, 0, −0.33	0.7 0.025 0.7 0.025
	S7	0, 0, −0.26	0.7 0.025 0.7 0.025
	S8	0, 0, −0.19	0.7 0.025 0.7 0.025
	S9	0, 0, −0.12	0.7 0.025 0.7 0.025
	S10	0, 0, −0.05	0.7 0.025 0.7 0.025

and entrance angles and the channel length. The glottal geometry shown in Table I was changed in order to produce such profiles. Table III shows only these modified sections and Fig. 3 presents theses geometric profiles in transversal view.

The subglottal pressure is 0.5884 kPa (or 6 cmH_2O) which is close to that provided by Guo and Scherer (1993). In their studies, a parabolic velocity profile was set in order to excite the airflow through their simulated larynx. Also, their profiles only considered one vocal fold, while in this work, the whole larynx (including the laryngeal ventricle and both false vocal folds) are geometrically represented.

Figure 4 shows the air pressure distribution along the whole larynx. In Guo and Scherer's two-dimensional larynxes, the pressures were directly extracted from the nodes in contact with the larynx walls. Because the proposed model uses a three-dimensional larynx, it is first divided into two halves, then the pressures are collected along the intersection of a coronal plane with one-half of the larynx.

These results are quite similar to the ones produced by Guo and Scherer (1993). The differences between the two approaches are related to the fact that the here simulated larynx is coarser than the actual one used by Guo and Scherer (1993). Increasing the mesh nodes would surely smooth the pressure distribution.

The maximum pressure gradient occurs just after the larynx section with minimal constriction. It is developed in the glottal entrance for divergent and parallel profiles and in the glottal output for convergent profiles.

It is possible to observe a secondary negative pressure region formed over the glottal walls becoming more evident although it is less intense (in magnitude) than the main negative pressure region (the same phenomenon is observed when the number of nodes in the airflow mesh is doubled—the result is just smoother then that obtained with the original mesh). When the true vocal fold diameters are increased by 0.02 cm, the pressure distribution becomes smoother.

This phenomenon occurs because tiny glottises cause large vortexes after large aerodynamic jumps (represented by the larynx) due to the demands of fluid incompressibility and convection. Consequently, the transglottal pressure drop is increased when the glottal diameter is reduced.

It was also observed that the longitudinal distribution of air pressure is not constant. The central portion of the vocal folds (approximately 50% of the glottis) is more affected by the air pressure gradient. This fact explains the horizontal phase difference in the vocal fold movements. Because the pressure in the anterior and posterior comissures is very low, it seems that the movements of these glottal regions are not

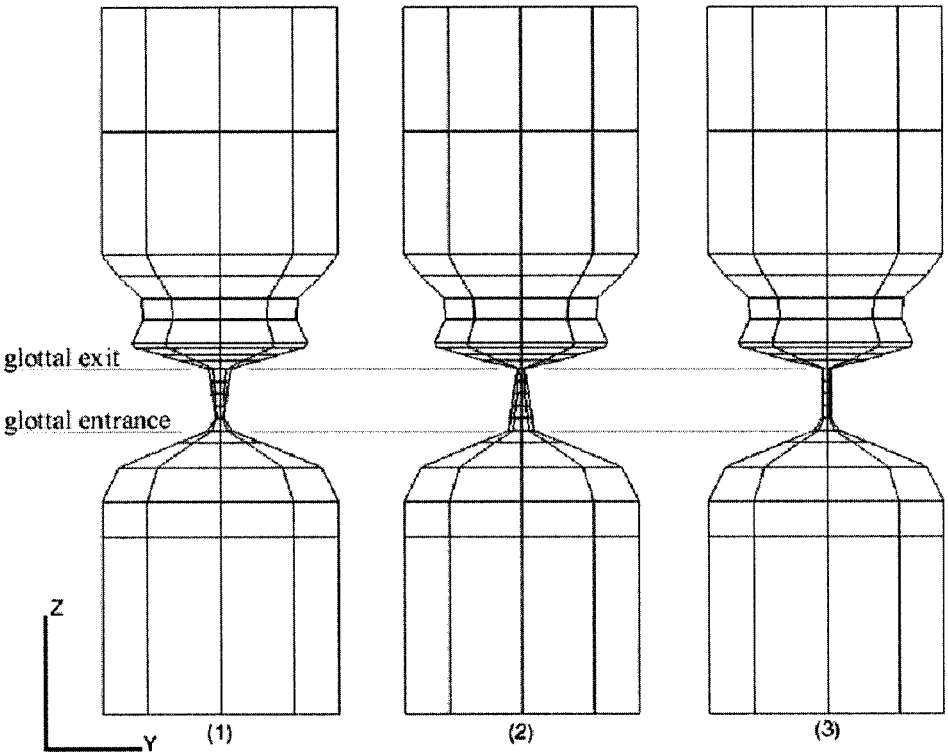


FIG. 3. Geometric description of larynxes with (1) divergent, (2) convergent, and (3) parallel profiles in transversal view.

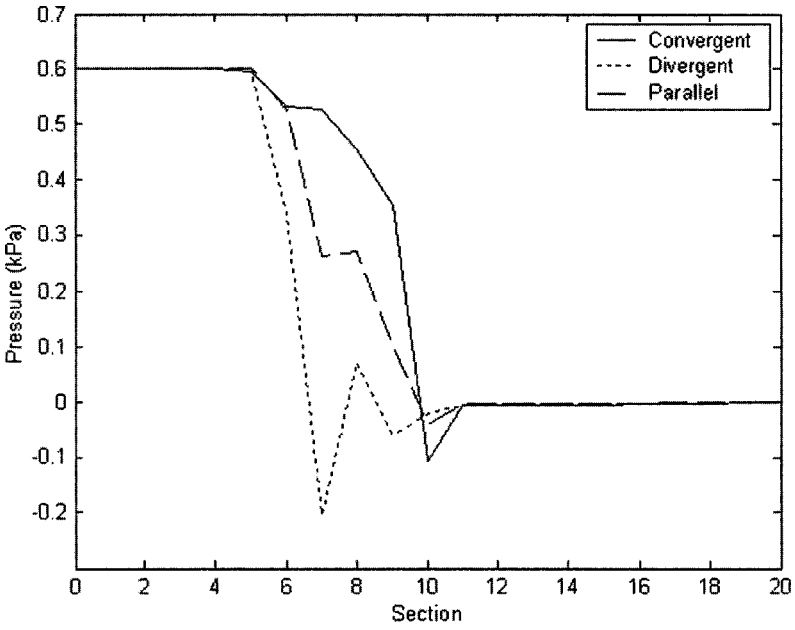


FIG. 4. Pressure distribution over the surface of larynxes with divergent, convergent, and parallel profiles.

excited by aerodynamic conditions (suction, for example) but by influence of the movements of the central portion of the true vocal folds.

The false glottal folds also change pressure distribution. Reducing their diameters increases the air pressure gradient along the larynx. Specifically negative pressures become ''more negative.'' Although this phenomenon happens for all three geometric profiles, the relationship between false vocal fold area and pressure gradient change is nonlinear. To clarify these findings, Fig. 5 shows the differences between three larynx configurations with divergent profiles where the geometry of the false vocal folds have changed. The geometry of all these larynxes is presented in Table IV.

The reduction of the laryngeal areas formed by the false vocal folds increases the transglottal pressure drop. Similar results are obtained for parallel and convergent laryngeal profiles. This phenomenon helps to understand how whispered voices are sustained. A simultaneous reducing of subglottal pressure and false vocal fold area allows the larynx to keep the true vocal folds vibrating, as demonstrated by Miller *et al.* (1988). It means that the minimum subglottal pressure condition to excite the self-vibration of the larynx during the phonation should consider the false vocal fold area at extremely low pressure drops.

B. Complete simulation

All simulations start with a larynx whose geometry is shown in Table I. The viscoelastic properties of the laryngeal tissues are presented in Table II. The lung pressure was 0.8 kPa. Different from Alipour *et al.* (2000), whose simulation was just restricted to one true vocal fold, the present model considered the whole larynx.

Figure 6 presents the synthesized glottal signal whose values are obtained by integrating the volumetric velocity along the exit laryngeal surface. Note the initial transient and

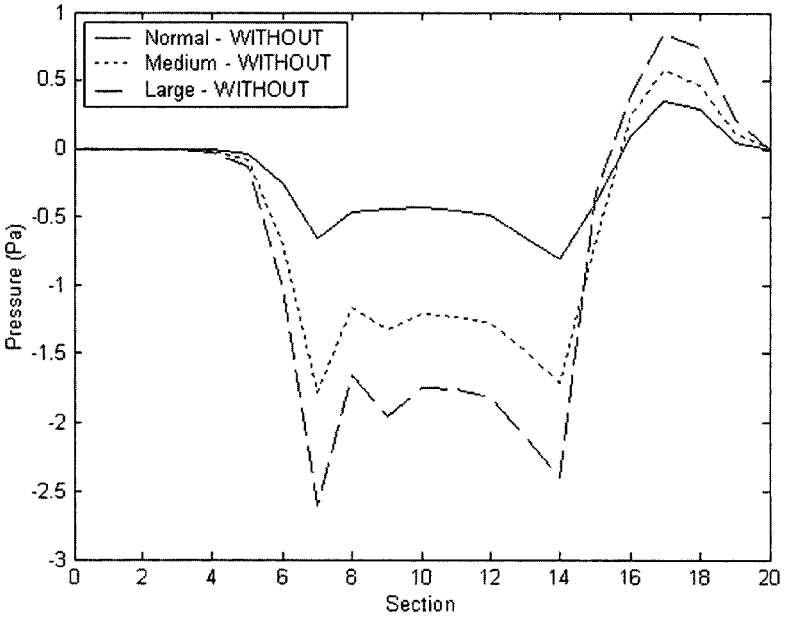


FIG. 5. Curves representing the difference between the pressure distribution of larynxes with three different false vocal fold shapes (normal, medium, and large) and the pressure distribution of a larynx without such a structure. All these larynxes have a divergent profile.

TABLE IV. Geometric description of larynxes with four different false vocal fold shapes which have divergent profiles.

Type of false vocal folds	Section	Center (cm)	Radius R1, R2, R3, R4 (cm)
"None"	S14	0, 0, 0.10	0.7 0.6 0.7 0.6
	S15	0, 0, 0.23	0.7 0.7 0.7 0.7
	S16	0, 0, 0.35	0.7 0.7 0.7 0.7
	S17	0, 0, 0.48	0.7 0.6 0.7 0.6
"Normal"	S15	0, 0, 0.23	0.7 0.45 0.7 0.45
	S16	0, 0, 0.35	0.7 0.46 0.7 0.46
"Medium"	S15	0, 0, 0.23	0.7 0.35 0.7 0.35
	S16	0, 0, 0.35	0.7 0.36 0.7 0.36
"Large"	S15	0, 0, 0.23	0.7 0.25 0.7 0.25
	S16	0, 0, 0.35	0.7 0.26 0.7 0.26

rising of volume velocity until becomes a stationary.

The calculated fundamental frequency is 164 Hz, which is higher than the one presented by Alipour *et al.* (2000). The open quotient for this simulation is 0.6347 while Alipour *et al.* (2000) reported 0.9000. The temporal variables are calculated for the last nine glottal signal cycles when the signal became steady. The difference between the quantities obtained in both simulations can be attributed to the viscoelastic and geometric property differences. It should be taken into account that the three-dimensional aspect of the proposed model affects the pressure distribution along the vocal folds, as demonstrated before, and causes a horizontal phase difference in the tissue movements. It seems that the presented larynx has more mass and therefore more inertial energy than Alipour's one, even considering the differences across the tissue viscosity properties (especially in body tissue specification). This additional inertial energy surely contributed to lower the open quotient value when compared to that obtained by Alipour *et al.* (2000).

Lung pressure affects the temporal properties of the glottal signal. Figure 7 shows the volumetric velocity curves for three larynxes excited by different lung pressures (0.7, 0.8, and 0.9 kPa). The open quotients obtained from these glottal signals are 0.6747, 0.6347, and 0.6249. The lung pressure increase causes higher extensions in the cover tissue movements and, consequently, higher kinetic energy. It means that the glottal closing interval becomes higher because the time of kinetic energy dissipation by the collision effects is increased. Therefore the open quotient is decreased.

The fundamental frequency of each glottal signal is 167.4, 164.0, and 161.3 Hz, showing a variation of −21 Hz/kPa from 0.7 to 0.8 kPa and −34 Hz/kPa from 0.8 to 0.9 kPa. The absolute values are quite similar to the 20.3943 to 30.5915 Hz/kPa (or 2 to 3 Hz/cmH_2O) ratio presented by Ishizaka and Flanagan (1972). However, this ratio is inversely proportional to lung pressure: an increase in lung pressure produces a decrease in the fundamental frequency.

However, the aerodynamic effects in fundamental frequency changes when the larynx become stiffer. Using a larynx with same geometry and higher transversal Young's constants for cover and ligament tissues, as shown in Table V, an increase of 1.5 Hz in the fundamental frequency when the lung pressure is increased from 0.8 to 0.9 kPa is produced. Figure 8 shows that this stiffer larynx excited by 0.8 and 0.9 kPa produces glottal signals with 188.8 and 190.3 Hz, respectively.

Such a dependence of the lung pressure versus fundamental frequency ratio to the laryngeal tissue stiffness (or viscoelastic properties, in a more general view) presents interesting questions on the interrelationship between both myoelastic and aerodynamic aspects of the vocal fold vibration. The results demonstrate that this ratio is nonlinear and dependent on tissue mechanics, although this ratio is quite similar to the ones presented by Ishizaka and Flanagan (1972) in their two-mass model.

Another important aspect is the larynx spatial movements. Alipour *et al.* (2000) presented the coordinates of a tissue point during a phonation, showing that the larynx movements are also periodic. Figure 9 shows the studied spatial points whose coordinates were extracted during a

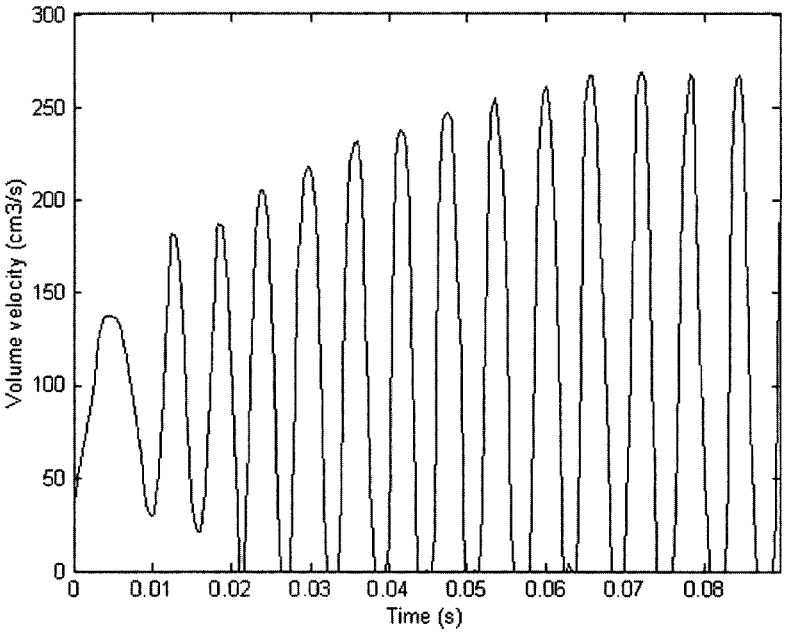


FIG. 6. Glottal signal obtained from a normal larynx excited by pressure drop of 0.8 kPa.

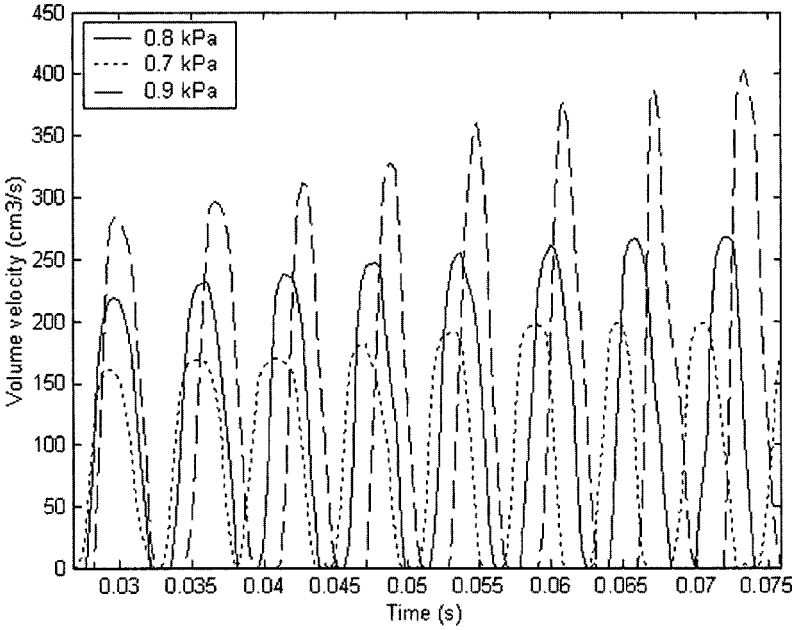


FIG. 7. Glottal signals obtained from normal larynxes excited by different pressure drops.

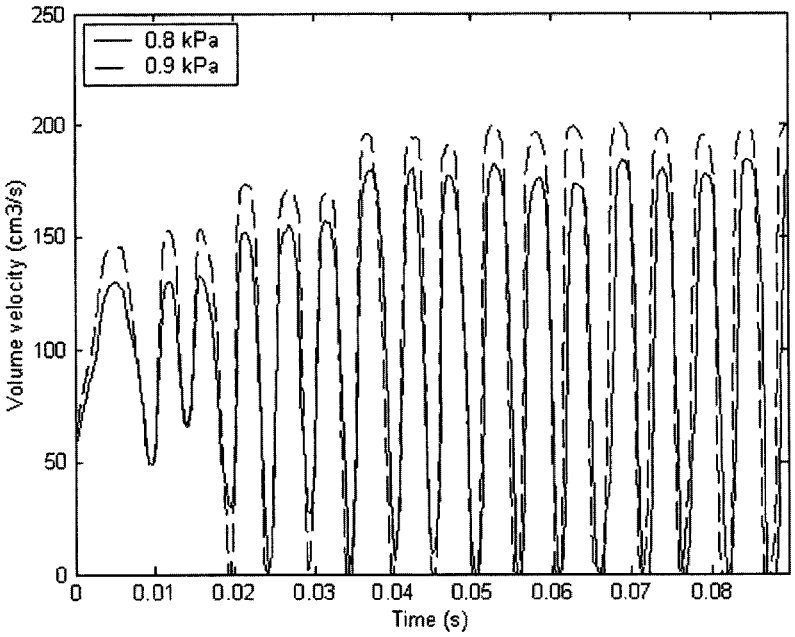


FIG. 8. Glottal signals obtained from stiffer larynxes excited by different pressure drops.

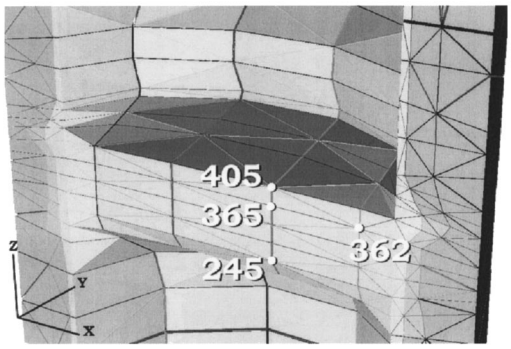


FIG. 9. Nodes whose movements are tracked along the time.

TABLE V. Description of the viscoelastic tissue parameters of a stiff larynx (specifically the transversal Young's module E for all tissues are higher than the ones presented for a normal larynx as described in Table II).

Tissue	E (kPa)	ν	E' (kPa)	μ'_m (kPa)	ν'	η (Pa.s)
Cover	3	0.73	200k	200k	0.73	1
Ligament	5	0.66	20	40	0.66	1
Body	20	0.45	20	30	0.45	15

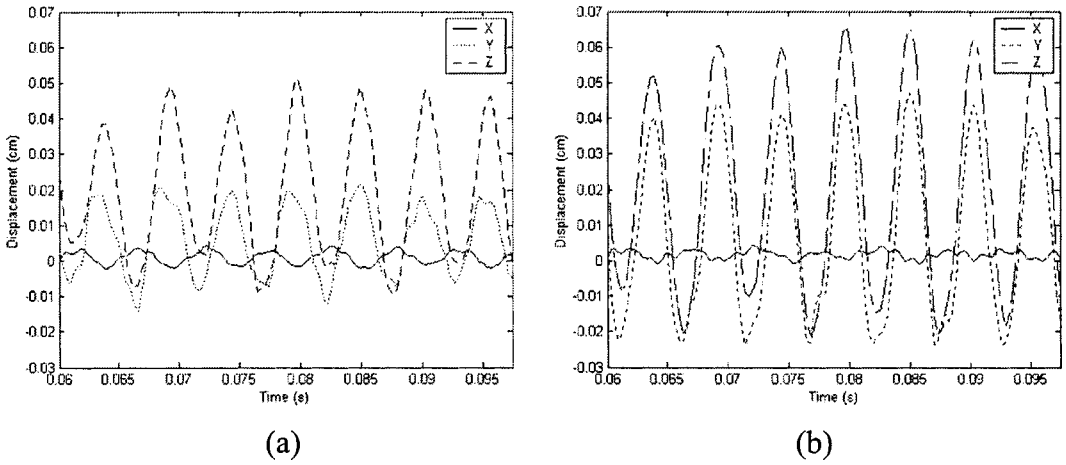


FIG. 10. Displacement of the nodes (a) 365 and (b) 405 along the three Cartesian axes.

complete simulation of the larynx with mechanical properties described in Table V.

Analyzing the coordinates of the points 365 and 405 (Fig. 10) the surface tissues near the glottal exit have higher amplitude than those near the glottal entrance. Considering that the mathematical modeling of the laryngeal tissues assumes that there is movement along the longitudinal axes, both points show displacements along this direction. Alipour *et al.* (2000) assumed such the movements in longitudinal direction are insignificant, allowing a reduction of mathematical complexity in their tissue modeling. Clearly their assumptions are correct, especially for the region near the glottis exit. However, the magnitudes along the three Cartesian axes becomes smaller as the distance from the glottis center becomes larger.

The periodicity degree seems to be higher for surface points located near the glottis exit. It can be related to the movement amplitude at that region. For nodes located far from the glottal exit, high-frequency components found in the longitudinal movements are no more insignificant in comparison with the movements in all other axes.

Additionally, an analysis of Fig. 11 reveals that the phase difference in the horizontal movement of the vocal folds are close to zero. This result is also found in Gunter (2003). The movement amplitudes of nodes 362 and 365 confirm that the main tissue displacements occur in the central portion of the vocal folds, which is caused by the pressure drop magnitude at this region.

Figure 12 shows that the movement of different points of vocal folds can present significant vertical phase difference in their movements. Concurrently with Fig. 11, movements in the *Y* direction are almost in-phase for the nodes 245, 365, and 405. Figure 12 clearly shows approximately 90° phase difference between nodes 245 and 405 in the *Z* direction. It is interesting to note that node 245 presents a curve whose its rise time is shorter than its fall time in the *Y* direction. This characterize the wavy behavior of the surface of vocal folds found in laryngoscope videos. These phenomena happen because node 245 is influenced by two concurrent forces: an aerodynamic force that always pushes it upward and a tissue force that is caused by the movements of nodes like 365 and 405. Therefore, the location where the vertical phase difference is a nonlinear function of mechanical and geometric properties of the laryngeal tissues and aerodynamic forces produced in phonation.

Considering the movement extension, the cover tissue of nodes at glottal exit (like node 405) suffers more compression and expansion then other laryngeal tissues. Gunter (2003) reports that collision forces acting at this node can be responsible for damages in tissue structures. The present re-

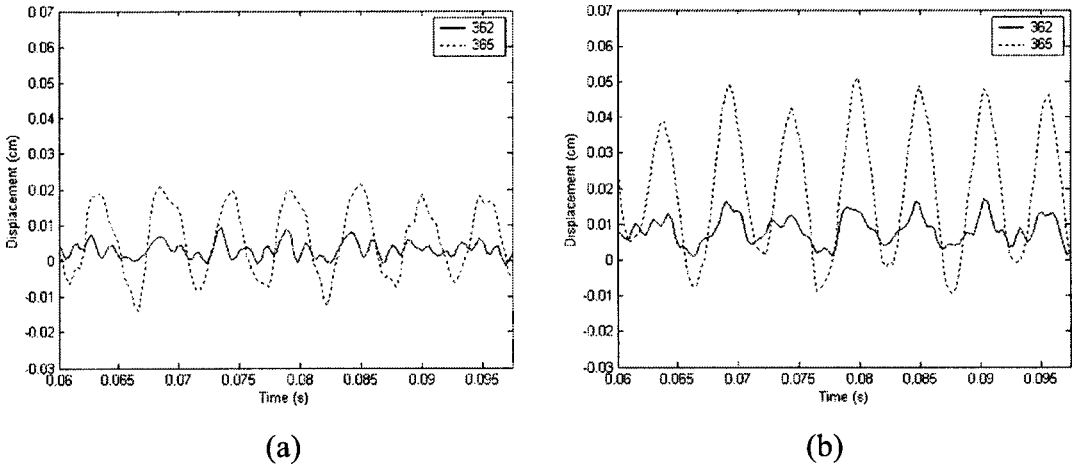


FIG. 11. Displacement of the nodes 362 and 365 along (a) *Y* and (b) *Z* directions.

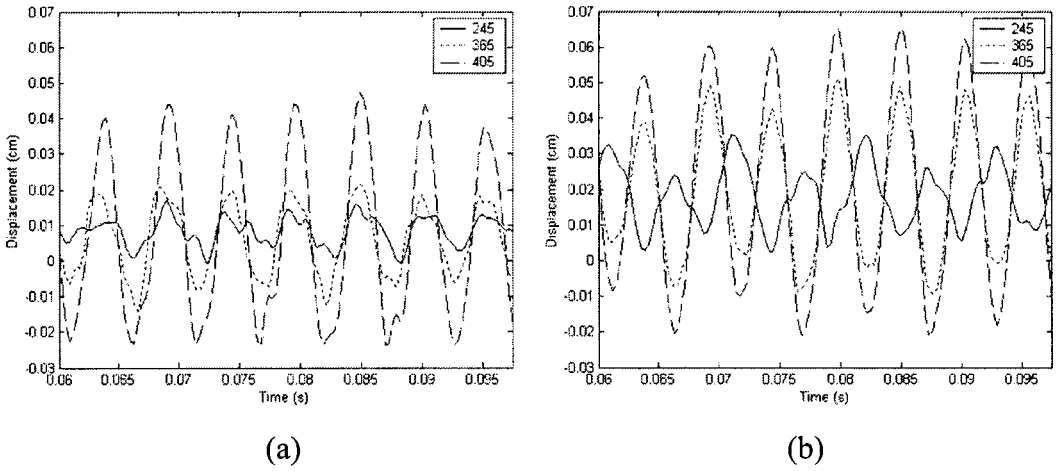


FIG. 12. Displacement of the nodes 245, 365, and 405 along (a) *Y* and (b) *Z* directions.

sults indicate that the large movements of glottal exit nodes can be a source of alterations in larynx tissues. It is important to consider that in phonatory conditions where both vocal folds barely reach each other, like in high-pitch voices, it is not the collision forces that are the cause for tissue damages (specially the superficial tissue layer of lamina propria) but the extreme nodal displacements as a consequence of high longitudinal and transverse stretch and aerodynamic forces.

In several larynx simulations, an additional phenomenon is observed: an incomplete glottal closure. Two spaces, located at both anterior and ''posterior'' comissures, are formed during the ''glottal closure'' time interval. These spaces are pointed by arrows in Fig. 13. These openings happen due to the incapacity of the aerodynamic forces to displace the laryngeal tissue located there during the phonation. Section III A shows that the air pressure applied over the larynx tissues is not constant and the main aerodynamic force is exerted over the central portion of the true vocal folds. Therefore, parts of these structures move first, bringing along comissure tissues. But this continuum effect is not enough to close the whole glottis. It is important to emphasize that the term ''posterior'' comissure refers to the point where both arythenoid cartilages meet each other during the phonatory process. By the fact that the present model is not representing all details—geometric and mechanical—of the larynx structures, careful consideration should be taken in order to analyze the presence of air leakage at ''posterior'' comissure.

Pelorson *et al.* (1994) considered such an incomplete closure in their model and mentioned that more realistic voices, from a perceptual view, are obtained when such airflow leakage is taken into account. As shown, the three-dimensional model presented here is able to reproduce such phenomena.

IV. CONCLUSIONS

This work presents the simulation of three-dimensional larynx models using the finite-element method as the main mathematical framework. The simulated models include false vocal folds and laryngeal ventricles which are first introduced in numeric simulation of larynx. The results show that these laryngeal structures affect the pressure distribution along the glottal walls and therefore modify the final glottal waveform.

The airflow velocities and pressures are obtained by numerically solving Navier–Stokes equations at each time step after the airflow mesh update. Mathematical strategies are used to speed up and control the convergence of the iterative methods employed to solve the sets of nonlinear equations. Different from Guo and Scherer (1993) and following Ishizaka and Flanagan (1972), the airflow is excited by a pressure drop.

Linear models describe the laryngeal tissues as Alipour *et al.* (2000). However, the collision between both true vocal folds during the glottal closure is modeled by calculating the surface forces that avoid the interpenetration between both tissues. This is the first time that this mathematical method is

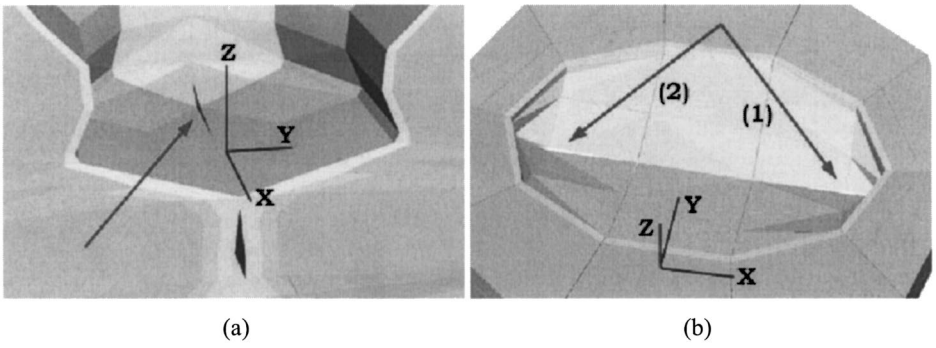


FIG. 13. Incomplete closure of the larynx: (a) at the ''posterior'' comissure and (b) at both the (1) anterior and (2) ''posterior'' comissures.

incorporated in whole larynx simulation under phonatory conditions although it is currently employed in mechanic computation.

The larynx meshes (both airflow and laryngeal tissues) are constructed by using hyper-ellipes. This method allows researchers to reconstruct larynx models in straightforward manner and reproduce the presented results. More realistic geometries can be obtained by manual manipulation of the larynx nodes or by automatic scanning of MRI images along with sophisticated automatic mesh generators.

Horizontal and vertical phase difference in the glottal tissue movements are captured by the larynx model and physically they are caused by the nonlinear air pressure distribution along the laryngeal surfaces and the viscoelastic constitution of the larynx tissues (specially in the glottal cavity).

Incomplete glottal closure during phonation are observed in some simulations, around anterior and posterior comissures. This phenomena is more evident in larynx with large glottal areas and/or with stiffer tissues. Pelorson *et al.* (1994) state that airflow escapes during phonation make the voice signal more perceptually realistic.

The lung pressure to fundamental frequency ratios obtained in the simulations with the present model are quite similar to the ones presented by Ishizaka and Flanagan (1972). However, the results show that these ratios are dependent on viscoelastic properties of the laryngeal tissues.

Statically the airflow for the three most significant laryngeal profiles are solved and the nonlinear behavior of the pressure distribution formed over the larynx walls justifies the vocal fold movements. The volume velocities are lower than the ones presented by Guo and Scherer (1993) due to the three-dimensional aspect of the larynx models presented here. In addition to the transversal funneling of the airflow along the larynx, there is another airflow funneling in sagittal plane. It is caused by the finite aspect of true three-dimensional larynxes and glottis shape which gets narrower in the direction of both comissures.

The next steps in larynx simulations require some model improvements. Larynx airflow and tissue meshes with higher numbers of nodes have to be used in order increase the accuracy of simulations although the inherit computational power expenses.

Additionally, sophisticated models to describe the viscoelastic nature of the larynx tissues should be studied, with special emphasis to their fibrous aspect. We believe that this is a key study that will bring more realistic results to the larynx simulations.

Another important aspect to improve the present model is to incorporate the airflow unsteady aspect into Navier–Stokes equations. The airflow mesh deforming also has to be put into the final partial differential equations in order to correctly compute the airflow velocities and pressures along the larynx. Work is ongoing to incorporate this missing aerodynamic aspect into the simulation model.

ACKNOWLEDGMENTS

This work was supported by FAPESP (The State of São Paulo Research Foundation, Brazil).

ANSYS (**2000**). *ANSYS 5.7 Theory Reference* (Ansys, Canonsburg).
Alipour, F., Berry, D., and Titze, I. (**2000**). ''A finite-element model of vocal-fold vibration,'' J. Acoust. Soc. Am. **108**, 3003–3012.
Bathe, K. (**1996**). *Finite Element Procedures* (Prentice–Hall, Upper Saddle River, NJ).
Berry, D., and Titze, I. (**1996**). ''Normal modes in a continuum model of vocal fold tissues,'' J. Acoust. Soc. Am. **100**, 3345–3354.
Berry, D., Herzel, H., Titze, I., and Krischer, K. (**1994**). ''Interpretation of biomechanical simulations and chaotic vocal fold oscillations with empirical eigenfunction,'' J. Acoust. Soc. Am. **95**, 3595–3604.
De Vries, M., Schutte, H., and Verkerke, G. (**1999**). ''Determination of parameters for lumped parameter models of the vocal folds using a finite-element method approach,'' J. Acoust. Soc. Am. **106**, 3620–3628.
Deblois, B. (**1996**). ''Quadratic, streamline upwinding for finite element method solutions to 2-D convective transport problems,'' Comput. Methods Appl. Mech. Eng. **134**, 107–115.
Demmel, J., Eisenstat, S., Gilbert, J., Li, X., and Liu, J. (**1999**). ''A supernodal approach to sparse partial pivoting,'' SIAM J. Matrix Anal. Appl. **20**, 720–755.
FIDAP (**1999**). *FIDAP 8 Theory Manual* (Fluent, Lebanon).
Gresho, P., and Sani, R. (**1998**). *Incompressible Flow and the Finite Element Method: Advection-diffusion and Isothermal Laminar Flow* (Wiley, Chichester).
Gunter, H. (**2003**). ''A mechanical model of vocal-fold collision with high spatial and temporal resolution,'' J. Acoust. Soc. Am. **113**, 994–1000.
Guo, C., and Scherer, R. (**1993**). ''Finite element simulation of glottal flow and pressure,'' J. Acoust. Soc. Am. **94**, 688–700.
Hirsch, C. (**1988**). *Numerical Computation of Internal and External Flows* (Wiley, Chichester), Vol. 2.
Huebner, K., and Thornton, E. (**1982**). *The Finite Element Method for Engineers* (Wiley, New York).
Ishizaka, K., and Flanagan, J. (**1972**). ''Synthesis of voiced sounds from a two-mass model of the vocal cords,'' Bell Syst. Tech. J. **51**(6), 1233–1268.
Johnson, A., and Tezduyar, T. (**1994**). ''Mesh update strategies in parallel finite element computations of flow problems with moving boundaries and interfaces,'' Comput. Methods Appl. Mech. Eng. **119**, 73–94.
Miller, J., Pereira, J., and Thomas, D. (**1988**). ''Fluid-flow through the larynx channel,'' J. Sound Vib. **121**, 277–290.
Narayanaswamy, O. (**1985**). ''Processing nonlinear multipoint constraints in the finite element method,'' Int. J. Numer. Methods Eng. **21**, 1283–1288.
Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). ''Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model,'' J. Acoust. Soc. Am. **96**, 3416–3431.
Peterson, J. (**1989**). ''The reduced basis method for incompressible viscous flow calculations,'' SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput. **10**, 777–786.
Sampaio, P. (**1991**). ''A Petrov-Galerkin formulation for the incompressible Navier-Stokes equations using equal order interpolation for velocity and pressure,'' Int. J. Numer. Methods Eng. **31**, 1135–1149.
Sani, R. L. and Gresho, P. M. (**1994**). ''Resume and remarks on the open boundary condition minisymposium,'' Int. J. Numer. Methods Fluids **18**, 1135–1149.
Tabata, M., and Fujima, S. (**1991**). ''An upwind finite element scheme for high-Reynolds-numfer flows,'' Int. J. Numer. Methods Fluids **12**, 305–322.
Titze, I., and Strong, W. (**1975**). ''Normal modes in vocal cord tissues,'' J. Acoust. Soc. Am. **57**, 736–744.
Titze, I., and Talkin, D. (**1979**). ''A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation,'' J. Acoust. Soc. Am. **66**, 60–74.
Tran, Q., Gerratt, B., Berke, G., and Kreiman, J. (**1993**). ''Measurement of Young's modulus in the in-vivo human vocal folds,'' Ann. Otol. Rhinol. Laryngol. **102**, 584–591.

Shaping function models of the phonatory excitation signal

Jean Schoentgen[a)]
Laboratory of Experimental Phonetics, Université Libre de Bruxelles, 50, Av. F.-D. Roosevelt, 1050 Brussels, Belgium

(Received 16 March 2003; revised 2 July 2003; accepted 4 August 2003)

The phonatory excitation signal is the acoustic signal that is generated at the glottis by the vibrating vocal folds and pulsatile airflow. A shaping function model is a nonlinear memoryless input–output characteristic that transforms a simple harmonic into the desired output. The model can be fitted linearly to observed or simulated template cycles. The instantaneous values of the excitation cycle centroid, amplitude as well as length, and the cues for phonatory identity are set via distinct parameters. The synthetic phonatory excitation signal is zero on average, as well as identically zero when the glottal airflow rate is constant. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1612485]

PACS numbers: 43.70.Bk, 43.70.Gr, 43.70.Ja [AL] Pages: 2906–2912

I. INTRODUCTION

The purpose of the article is the presentation of a model of the phonatory excitation signal. The phonatory excitation or glottal source signal is the acoustic signal that is generated by the vibrating glottis and pulsatile glottal airflow. The model is based on a nonlinear memoryless input–output characteristic that transforms a cosine function into the synthetic glottal source signal. The memoryless input–output characteristic is a truncated power series that is called a shaping or distortion function (Hartmann, 1997).

The spectral centroid, which is a cue of the brightness of a sound, is used to discuss the shape of the glottal source signal (Van Son and Van Santen, 1997; Harrington and Cassidy, 1999). Here, the spectral centroid is the ratio of the weighted and unweighted averages of the amplitudes of the harmonics. The weights are the orders of the harmonics.

The most popular models of the phonatory excitation signal, which are not unit pulses, are the Liljencrants–Fant and Klatt models (Fant *et al.*, 1985; Klatt and Klatt, 1990). These consist of curves whose parameters are chosen to describe simulated or observed phonatory cycles. The primitive of the Klatt model is used later in the text as a template of the glottal flow rate pulse. The properties of the Liljencrants–Fant model are discussed hereafter.

The Liljencrants–Fant model consists of the concatenation of an exponentially increasing sinusoid and a decaying natural exponential that simulates the so-called return phase. The model variable is time and the number of parameters is five, including the total cycle length.

Problems that must be solved when analyzing or synthesizing phonatory excitation signals by means of such a model are the following. First, the five model parameters are not independent. The reasons are that the phonatory excitation signal is an acoustic signal that must be zero on average as well as identically zero when the glottal flow rate is constant and the glottis immobile. Consequently, when one model parameter is changed deliberately to generate dynamic prosodic or phonatory quality cues, the values of the remaining parameters must be recalculated heuristically (Fant *et al.*, 1994).

Second, the model must be fitted to observed phonatory cycles by nonlinear constrained optimization (Strik, 1998).

Third, the interpretation of some of the parameters of the model is not obvious. Also, the distinction between parameters that generate prosodic and phonatory quality cues on the one hand and phonatory identity cues on the other is lost because the parameters are interdependent.

Fourth, the signal phase is not a model parameter. This may be a problem because the instantaneous angular frequency, which is the derivative with respect to time of the signal phase, is the physical quantity that is sufficient for describing changing cycle lengths (Schoentgen, 2001). When the total cycle length is used as a model parameter in place of the instantaneous length, any updates of the cycle length must be synchronized with the onsets or offsets of the excitation cycles. This synchrony has no physiological basis and adds programming overhead, needed to monitor the excitation cycles so as to update the model parameters once per cycle (Lalwani and Childers, 1991).

To conclude, connected-curve models may be difficult to fit to observed cycles and they do not enable the instantaneous values of the spectral centroid, cycle frequency, or cycle amplitude to be set directly. It is therefore proposed to simulate the phonatory excitation signal by means of shaping functions. These are truncated power series that transform a trigonometric function into the desired output.

The motivation for the proposal lies in the following desirable properties of such a model. First, the shaping model can be linearly fitted to observed or simulated template cycles. The fit is algorithmic and constraint-free. Second, the instantaneous phonatory frequency is set by a single parameter. A change in the instantaneous frequency of phonation leaves the values of the amplitude and spectral centroid of the cycle unaffected. Third, the value of the spectral centroid depends on a single parameter, which leaves the phonatory frequency unaffected. Fourth, the cycle amplitude is set by a linear scaling parameter. Fifth, the parameters of the model fall into two categories, that is, those that set the

a)Electronic mail: jschoent@ulb.ac.be

instantaneous values of the amplitude, spectral centroid and phonatory frequency on the one hand, and the power series coefficients that fix the default cycle shape on the other. Sixth, the excitation cycle is zero on average, as well as identically zero when the glottal airflow rate is constant.

In an earlier article, we have shown that shaping function (1) is a building block for a nonlinear model of the glottal airflow rate (Schoentgen, 1990). Symbol M is the order of the power series, the constants c_i are the power series coefficients, and n is the time index:

$$s_e(n)=c_0+c_1x(n)+c_2x^2(n)+\cdots+c_Mx^M(n). \tag{1}$$

The reason for the presence of nonlinear terms in expression (1) is that the dynamics of the vibration of the vocal folds is nonlinear and the reason for the omission of any delayed samples is that the length of the vibrating glottis is small compared to a typical speech wavelength. This model has the first five desired properties that are listed above. The sixth property takes into account that the glottal source signal is an acoustic signal. The purpose of this article is to show how building block (1) must be transformed so as to satisfy the sixth property automatically, while keeping intact the desirable properties of the earlier model.

II. MODELS

A. Shaping function models of cyclic signals

The requirement that the model should comprise the signal phase as a parameter suggests equating driving function $x(n)$ with $\cos[\theta(n)]$. This turns power series (1) into a model for even signals, because the powers of a cosine function are even, i.e., symmetric about the origin of the time axis. When the Fourier series coefficients a_i of the signal exist, the power series coefficients c_i can be obtained via the following system of equations (Schoentgen, 1990) (symbol M_e is a constant square matrix with $M+1$ rows, M is the number of harmonics):

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \cdots \\ a_M \end{pmatrix} = M_e \begin{pmatrix} c_0 \\ c_1/2 \\ c_2/4 \\ c_3/8 \\ \cdots \\ C_M/2^M \end{pmatrix}. \tag{2}$$

An odd function is antisymmetric about the origin of the time axis. An odd signal, s_0, can nevertheless be expanded into powers of cosines by observing that the ratio of an odd function and a sine function is even:

$$\frac{s_0(n)}{\sin[\theta(n)]}=d_0+d_1x(n)+d_2x^2(n)+\cdots+d_Mx^M(n). \tag{3}$$

When the Fourier series coefficients b_i exist, the power series coefficients d_i are obtained via the following system of equations (Schoentgen, 1990) (symbol M_0 is a constant square matrix with $M+1$ rows):

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ \cdots \\ b_{M+1} \end{pmatrix} = M_0 \begin{pmatrix} d_0/2 \\ d_1/4 \\ d_2/8 \\ d_3/16 \\ \cdots \\ d_M/2^{M+1} \end{pmatrix}. \tag{4}$$

The coefficients of matrices M_e and M_0 can be read off from a Pascal arithmetical triangle. The nonzero constants in matrix M_e are indeed identical to the coefficients of a Pascal triangle that is positioned such that its bisector is identical with the first row of the matrix. The nonzero constants in matrix M_0 are identical to the differences of two neighboring coefficients in a Pascal triangle (Schoentgen, 1990):

$$M_e=\begin{pmatrix} 1 & 0 & 2 & 0 & 6 & 0 & 20 & \cdots \\ 0 & 1 & 0 & 3 & 0 & 10 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 4 & 0 & 15 & \cdots \\ 0 & 0 & 0 & 1 & 0 & 5 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 1 & 0 & 6 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

$$M_0=\begin{pmatrix} 1 & 0 & 1 & 0 & 2 & 0 & 5 & \cdots \\ 0 & 1 & 0 & 2 & 0 & 5 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 3 & 0 & 9 & \cdots \\ 0 & 0 & 0 & 1 & 0 & 4 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 1 & 0 & 5 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

A signal that is neither even nor odd can be represented by a combination of shaping functions (1) and (3). The reason is that an arbitrary one-variable function can be decomposed into the sum of an even and an odd function. The expansion of a signal by its Fourier series is based on that property (Boyd, 2001). A Fourier series indeed decomposes a signal into its even component (the sum of cosines) and odd component (the sum of sines). Consequently, an arbitrary signal whose Fourier series coefficients are known can be expanded into a sum (5) of two power series (1) and (3). The coefficients of the power series are given by relations (2) and (4), which link the coefficients of shaping function (1) to the coefficients of the even sum of cosines and the coefficients of shaping function (3) to the coefficients of the odd sum of sines,

$$\begin{aligned} s(n)&=s_e(n)+s_0(n) \\ &=\sum_{i=0}^{M} c_i\cos^i[\theta(n)]+\sin[\theta(n)]\sum_{i=0}^{M} d_i\cos^i[\theta(n)]. \end{aligned} \tag{5}$$

When the aim is the generation rather than the expansion of signals, model (5) may be turned into a more versatile model (6) by inserting a cosine function with arbitrary amplitude A, as well as a linear scaling parameter G:

$$s(n)=G\left\{\sum_{i=0}^{M} c_i A^i \cos^i[\theta(n)] + A\sin[\theta(n)]\sum_{i=0}^{M} d_i A^i \cos^i[\theta(n)]\right\}. \quad (6)$$

B. Input–output relations

Model (6) generates cycles whose instantaneous amplitudes, frequencies, and spectral centroids can be chosen. The instantaneous frequency of an output cycle is the instantaneous frequency of the driving cosine and the cycle amplitude is scaled by linear gain G. The cycle shape depends on amplitude A of the driving cosine and on polynomial coefficients c_i and d_i. The latter depend on the Fourier series coefficients of a template cycle via matrix relations (2) and (4).

The values of the Fourier coefficients depend on the time origin of the template cycle. Equations (7) give the Fourier series coefficients of a cycle that has been time-shifted. Pair (a,b) is the coefficients of a template cycle and pair (a',b') is the coefficients of the same cycle that has been shifted by time t_0. Symbol $\Phi_0=2\pi t_0/T$ is the corresponding phase shift when T is the cycle length:

$$a'_n=a_n\cos(n\Phi_0)+b_n\sin(n\Phi_0),$$
$$b'_n=b_n\cos(n\Phi_0)-a_n\sin(n\Phi_0). \quad (7)$$

Matrix relations (8) give the Fourier series coefficients (a',b') of the model output for all possible template cycle origins and driving amplitudes as a function of the Fourier series coefficients (a,b) of the template cycle. Matrix relations (8) are obtained as follows. First, relations (7) are rewritten under matrix form. Second, matrix relations (2) and (4) are rewritten taking into account that coefficients c_i are replaced by c_iA^i and coefficients d_i by d_iA^{i+1} when model (6) is driven by a cosine whose amplitude A is arbitrary. Third, original relations (2) and (4) are inverted to replace polynomial coefficients (c,d) by Fourier coefficients (a,b) of the template cycle. Triangular matrixes M_e and M_o are invertible because the product of the elements of the main diagonal is equal to unity. Matrix M_e is replaced by matrix $M_{e,*}$, from which the first row and column have been removed, because in relations (7) the odd Fourier series coefficients do not involve even coefficient a_0:

$$\begin{bmatrix} a'_0 \\ a'_1 \\ a'_2 \\ \dots \\ a'_M \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & \cos\Phi_0 & 0 & \dots & 0 \\ 0 & 0 & \cos 2\Phi_0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \cos M\Phi_0 \end{bmatrix} \times M_e \begin{bmatrix} 1 & 0 & 0 & \dots & \dots \\ 0 & A & 0 & \dots & \dots \\ 0 & 0 & A^2 & \dots & \dots \\ 0 & 0 & 0 & \dots & \dots \\ \dots & \dots & \dots & \dots & A^M \end{bmatrix} M_e^{-1} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \dots \\ a_M \end{bmatrix}$$
$$+ \begin{bmatrix} \sin\Phi_0 & 0 & 0 & \dots & 0 \\ 0 & \sin 2\Phi_0 & 0 & \dots & 0 \\ 0 & 0 & \sin 3\Phi_0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \sin M\Phi_0 \end{bmatrix} \times M_o \begin{bmatrix} A & 0 & 0 & \dots & 0 \\ 0 & A^2 & 0 & \dots & 0 \\ 0 & 0 & A^3 & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & \dots & A^M \end{bmatrix} M_o^{-1} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \dots \\ b_M \end{bmatrix}, \quad (8)$$

$$\begin{bmatrix} b'_1 \\ b'_2 \\ b'_3 \\ \dots \\ b'_M \end{bmatrix} = \begin{bmatrix} \cos\Phi_0 & 0 & 0 & \dots & 0 \\ 0 & \cos 2\Phi_0 & 0 & \dots & 0 \\ 0 & 0 & \cos 3\Phi_0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \cos M\Phi_0 \end{bmatrix} \times M_o \begin{bmatrix} A & 0 & 0 & \dots & 0 \\ 0 & A^2 & 0 & \dots & 0 \\ 0 & 0 & A^3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & A^M \end{bmatrix} M_o^{-1} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \dots \\ b_M \end{bmatrix}$$
$$- \begin{bmatrix} \sin\Phi_0 & 0 & 0 & \dots & 0 \\ 0 & \sin 2\Phi_0 & 0 & \dots & 0 \\ 0 & 0 & \sin 3\Phi_0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \sin M\Phi_0 \end{bmatrix} \times M_{e,*} \begin{bmatrix} A & 0 & 0 & \dots & \dots \\ 0 & A^2 & 0 & \dots & \dots \\ 0 & 0 & A^3 & \dots & \dots \\ 0 & 0 & 0 & \dots & \dots \\ \dots & \dots & \dots & \dots & A^M \end{bmatrix} M_{e,*}^{-1} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \dots \\ a_M \end{bmatrix}.$$

By inspection, one may infer from relations (8) the following. First, when phase $\Phi_0=0$ and driving amplitude A

$=1$, coefficients a_i' and b_i' are equal to a_i and b_i. That is, model (6) outputs the template cycle shape.

Second, when phase $\Phi_0 \neq 0$ and driving amplitude $A=1$, relations (8) are identical with relations (7) in matrix form. That is, model (6) outputs the template cycle shifted by time t_0.

Third, when driving amplitude $A<1$, powers A^i become negligibly small when exponent $i \geqslant k$, where threshold k depends on the value of driving amplitude A. Consequently, rows k to M of the diagonal matrices that involve the powers of A are nearly zero. Further, matrices M_e and M_o as well as M_e^{-1} and M_o^{-1} are triangular. The Fourier series coefficients a_i' and b_i' of order k to M are therefore omitted from the output. That is, model (6) outputs a cycle shape whose spectral centroid decreases with driving amplitude A.

Fourth, when phase $\Phi_0 \neq 0$ and driving amplitude $A<1$, the previous property as well as the property preceding the previous one are involved. That is, the output of model (6) is time-shifted and the spectral centroid decreases with A as before.

C. Shaping function models of the glottal source signal

A property of model (6) is that the average of an output cycle is different from zero when the value of the driving amplitude is different from unity. This is the case even when the average of the template cycle is zero. This is shown by approximation (9) of model (6), which is valid when amplitude $A \approx 0$. Output (9) is incompatible with the observation that acoustic waves do not implicate a net change of atmospheric pressure:

$$s(n) \approx G\{c_0 + c_1 A \cos[\theta(n)] + d_0 A \sin[\theta(n)]\}. \tag{9}$$

This section shows how to transform model (6) so as to force the model output to be zero on average, as well as identically zero when the glottal airflow rate is constant. In the framework of glottal source signal analysis and synthesis, many authors use the terms flow rate or volume velocity interchangeably because these are hypothesized to have the same form, which is a pulse that is skewed to the right of the time axis (Childers, 2000). In this article, the term flow rate is preferred. Several models of the flow rate or volume velocity signal exist (Rosenberg, 1971; Fant *et al.*, 1985; Klatt and Klatt, 1990).

Generally speaking, the shape of the phonatory excitation signal is heuristically predicted by the derivative with respect to time of the flow rate or volume velocity (Fant, 1997; Stevens, 1998; Childers, 2000). This would imply that the amplitude of the phonatory excitation evolves proportionally to the instantaneous phonatory frequency. The reason is that the derivative with respect to time of a cyclic signal is proportional to the instantaneous frequency of the signal.

However, extant models enable the cycle length as well as the amplitude of the negative-going peak of the glottal excitation signal to be set independently (Fant *et al.*, 1985; Klatt and Klatt, 1990). Published data indeed confirm that phonatory frequency and amplitude of the negative-going peak do not evolve proportionally. Since the negative-going peak co-determines the sound pressure level, these data can be corroborated indirectly by inspecting phonetograms, which display a speaker's sound pressure level and phonatory frequency (Gauffin and Sundberg, 1989).

Possible explanations for the lack of proportionality between phonatory frequency and peak amplitude are that the latter depends on phonatory frequency as well as subglottal pressure. Another reason is that the control of the phonatory frequency as such does not only involve the frequency of vibration of the glottal walls, but also the shape of the glottal space that co-determines the pulse shape of the glottal flow rate and therefore the pulse amplitude of the phonatory excitation (Gauffin and Sundberg, 1989; Titze, 1989).

In the framework of speech synthesis, instantaneous frequency and amplitude of the phonatory excitation signal must therefore be controlled independently to simulate real speech. A strictly proportional evolution of instantaneous frequency and amplitude would consequently be perceived as a nuisance by the experimenter. The undesired dependency of the signal amplitude on the signal frequency can be avoided in the framework of the shaping function model by replacing the derivative of model (6) with respect to time by a derivative, $s'(n)$, with respect to phase:

$$\begin{aligned} s'(n) = G\Bigg\{ & -A \sin[\theta(n)] \sum_{i=1}^{M} c_i A^{i-1} i \cos^{i-1}[\theta(n)] \\ & + \sum_{i=0}^{M} d_i A^{i+1} \cos^{i+1}[\theta(n)] \\ & - A^2 \sin^2[\theta(n)] \sum_{i=1}^{M} d_i A^{i-1} i \cos^{i-1}[\theta(n)] \Bigg\}. \end{aligned} \tag{10}$$

Coefficients c_i and d_i of model (10) are the coefficients of model (6) of the template flow rate. Properties that are not shared by models (6) and (10) are the following. First, phonatory excitation (10) is identically zero when the template flow rate is constant. The reason is that model (10) is a model of the derivative of a model of the glottal flow rate. Second, the average of a cycle of phonatory excitation (10) is zero because model (10) involves the derivative of a template whose initial and final samples are identical owing to cyclicity. Third, output (10) is a sinusoidal function when the driving amplitude is small, and zero when the driving amplitude is zero. This is shown by approximation (11) of model (10), which is valid when the amplitude $A \approx 0$. Model (10) therefore possesses the desired properties that are posited in the introduction section:

$$s'(n) \approx G\{-c_1 A \sin[\theta(n)] + d_0 A \cos[\theta(n)]\}. \tag{11}$$

III. METHODS

A. Flow rate template

The template cycle of the flow rate is based on a model proposed by Klatt (Klatt and Klatt, 1990). The cycle consists of a zero closed-glottis flow and a polynomial $et^2 - ft^3$ for the open-glottis flow. Symbol t is the time variable and constants e and f depend on the desired pulse length and maxi-

mal flow rate. Subsequently, the flow rate cycle is low-pass filtered by a second-order linear filter to simulate the gradual return of the pulse towards zero when the glottis closes. The sampling frequency of the flow rate template is 20 kHz and the number of harmonics 45. The cycle origin agrees with $t=0$ at which time the flow rate is zero. The Fourier sum coefficients are computed by a conventional method (Duffy, 1997). Truncation errors should be kept as small as possible when computing the Fourier series coefficients numerically. Also, care should be exercised to avoid spurious discontinuities at the beginning and end of the cycle when isolating a simulated or observed discrete flow rate pulse.

B. Shaping function model of the phonatory excitation signal

Synthesis of the phonatory excitation consists in inserting the power series coefficients c_i and d_i of the flow rate template into shaping function (10), driving the shaping function by a cosine that has the desired instantaneous frequency and amplitude, and scaling the output via the linear gain. The sampling frequency is 20 kHz.

To avoid aliasing, the product of the frequency of the driving cosine times the order of the shaping functions must be less than half the sampling frequency when the amplitude and frequency of the driving cosine are constant. For the same reason, the product of the upper bound of the effective bandwidth of the driving cosine times the order of the shaping function must be less than half the sampling frequency when the parameters of the driving cosine are modulated (Schoentgen, 2003).

C. Control of the spectral centroid of the phonatory excitation signal

This section summarizes the properties of the synthetic output when shaping function model (10) is driven by a cosine whose amplitude is different from unity. First, driving the model by a cosine whose amplitude is different from unity is an option. When the amplitude is unity, model (10) outputs cycles the instantaneous frequency of which is identical to the instantaneous frequency of the driving cosine and whose shape is the derivative with respect to phase of the template flow rate. The output amplitude may be controlled via linear gain G.

Second, the amplitude of the driving cosine must lie in practice between 0 and $+1$. The reason is that polynomial shaping functions (1) and (3) are defined between -1 and $+1$, because the polynomial coefficients are calculated to restitute the template cycle when the amplitude of the driving cosine is unity.

Formally, model (6) or (10) may be driven by a cosine function whose amplitude exceeds unity. Then the model outputs synthetic cycle shapes that are extrapolated and that are only plausible when the excess amplitude is small. The reason is that the shaping functions involve powers A^i of the driving amplitude. An arbitrarily large increase of the driving amplitude would therefore enable the spectral centroid of the cycle shape of model (6) to be increased to its maximum, the asymptote of which is M. This asymptote is approached when driving amplitude A becomes so large that all powers of A less than the maximum power become negligible.

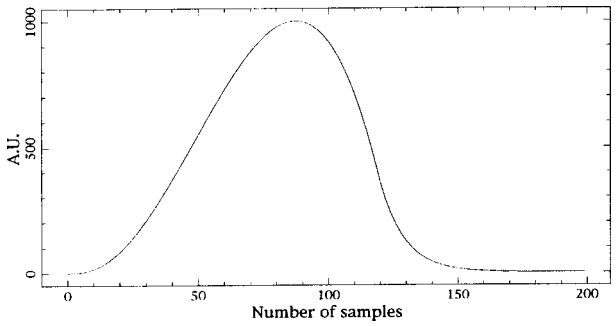


FIG. 1. Klatt flow rate template that is at the base of the shaping function model of the phonatory excitation. The vertical axis is in arbitrary units; the horizontal axis is in number of samples; the sampling frequency is 20 kHz.

The spectral centroid of the flow rate or phonatory excitation must, however, be less than $(M+1)/2$ because the slope of the envelope of the partials must be negative (Monsen and Engebretson, 1977). The centroid threshold $(M+1)/2$ corresponds to a spectrum that is flat. This threshold is reached for driving amplitudes in excess of unity that become smaller when the model order increases. In practice, the cycle shapes that are outputted by the model become implausible before this threshold is reached.

Third, Eqs. (8) show that the relation is one-to-one between the driving amplitude $0 \leqslant A \leqslant 1$ and the size of the partials of the output of model (6). The amplitudes of the partials decrease monotonically with the driving amplitude and the decrease is faster the higher the order of the partials. Consequently, model (11) shows that small driving amplitudes produce outputs whose amplitudes and spectral centroids are small. Models (6) and (10) of the flow rate and phonatory excitation therefore belong with the laryngeal vibrator to the class of systems that output signals whose spectral centroid values decrease with overall amplitude (Krishnamurthy and Childers, 1986; Stevens, 1998). Property (11) is therefore a desirable feature for simulating vocal onsets and offsets.

Fourth, when the values for the spectral centroid and overall signal amplitude must be set independently, the value of the spectral centroid may be fixed by the amplitude of the driving cosine and the overall amplitude by the linear gain. In practice, this scheme may be implemented by pretabulating for a given model the relations between driving amplitude, spectral centroid, and output amplitude when the linear gain is equal to unity. During synthesis, linear gain and driving amplitude may then be selected to obtain the desired output amplitude and spectral centroid.

IV. RESULTS

Figure 1 is the shape of the Klatt flow rate template. Figure 2(a) is a phonatory excitation signal that has been obtained by inserting the power series coefficients of the template into shaping function (10). The constant phonatory frequency is equal to 100 Hz. Figure 2(b) is the same model with the instantaneous frequency of the driving function evolving linearly from 150 to 100 Hz in the central 50% of

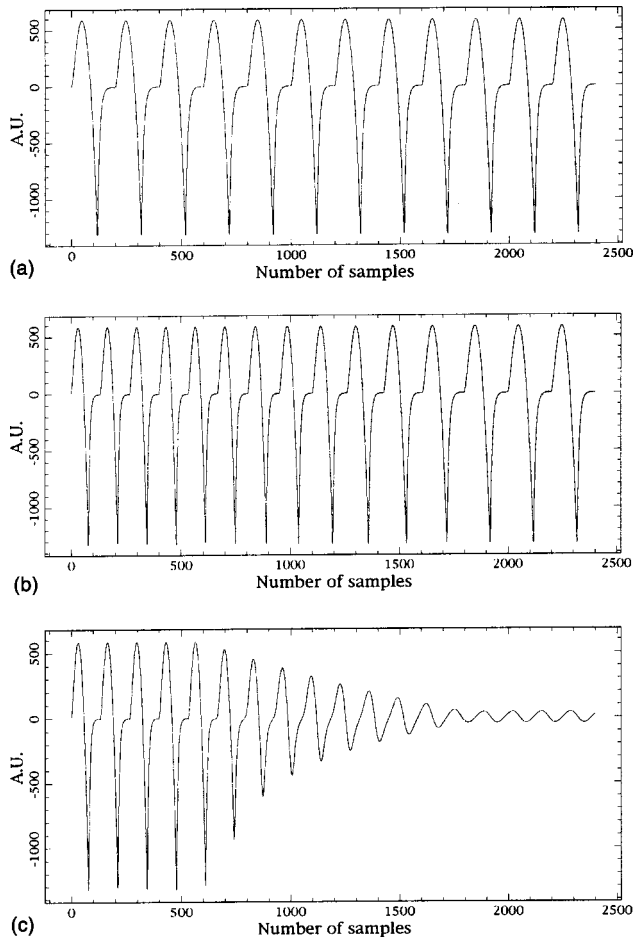


FIG. 2. (a) Phonatory excitation signal that is obtained by the Klatt flow rate template and shaping function model (10). The constant phonatory frequency is equal to 100 Hz. (b) The same model with the instantaneous frequency of the driving function evolving linearly in the central 50% of the graph from 150 to 100 Hz. (c) The same model with the amplitude of the driving function evolving linearly in the central 50% of the graph from 1 to 0.1. The sampling frequency is 20 kHz; the vertical axis is in arbitrary units; the horizontal axis is in number of samples.

the graph. Finally, Fig. 2(c) is the same model with the amplitude of the driving function evolving linearly from 1 to 0.1 in the central 50% of the graph. As predicted by relation (11), the shape of the synthetic excitation is nearly sinusoidal when the amplitude of the driving cosine is small. The phonatory frequency is equal to 150 Hz.

V. DISCUSSION AND CONCLUSION

In this article, the template for the flow rate pulse is based on the Klatt model. In Schoentgen (1990) shaping function models have been constructed based on observed glottal source signals that have been numerically integrated to obtain heuristic flow rate shapes. These models demonstrate that it is possible to use observed signals as templates. One drawback is that observed signals are frequently noisy. A consequence is that the synthetic signals may be perceived as hoarse when the additive noise has not been removed.

The motivation for the development of the shaping function model of the phonatory excitation signal is that the model has desirable properties that are the following. It can be fitted linearly and noniteratively to observed or simulated flow rate cycles. The instantaneous frequency of the synthetic phonatory excitation can be varied continuously and asynchronously. The instantaneous amplitude, frequency, and spectral centroid of the synthetic excitation, as well as the shaping function coefficients that encode phonatory identity, can be set independently. Finally, the synthetic excitation signal is zero on average and identically zero when the glottal flow rate is a constant.

These properties are desirable for the following reasons. First, extant models are fitted via nonlinear constrained optimization. The linear noniterative fit of the shaping function model precludes the risk, associated with optimization, of obtaining suboptimal solutions. Second, the continuous control of the instantaneous frequency of the phonatory excitation cycles omits the a-physiological synchronization of the updating of the total glottal cycle lengths with the cycle onsets or offsets, as well as the concomitant programming overhead. Third, the built-in property that the cycle is zero on average and that it is identically zero when the glottal flow rate is a constant confers to the output cycles the properties of an acoustic signal for any choice of the model parameters. In models that are without that property, when one parameter is changed deliberately to simulate dynamic prosodic or phonatory quality cues, all the model parameters are heuristically recalculated sample by sample or, alternatively, the condition is relaxed that the numerically simulated speech signal must be an acoustic signal (Fant *et al.*, 1994; Gerrat and Kreiman, 2001).

ACKNOWLEDGMENTS

This work was supported by the National Fund for Scientific Research, Belgium, and has been carried out in the framework of an *Action de Recherche Concertée (98-02; 226), Communauté Française de Belgique.*

Boyd, J. P. (**2001**). *Chebyshev and Fourier Spectral Methods* (Dover, Mineola, NY), pp. 159–171.

Childers, D. G. (**2000**). *Speech Processing and Synthesis Toolboxes* (Wiley, New York), pp. 308–319.

Duffy, D. G. (**1997**). *Advanced Engineering Mathematics* (CRC, Boca Raton), pp. 97–111.

Fant, G. (**1997**). ''Acoustical analysis of speech,'' in *Encyclopedia of Acoustics*, edited by M. J. Crocker (Wiley, New York), Vol. IV, pp. 1589–1598.

Fant, G., Liljencrants, J., and Lin, Q. (**1985**). ''A four-parameter model of glottal flow,'' Speech Transmiss. Lab.-Quart. Progr. Stat. Rep. Vol. 4, pp. 1–13.

Fant, G., Kruckenberg, A., Liljencrants, J., and Bavegard, M. (**1994**). ''Voice source parameters in continuous speech. Transformation of LF-parameters,'' in *Proc. Int. Conf. Spoken Language Processing*, Yokohama, Vol. 3, pp. 1451–1454.

Gauffin, J., and Sundberg, J. (**1989**). ''Spectral correlates of glottal voice source waveform characteristics,'' J. Speech Hear. Res. **32**, 556–565.

Gerrat, B. R., and Kreiman, J. (**2001**). ''Measuring vocal quality with speech synthesis,'' J. Acoust. Soc. Am. **110**, 2560–2566.

Harrington, J., and Cassidy, S. (**1999**). *Techniques in Speech Acoustics* (Kluwer, Dordrecht), p. 173.

Hartmann, W. M. (**1997**). *Signals, Sound, and Sensation* (AIP, Woodbury, NY), pp. 491–521.

Klatt, D., and Klatt, L. (**1990**). ''Analysis, synthesis, and perception of voice quality variations among female and male talkers,'' J. Acoust. Soc. Am. **87**, 820–857.

Krishnamurthy, A. K., and Childers, D. G. (**1986**). ''Two-channel speech analysis,'' IEEE Trans. Acoust., Speech, Signal Process. **34**, 730–743.

Lalwani, A. L., and Childers D. G. (**1991**). ''Modeling vocal disorders via formant synthesis,'' in Proceedings IEEE Conf. Acoustics, Speech, Signal Processing, pp. 505–508.
Monsen, R. B., and Engebretson, A. M. (**1977**). ''Study of variations in the male and female glottal wave,'' J. Acoust. Soc. Am. **62**, 981–993.
Rosenberg, A. (**1971**). ''Effect of glottal pulse shape on the quality of natural vowels,'' J. Acoust. Soc. Am. **49**, 583–590.
Schoentgen, J. (**1990**). ''Nonlinear signal representation and its application to the modeling of the glottal waveform,'' Speech Commun. **9**, 189–201.
Schoentgen, J. (**2001**). ''Stochastic models of jitter,'' J. Acoust. Soc. Am. **109**, 1631–1650.
Schoentgen, J. (**2003**). ''On the bandwidth of a shaping function model of the phonatory excitation signal,'' in *Proceedings ISCA Tutorial and Research Workshop Nonlinear Speech Processing*, edited by F. Bimbot (Le Croisic, France), pp. 123–126.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA), pp. 55–100.
Strik, H. (**1998**). ''Automatic parameterization of differentiated glottal flow: Comparing methods by means of synthetic flow pulses,'' J. Acoust. Soc. Am. **103**, 2659–2669.
Titze, I. (**1989**). ''A four-parameter model of the glottis and vocal fold contact area,'' Speech Commun. **8**, 191–201.
Van Son, R. J. J. H., and Van Santen, J. P. H. (**1997**). ''Word-level prosodical marking of consonant duration and spectral balance,'' Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam, Vol. 21, pp. 21–36.

Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers

Christopher J. Darwin[a)]
Experimental Psychology, University of Sussex, Brighton BN1 9QG, United Kingdom

Douglas S. Brungart
Air Force Research Laboratory, Human Effectiveness Directorate, Wright-Patterson AFB, Ohio 45433

Brian D. Simpson[b)]
Veridian, 5200 Springfield Pike, Suite 2000, Dayton, Ohio 45431

(Received 15 June 2002; revised 9 August 2003; accepted 11 August 2003)

Three experiments used the Coordinated Response Measure task to examine the roles that differences in *F*0 and differences in vocal-tract length have on the ability to attend to one of two simultaneous speech signals. The first experiment asked how increases in the natural *F*0 difference between two sentences (originally spoken by the same talker) affected listeners' ability to attend to one of the sentences. The second experiment used differences in vocal-tract length, and the third used both *F*0 and vocal-tract length differences. Differences in *F*0 greater than 2 semitones produced systematic improvements in performance. Differences in vocal-tract length produced systematic improvements in performance when the ratio of lengths was 1.08 or greater, particularly when the shorter vocal tract belonged to the target talker. Neither of these manipulations produced improvements in performance as great as those produced by a different-sex talker. Systematic changes in both *F*0 and vocal-tract length that simulated an incremental shift in gender produced substantially larger improvements in performance than did differences in *F*0 or vocal-tract length alone. In general, shifting one of two utterances spoken by a female voice towards a male voice produces a greater improvement in performance than shifting male towards female. The increase in performance varied with the intonation patterns of individual talkers, being smallest for those talkers who showed most variability in their intonation patterns between different utterances. [DOI: 10.1121/1.1616924]

PACS numbers: 43.71.Bp, 43.71.Es [PFA] Pages: 2913–2922

I. INTRODUCTION

Over the past 50 years, numerous researchers have studied the ability of human listeners to extract information from a target speech signal that is masked by one or more competing talkers [see Ericson and McKinley (1997) and Bronkhorst (2000) for recent reviews of this literature]. This ''cocktail party'' listening task is particularly difficult when the target and masking speech signals are mixed together into a single channel and then presented monaurally or diotically over headphones. In this condition, binaural speech segregation cues, which are typically available in real-world listening situations, are absent and listeners must rely on monaural cues to perform the task. Although there is some evidence that listeners can use differences in the overall levels of the two voices to perform the segregation task (Egan *et al.*, 1954; Brungart, 2001), the most powerful monaural speech segregation cues seem to be related to differences in the vocal characteristics of the competing talkers (Bregman, 1990; Darwin and Hukin, 2000; Brungart, 2001). Brungart, for example, found that phrases spoken by different-sex talkers were substantially easier to segregate than phrases spoken by same-sex talkers, and that phrases spoken by two different same-sex talkers were substantially easier to segregate than two phrases spoken by the same talker.

To this point, however, little is known about the relative contributions that different voice characteristics make to the voice segregation process. Differences in the overall long-term spectra of the competing speech signals do not seem to have much influence on performance: Festen and Plomp (1990) found little difference between the intelligibility of a speech signal masked by speech-shaped noise with the overall spectrum of a same-sex talker and the intelligibility of a speech signal masked by speech-shaped noise with the overall spectrum of a different-sex talker.

Two relatively simple physical characteristics can be manipulated to change the apparent gender of a voice: the fundamental frequency (*F*0) range, and the vocal-tract length. Women's voices are typically a little under an octave higher in *F*0 than men's, and women's formant frequencies are around 16% higher than men's as a result of male vocal tracts being longer (Peterson and Barney, 1952). Voice individuality is lost if formants are shifted by 8% (Kuwabara and Takagi, 1991), but the successful digital transformation of voice gender normally requires manipulation of both *F*0 and vocal-tract length (Atal and Hanauer, 1971).

Of these two parameters, only differences in the *F*0 of the talkers and differences in the prosodic features of the

[a)]Electronic mail: cjd@biols.susx.ac.uk
[b]Currently at Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio.

competing utterances have been shown to produce some improvement in voice segregation (Brokx and Nooteboom, 1982; Scheffers, 1983; Assmann and Summerfield, 1990; Bird and Darwin, 1998; Darwin and Hukin, 2000). We know of no studies that have systematically examined both *F*0 and vocal-tract length while preserving the small temporal and prosodic variations that normally occur in repeated utterances spoken by the same talker. In this series of experiments, we examine diotic speech segregation when each of two phrases is spoken by the same talker, similar to the ''same-talker'' (TT) condition examined by Brungart (2001). However, here the speech signals were electronically modified to introduce differences in *F*0 (experiment 1), vocal-tract length (experiment 2), or both *F*0 and vocal-tract length (experiment 3) between the two competing phrases. The results provide valuable insights into the roles that *F*0, vocal-tract length, and target and masker gender play in two-talker speech segregation.

II. EXPERIMENT 1: CHANGES IN *F*0 ONLY

A. Method

The stimuli were derived from the publicly available Coordinate Response Measure speech corpus (Bolia *et al.* 2001). This corpus, which has been used in previous multitalker listening experiments (Brungart, 2001; Brungart *et al.*, in press), consists of sentences of the form ''Ready ⟨call sign⟩ go to ⟨color⟩ ⟨number⟩ now'' spoken with all 32 possible combinations of four colors (''red,'' ''blue,'' ''white,'' and ''green'') and eight numbers (1–8). The present experiments used all eight talkers available in the corpus (four male, four female), but only used four of the eight available call signs (''Arrow,'' ''Tiger,'' ''Eagle,'' and ''Baron''). Thus, a total of 1024 different sentences (8 talkers×4 call signs×4 colors×8 numbers) were used to produce the stimuli employed in the experiments

These 1024 sentences were down-sampled from 40 to 20 kHz, and then processed with the PSOLA algorithm (Moulines and Charpentier, 1990), as implemented in Macintosh version 3.9.28 of the Praat software package (Boersma and Weenink, 1996), to produce six new sets of speech files with *F*0 contours shifted by various numbers of semitones. The sentences spoken by female talkers were shifted by −9, −3, −1, 0, 1, or 3 semitones. The sentences spoken by male talkers were shifted by −3, −1, 0, 1, 3, or 9 semitones. The resulting speech quality was excellent for the range of *F*0 shifts that we used in the experiment.

Each stimulus consisted of a diotic mixture of two sentences spoken by the same talker: one sentence (the ''target'') contained the call sign ''Baron'' and a randomly selected color and number; the other sentence (the ''masker'') was randomly selected from all of the sentences in the corpus with a different call sign, color, and number than the target sentence. The *F*0 shifts of the two sentences were selected from the combinations shown in Table I to produce one of eight different absolute *F*0 differences: 0, 1, 2, 3, 4, 6, 9, or 12 semitones. The masking speech was also scaled to set its rms power to one of six different signal-to-noise ratios (SNRs) relative to the rms power of the target speech: −6, −3, 0, +3, +6, or +9 dB. Finally, as in previous CRM experiments, the overall output was randomly roved over a 6-dB range (in 1-dB increments) before a D/A converter (Tucker-Davis Technologies DD-1) was used to present the stimulus to the listener through headphones (Sennheiser HD-540) at a comfortable listening level (60–70 dB SPL).

TABLE I. *F*0 differences (in semitones) tested in the stimuli of experiment 1. Each pair of sentences in a trial was chosen from the same talker. The *F*0 shifts of each talker were selected differently for male and female talkers to place most of the *F*0 shift into the range between male and female speech. Thus, male talkers were generally shifted up more than down, and female talkers were generally shifted down more than up. Note that, on a given trial, the target sentence was equally likely to receive the higher or lower *F*0 shift.

	Male		Female	
Δ*F*0 (Semitones)	Lower *F*0	Higher *F*0	Lower *F*0	Higher *F*0
0	0	0	0	0
1	0	+1	0	−1
2	−1	+1	+1	−1
3	0	+3	0	−3
4	−1	+3	+1	−3
6	−3	+3	+3	−3
9	0	+9	0	−9
12	−3	+9	+3	−9

The experiment was conducted with the listeners seated at the CRT of a control computer in a sound-treated listening room. The listeners' task in each trial was to identify the color and number spoken in the target phrase, which was identified by the presence of the call sign ''Baron.'' Responses were made by using the computer mouse to select the appropriate color–number combination from an array of 32 colored numbers shown on the CRT. Nine paid volunteer subjects (four male, five female) with normal hearing participated in the study. All were native English speakers from the midwestern United States. Each of these listeners completed ten trials for each of the 384 possible stimulus configurations of the experiment (8 talkers×8 *F*0 differences×6 SNRs), for a total of 3840 trials per listener. These trials were completed in blocks of 192 trials, with separate blocks for the male and female talkers in the corpus, and with the trials within each block randomly balanced to have an equal number of trials for each value of each of the three independent variables (talker, *F*0 difference, and SNR). Each block took approximately 10 min to complete and each listener participated in two to three blocks per day over a 2-week period.

B. Results

1. Overall effects of shift in F0

The overall results averaged across all eight talkers are shown in Fig. 1. The left panel of the figure shows the probability of a correct identification of both the color and number in the target sentence as a function of the signal-to-noise ratio and the absolute *F*0 separation between the two sentences in semitones.

The 0-semitone data represent a stimulus condition that is very similar to the ''same talker'' (''TT'') condition presented in Brungart (2001). The only difference is that the

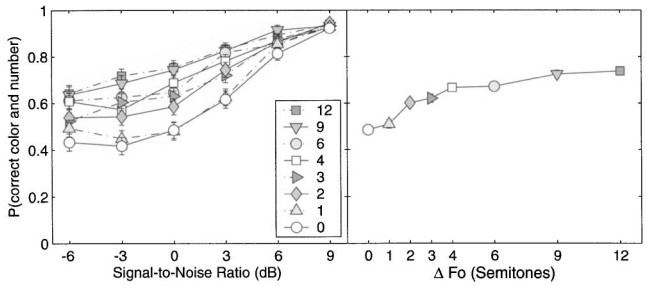


FIG. 1. The left panel shows the probability of a correct identification of the number and color in the target sentence as a function of SNR for eight different separations in semitones of the *F*0 contours between the target and the masker sentence. The right panel shows performance averaged across the lowest four SNRs (−6, −3, 0, and 3 dB) as a function of the separation in *F*0. The data have been averaged over all eight talkers. The error bars represent the 95% confidence intervals of the mean.

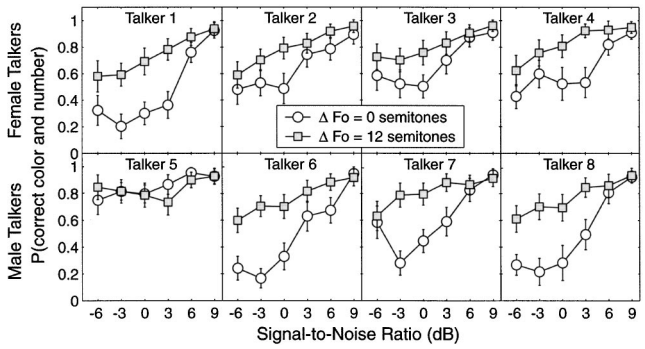


FIG. 2. Data as in the left panel of Fig. 1 but from individual talkers with either zero semitones (open circles) or 12 semitones (shaded squares) of separation in *F*0 between target and masker sentences. The upper row gives the results from the four female talkers (1–4) and the lower row from the four male talkers (5–8).

speech in this experiment has been analyzed and resynthesized by the PSOLA algorithm (without introducing any shift in *F*0). This analysis and synthesis did not appear to have any substantial effects on performance in the CRM task: the present results are nearly identical to those from the previous experiment, with performance declining as SNR decreases from over 90% at +9 dB to around 40% at 0 dB and then leveling out at SNRs less than 0 dB. The results of a *posthoc* pairwise comparison test (Fisher LSD) performed on the arcsine-transformed percentages of correct responses from the individual subjects confirmed that this plateauing was a significant effect: the 0-, −3-, and −6-dB conditions did not differ significantly at the $p<0.05$ level, while the 0, +3, +6, and +9 conditions were all significantly different from one another. The somewhat paradoxical leveling out of performance for SNRs less than 0 dB was found in early experiments on speech intelligibility in the presence of a single competing talker (Egan *et al.*, 1954; Dirks and Bower, 1969) and, as discussed in Brungart (2001), is likely due to a difference in level providing a cue to sound source identity, which compensates for increased energetic masking.

The new result in this figure is the gradual improvement in performance that occurred when target and masker sentences were artificially separated in *F*0 ($F_{1,8}=45.4$, $p<0.0001$, averaged over the lowest four SNRs). This effect is highlighted in the right panel of Fig. 1, which shows the probability of a correct color and number response averaged across the four lowest SNR values (−9 to +3 dB) as a function of the absolute *F*0 separation between the two talkers. (The +6- and +9-dB values were excluded from the average since performance was asymptoting, leaving little room for *F*0 differences to improve performance.) Although the improvement in performance from 0- to 1-semitone separation was only marginally significant ($F_{1,8}=5.9$, $p<0.05$, averaged over the lowest four SNRs), increasing the separation to 2 semitones improved performance by 12 percentage points, and increasing the separation to 12 semitones produced a 24 percentage point improvement in overall performance.

Note that, as the *F*0 separation was increased, the plateau in performance that occurred in the 0-semitone condition at SNRs less than 0 dB gradually disappeared. Statistical support for this observation is provided by the fact that increasing the SNR from −6 to 0 dB produced a significant improvement in performance at *F*0 separations of 3, 4, 6, 9, and 12 semitones (*post hoc* Fisher LSD test on the arcsine-transformed individual subject data, $p<0.05$), but not at *F*0 separations of 0, 1, or 2 semitones. At separations of 9 or 12 semitones, performance declined monotonically from over 90% at +9 dB SNR to around 65% at −6 dB SNR (left panel). A similar result was seen in the earlier Brungart (2001) study, where performance plateaued at negative SNRs in the same-talker (TT) and same-sex (TS) masking conditions, but decreased gradually in the different-sex (TD) masking condition. Thus, the overall shape of the 9- and 12-semitone curves is similar to the TD condition of the earlier study. However, despite the fact that the mean *F*0 values of the male and female talkers in the CRM corpus differ by almost exactly 12 semitones (104 Hz versus 206 Hz), the maximum improvement in performance afforded by a 12-semitone shift in *F*0 (28% at 0 dB SNR) in this experiment was only half as large as the improvement from the TT to the TD condition in the earlier experiment. Thus, it is clear that differences in *F*0 alone cannot account for the difference in performance between the same-talker and different-sex masking conditions of the earlier two-talker study.

The data were also examined to determine what effect the relative values of *F*0 for the target and masker had on performance. There was no consistent difference between conditions with a higher-pitched target and those with a higher-pitched masker. In addition, there was no interaction between relative *F*0 and the sex of the talker.

Nearly all of the incorrect responses in the experiment consisted of the color and/or number that occurred in the masker sentence. Only 1.8% of the incorrect responses were not such intrusions. This result indicates that the performance benefits that occurred with separation in *F*0 were due to improvements in the listeners' ability to allocate the colors and numbers in the stimulus to the appropriate call-signs, and not due to an increase in their ability to correctly perceive the colors and numbers in the stimulus.

2. Differences between talkers

The overall results conceal considerable variation between individual talkers. Figure 2 shows the probability of a correct identification separately for each talker. The data

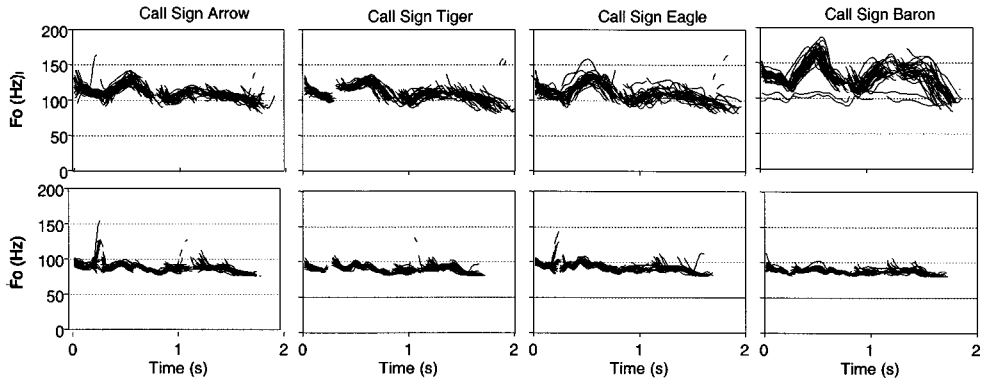


FIG. 3. Plots of the *F*0 contours for two male talkers. The upper row is talker 5, the lower row talker 8. Each plot overlays the *F*0 contours from all 32 sentences that used a particular call sign. The target call sign (''Baron'') is no. 7. The extracted *F*0 contours have been smoothed by low-pass filtering at 10 Hz.

from female talkers are shown in the upper row, with males in the lower row. As an extreme example of the differences between talkers, compare talker 5 (bottom left) with talker 8 (bottom right). For talker 5, an *F*0 separation of 12 semitones between the target and masker phrases did not improve performance over the condition in which there was no *F*0 separation; performance to his utterances is already very high with 0-semitone separation. By contrast talker 8's performance is much lower with 0-semitone separation than is talker 5's, and improves dramatically when the *F*0 separation is increased to 12 semitones.

These striking differences between talkers are at least partly due to their different intonation patterns. The intonation patterns for talkers 5 and 8 are shown in Fig. 3. Talker 8 (bottom row) spoke all his sentences with a rather flat intonation, which varied very little across the different call signs. Consequently, listeners were not able to use either instantaneous differences in *F*0 or differences in the overall contour to help them follow the target rather than the masker sentence. Introducing an artificial overall shift in *F*0 between the target and masker sentences, however, would have allowed listeners to use *F*0 differences in this way. Talker 5, on the other hand, used large excursions of *F*0 in his intonation, which, although similar for sentences with the same call-sign, differed substantially between call-signs: in particular, his intonation for the target call-sign ''Baron'' is distinctively different from the others. Even without any artificial shift in *F*0, this talker's speech provided substantial *F*0 differences between the target and masker sentences to help a listener in tracking the target sentence.

We now provide a quantitative test of whether the improvement with *F*0 separation is greater for talkers who show less initial difference in the original *F*0 between the target and masker sentences. Figure 4 shows a strong inverse correlation ($r^2=0.93$, df$=7$, $p<0.001$) between the improvement between 0 and 12 semitones averaged across SNR for individual talkers, and the average rms instantaneous difference in *F*0 between target and masker sentences for that talker. We estimated this rms difference with Praat by taking 50 random target-masker pairs for each talker, extracting their *F*0 contours (with 0.01-s intervals between points), smoothing these contours with a low-pass 10-Hz filter and then taking the rms difference in Hz for *F*0 at each time slice throughout the sentence pair. Time slices for which either sentence was silent or voiceless did not contribute to the rms average. These averages were themselves averaged across the 50 random sentence pairs to give the data for each talker plotted in Fig. 4.

C. Discussion

1. Effect of F0 separation

The main effect of *F*0 separation is that performance gradually improves as *F*0 separation increases, with little improvement at 1-semitone separation, and maximum improvement at 12 semitones. It is important to bear in mind that these nominal semitone separations in *F*0 have been added to the existing natural variations between utterances of the same talker. This natural variation is of the order of 5% to 10% for seven of the eight talkers and is probably responsible for the fact that overall improvement in the 1-semitone condition was barely detectable. The remaining talker (talker 5) had an unusually large variation in *F*0 between his target and masker sentences, and performance remained essentially unchanged as the separation in *F*0 varied from 0 to 12 semitones.

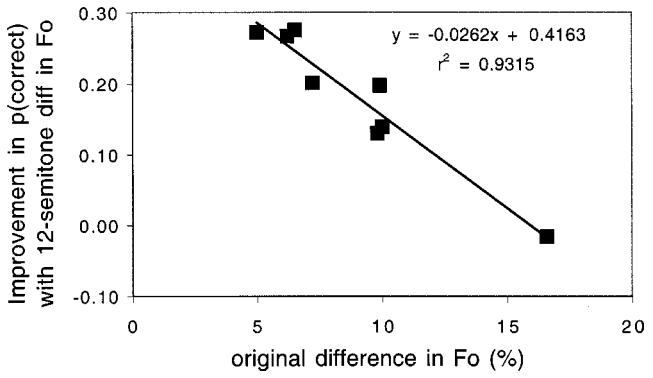


FIG. 4. The abscissa shows the rms % difference in *F*0 between the target voice and the masking voice for a particular talker condition with no alteration to *F*0. The ordinate shows the average change in performance for that talker across all six S/N ratios between an *F*0 shift of zero and 12 semitones. The figure shows a strong inverse relationship between the initial *F*0 difference between the target voice and the masker voice for a particular talker and the improvement in performance obtained by imposing a 12-semitone difference in *F*0 between the two voices.

The overall improvement in performance with *F*0 separation of the target and masking sentences that we have found in this experiment is compatible with previous experiments that have shown that a difference in *F*0 can help listeners segregate a target sentence from competing speech.

Brokx and Nooteboom (1982) asked their listeners (in their first experiment) to recall semantically anomalous sentences of fixed syntactic form (''*The town swims now in a sheep*'') that had been resynthesized (through a formant synthesizer after LPC-based formant tracking) on a monotone *F*0 contour. The sentences were played against a continuous background of speech—a short story read by the same talker—that had also been reduced to a monotone. The signal-to-noise ratio between the sentences and the background speech was either 0, −5, −10, or −15 dB. Content word identification (allowing a single phoneme alteration) of the sentences was measured as a function of the *F*0 difference between the sentences and the background speech. Brokx and Nooteboom found that word identification improved roughly linearly from about 40% to about 60% as the *F*0 difference increased from 0 to 3 semitones, but then identification dropped to about 50% when the separation was 12 semitones.

Two features of our data are different from these results of Brokx and Nooteboom. First, performance in our experiment barely increases overall between the 0- and 1-semitone conditions. Second, in our data the 12-semitone condition gives the best performance; in Brokx and Nooteboom's it did not. Both of these differences have similar explanations, which depend on the natural pitch contours used in our experiment.

In the Brokx and Nooteboom experiment, the *F*0 contours corresponding to a 0-semitone shift were identical, whereas in our experiment they had the variation inherent in natural utterances. As can be seen from the abscissa values in Fig. 4 the target and masker sentences had an average instantaneous difference of between about 5% and 17%, respectively (about 1 and 3 semitones), depending on the talker. These instantaneous differences in *F*0 prevent the fusion, and corresponding decrease in intelligibility (Scheffers, 1983; Assmann and Summerfield, 1989), that occurs with identical *F*0 values. They also provide some differential *F*0-contour information to help a listener to track a particular sound source over time. A small difference in *F*0 will thus have more of an effect on performance with Brokx and Nooteboom's monotone sentences than with ours. The reduction in performance with a one-octave (12-semitone) *F*0 separation in the Brokx and Nooteboom experiment has a similar explanation. The exact octave relationship between the two monotone utterances again causes fusion of the two sounds into a less-intelligible whole. In our experiments the 12-semitone separation was imposed on top of the natural variation in *F*0 and so there was no fusion; moreover, the large *F*0 difference between the sentences probably allowed listeners to track a particular sound source effectively.

Brokx and Nooteboom's (1982) second experiment used stimuli whose *F*0 contours were more similar to ours. Although the content of the utterances was the same as in their first experiment, the speech remained natural rather than being resynthesized by LPC coding, and different *F*0 conditions were obtained by the same talker naturally producing four different types of intonation: (1) normal intonation in the same *F*0 range as was used for the background passage (averaging about 110 Hz); (2) normal intonation at a high pitch (averaging about 160 Hz); (3) intended monotone at about 110 Hz; and (4) intended monotone at about 220 Hz. These four intonation conditions were presented at six different SNRs from 0 to −15 dB. With normal (rather than monotonous) intonation, performance was considerably better when the *F*0 of the target sentences was in a higher range than the background passage (72%) compared with when it was in the same range (55%). However, when the speech was spoken on a monotone, performance improved very little on the higher *F*0.

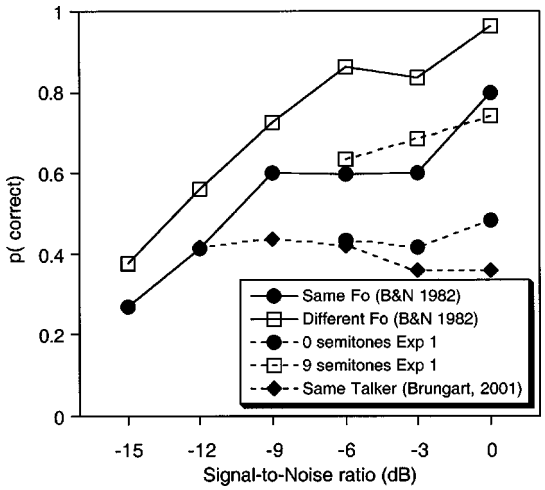


FIG. 5. Comparison of the data from experiment 1 with data from earlier experiments. Data joined by solid lines are from experiment 2 of Brokx and Nooteboom (1982) and show the probability of correct recall of content words from naturally spoken semantically anomalous sentences against a background of speech from the same talker in the same (solid circles) or different (open squares) pitch range. Data joined by dashed lines come from the same talker condition (solid diamonds) of Brungart (2001) and from the 0- (solid circles) or 9-semitone (open squares) conditions of experiment 1.

The *F*0 separation of the first two conditions of Brokx and Nooteboom's second experiment are similar to our 0- and 9-semitone manipulations. Figure 5 compares the results from these conditions at various SNRs. The figure also shows data from the same-talker condition of Brungart (2001) which was very similar to the present 0-semitone condition, but, like the Brokx and Nooteboom experiments, used a wide range of SNRs. Although overall performance in our data is lower (perhaps because of the rhythmic and semantic similarity between our target and mask utterances), the improvement in correct identifications with a difference in *F*0 is similar (c. 30%) in both experiments.

Larger changes in intelligibility of speech with differences in *F*0 have been reported by Bird and Darwin (1998). Their listeners had to recall the shorter of two simultaneous sentences that had been constructed to contain few stop and fricative consonants (e.g., ''*I only moan in the morning*''). The sentences were spoken on a monotone and resynthesized using LPC or PSOLA to have different monotonous *F*0 values. Intelligibility with the lower-quality LPC resynthesis in-

creased from about 20% correct words to 80% as the *F*0 difference increased from 0 to 8 semitones. With the higher-quality PSOLA resynthesis (and a new talker) the change was a little less (from 35% to 75%) from 0 to 10 semitones. The rather larger effect of a difference in *F*0 found in this experiment compared with our present experiment is probably due to fusion of the strictly monotone sentences at 0-semitone separation, together with the greater reliance on *F*0 to track an individual sentence in the absence of the onset and offset cues provided by stop and fricative consonants.

III. EXPERIMENT 2: CHANGES IN VOCAL TRACT LENGTH ONLY

The results of experiment 1 clearly show that artificial separations in the *F*0 values of the talkers can produce substantial performance improvements in a multitalker listening task. However, even a 12-semitone change in *F*0 did not produce as large an improvement in performance as a change in the sex of the masking talker. Thus, it is clear that differences in *F*0 alone cannot account for a listener's ability to segregate different-sex talkers. A second experiment was conducted to examine the effect of another perceptual property that listeners may be able to use to segregate different-sex talkers: the length of the vocal tract.

A. Method

The primary difference between experiments 2 and 1 is that the *F*0 shifts that were examined in the first experiment were replaced by changes in the vocal-tract lengths of the target and masking talkers. As was the case with the *F*0 shifts, these vocal-tract shifts were implemented by processing each of the 1024 sentences in the CRM corpus with the Praat software package. The apparent vocal-tract length of each talker was changed by a factor of vt for each utterance by (1) multiplying *F*0 by vt and duration by $1/vt$ (using PSOLA), (2) resampling at the original sampling frequency multiplied by vt, and then (3) playing the samples at the original sampling frequency. The end effect of this manipulation was to maintain the same duration and *F*0 of the original utterance but to scale the spectral envelope by vt. Scaling the spectral envelope is not identical to a change in vocal-tract length since it scales all those factors that are responsible for the spectral envelope. These factors include for example the spectral envelope characteristics of the voice source (such as spectral tilt) as well as the vocal-tract transfer function. However, the vocal-tract resonances are the main factor responsible for spectral envelope shape, and for simplicity we will refer to the manipulation as a change in vocal-tract length.

The range of vocal-tract ratios that we used is based on the average formant-frequency ratio between female and male voices reported by Peterson and Barney (1952) (of around 16%). Each utterance was processed with the following values of vt: 1.16, 1.08, 1.04, 1.02, 1.0, 0.98, 0.96, 0.92, 0.84. Pairs of utterances from the same original talker were then selected on each trial to produce one of the nine relative values of vt shown in Table II. These relative values were used in an experiment that was otherwise identical to experiment 1. Lengthening the vocal tract leads to a more male-sounding voice (though without the usual lower pitch), shortening it to a more female-sounding voice.

TABLE II. Vocal-tract length differences tested in the stimuli of experiment 2. Each pair of sentences in a trial was chosen from the same talker. The scaling values were selected differently for male and female talkers to place most of the scaled vocal tracts into the range between typical male and female talkers. Thus, male talkers were generally scaled down more than up, and female talkers were generally scaled up more than down. Note that, on a given trial, the target talker was equally likely to receive the higher or lower vt scaling.

	Male		Female	
vt Ratio	Longer VT	Shorter VT	Longer VT	Shorter VT
1.0	1.0	1.0	1.0	1.0
1.02	1.0	0.98	1.02	1.0
1.04	1.02	0.98	1.02	0.98
1.08	1.04	0.96	1.04	0.96
1.13	1.04	0.92	1.08	0.96
1.16B	1.08	0.92	1.08	0.92
1.16A	1.0	0.84	1.16	1.0
1.38	1.16	0.84	1.16	0.84

Eight of the nine paid volunteer listeners who were used in the first experiment had also participated in experiment 2. Four of the listeners were males, and four were females. Each completed ten trials for each of the 384 possible stimulus configurations of the experiment (8 talkers×8 vt differences×6 SNRs), for a total of 3840 trials. These trials were completed in blocks of 192 trials, with the trials within each block randomly balanced to have an equal number of trials for each value of each of the three independent variables (talker, vt difference, and SNR). Each block took approximately 10 min to complete, and each listener participated in two to three blocks per day over a 2-week period.

B. Results

The left panel of Fig. 6 shows the probability of a correct identification of both the color and number of the target sentence as a function of the signal-to-noise ratio and the ratio of vocal-tract length changes. The baseline condition with a vt ratio of 1.0 (open circles in Fig. 6) produced results very similar to those obtained in previous same-talker con-

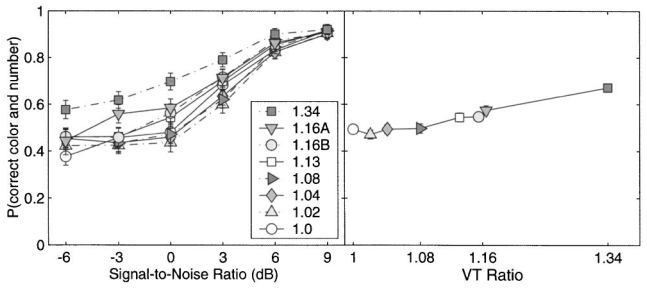


FIG. 6. The left panel shows the probability of a correct identification in experiment 2 of the number and color in the target sentence as a function of SNR for eight different ratios of vocal-tract length between the two sentences. The ratio is always taken so as to be unity or greater. The right panel shows performance averaged across the lowest four SNR values (−6, −3, 0, and 3 dB) as a function of the vt ratio. The data have been averaged over all eight talkers. The error bars represent 95% confidence intervals.

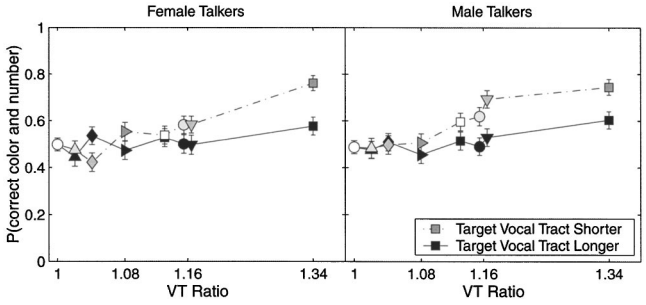


FIG. 7. Data as in the right panel of Fig. 6, except the results have been shown separately for female talkers (left panel) and male talkers (right panel), with separate curves representing trials where the target voice had a shorter vocal tract (open and shaded symbols) or a longer vocal tract (black symbols) than the masking voice.

figurations, with a decrease in performance from +9 to 0 dB SNR and then a plateau in performance at SNR values less than 0 dB.

The right panel of Fig. 6 shows the percentage of correct responses as a function of vt ratio averaged across the lowest four SNR values tested in the experiment (SNR$\leqslant$+3 dB). At 0-dB SNR, detectable improvement only appeared for vt ratios of 1.13 ($p<0.0001$) or greater. Changes of vt ratio of 1.08 or less gave no improvement in performance ($F_{3,21}=1.94$, $p>0.1$). For the largest vt ratio (1.38), however, there was a substantial improvement in performance compared to the baseline condition with equal-length vocal tracts: the percentage of correct responses increased from about 50% correct to 70% correct. This improvement is comparable to that produced by an *F*0 shift of 9 semitones in experiment 1 both in the magnitude of the improvement at 0 dB SNR and in the overall shape of the curve (filled squares in left panel of Fig. 6).

The results so far have not distinguished between trials where the target sentence had the shorter or the longer vocal tract. This distinction is made in Fig. 7, which is similar to the right panel of Fig. 6, but breaks the results down by the relative length of the target vocal tract (shaded and black symbols in each panel) and by the sex of the talker (females in the left panel, males in the right panel). The main result shown in the figure is that performance was consistently higher in trials where the target talker had a shorter vocal tract. Indeed, it appears that manipulating the vt ratio had very little effect on the proportion of correct responses when the masking talker had the shorter vocal tract. It is interesting to note that this effect was equally strong for both male and female talkers. Thus, the performance advantage seen for the shorter vocal tract does not appear to depend on the initial vocal-tract length of the talker. This result suggests that listeners may be biased to focus their attention on the talker with the shorter vocal-tract length when no other cues are available to segregate the two talkers.

As was the case in the first experiment, talker 5 produced the best performance in the baseline condition and showed the least improvement with increased vt ratio. At 0 dB SNR, there was no improvement in performance between the 1.0 and 1.34 vt ratios with talker 5, compared to an average improvement of 30 percentage points for the other seven talkers in the experiment. The large natural variations in *F*0 that occurred in the speech from talker 5 probably provided sufficient segregation such that no additional benefit was gained from changing the vocal-tract length of that talker.

As in experiment 1, the vast majority of incorrect responses contained the color and/or number from the masker sentence. Only 2.7% of the errors were not such intrusions. The improvements that we see with vt separation are thus improvements in the listeners' ability to allocate correctly perceived colors and numbers to the appropriate call-sign sentence.

C. Discussion

This experiment asked whether changes in vocal-tract length between the talkers of the target and masker sentences could improve performance. The main result of this experiment is that although small changes (8% or less) do not improve performance, larger changes do. Across the four lowest SNRs, a change of 13% in vt length increases performance from about 50% to 55% ($p=0.001$) and a 38% change increased it further to 67% ($p<0.0001$).

This result is entirely consistent with previous experiments by Darwin and Hukin (2000). They found that small differences in vocal-tract length ($\leqslant$8%) did not help listeners to allocate a target word to the target carrier sentence, but a difference of $\pm$17% helped substantially. It may be the case that small changes ($\pm$4%) in vocal-tract length are not effective in segregation because listeners do not hear them as a change in the identity of the talker. Kuwabara and Takagi (1991), for example, found that an upward or downward shift of 8% in the first three formant frequencies was sufficient to reduce individual voice recognition to chance. The smaller, bi-directional changes that we used may not be sufficient to change the identity of talker's voices that were familiar to our listeners.

D. Interim discussion

These two experiments have identified two factors which are likely to have contributed to the improvement in performance found by Brungart (2001) for different sex talkers relative to identical talkers. A difference in talker sex produces a difference in overall *F*0 of the target and masker sentences of a little under an octave and a change in vocal-tract length of about 16% (Peterson and Barney, 1952). Artificially manipulating the *F*0 and the vocal-tract length separately also gave substantial increases in performance. However, it is unlikely that these two factors alone are entirely responsible for the performance differences originally found. If we look simply at the data gathered with a SNR of 0 dB, and assume that the effects of *F*0 and of vocal-tract length are additive, then we find that together they provide less improvement than was found with different-sex talkers. Specifically, we converted the probability of a correct response to d' values using published tables (Hacker and Ratcliff, 1979), assuming a choice between four orthogonal alternatives (since almost all errors were the color or number from the masker sentence). Changing the sex of the talker

TABLE III. *F*0 ratio (and equivalent semitone) and vt values used to modify male voices in experiment 3.

Male original	*F*0 ratio	Semi tones	vt
Super-male	0.74	−5.2	1.08
Male	1.00	0.0	1.00
Quarter-female	1.17	2.7	0.96
Half-female	1.35	5.2	0.92
Almost female	1.53	7.4	0.88
Female	1.70	9.2	0.84
Super-female	2.05	12.4	0.76

TABLE IV. *F*0 ratio (and equivalent semitone) and vt values used to modify female voices in experiment 3.

Female original	*F*0 ratio	Semi tones	vt
Super-male	0.49	−12.4	1.24
Male	0.59	−9.2	1.16
Almost male	0.66	−7.2	1.12
Half-male	0.74	−5.2	1.08
Quarter-male	0.85	−2.7	1.04
Female	1.00	0.0	1.00
Super-female	1.35	5.2	0.92

increases d' by about 1.91; a change of 12 semitones in *F*0 increases d' by 0.97, and a change in vocal-tract length of 1.16 increases d' by 0.46. We can combine these two d' measures in one of two ways in order to obtain an expected d' when both cues are present. Any d' scores can be linearly added if the underlying noise distribution is the same for both measures, or orthogonally added (as the square root of the sums of their squares) if the underlying noise distributions are independent. With each of these ways of combination there is a shortfall attributable to other voice characteristics, which is between about 0.48 (assuming linear additivity) and 0.84 (assuming orthogonality). These voice characteristics may also be at least partly responsible for the 0.67 improvement in d' that is found in Brungart's data (2001) for different talkers of the same sex (TS) compared with identical talkers (TT). One of the ways in which talkers differ is in the timing of their speech, but there may also be a super-additive effect of *F*0 and vocal-tract length differences, when they are combined in a natural way. The next experiment examines improvement on the task when both *F*0 and vocal-tract length co-vary in a natural way.

IV. EXPERIMENT 3: CHANGE IN BOTH *F*0 AND VOCAL TRACT LENGTH

The first two experiments examined the effects that isolated changes in *F*0 and vocal-tract length have on a listener's ability to segregate two talkers. However, neither of these manipulations alone was found to produce the level of performance that occurs when the target and masking sentences are spoken by different-sex voices that differ both in *F*0 and in vocal-tract length. Thus, a third experiment was conducted that varied both *F*0 and vt in concert to simulate a smooth transition between a same-sex and different-sex masking voice.

A. Methods

Most of the procedures for the third experiment were the same as for the first two, but the stimuli consisted of speech phrases in which both the vt ratio and the *F*0 frequency were manipulated at the same time. The particular vt and *F*0 values that we used are shown in Tables III and IV. Another difference with the first two experiments is that here on each trial listeners heard one message with the *F*0 and vt unchanged; these unchanged values are referred to as "male" in Table III and "female" in Table IV. From the original male voices we generated "female" voices with *F*0 multiplied by 1.7 and vt by 0.84. These values correspond to the average female/male ratios for the formant and *F*0 data reported by Peterson and Barney (1952). Similar, inverse changes were made to the original female voices to give new "male" voices. Three intermediate voices were then linearly interpolated between each original voice and its changed counterpart (which we have arbitrarily labeled as "quarter-," "half-," and "almost" shifts to the opposite gender in the table), and two more extreme values were linearly extrapolated (the super males and super females of Tables III and IV).

The experimental procedure was similar to the first two experiments. Eight paid volunteer listeners (four male, four female) participated in experiment 3. All but two had previously participated in the first two experiments. Each listener completed ten trials for each of the 336 possible stimulus configurations of the experiment (8 talkers×7 gender differences×6 SNRs), for a total of 3360 trials. These trials were completed in blocks of 168 trials, with each block randomly balanced to have an equal number of trials for each value of each of the three independent variables (talker, gender difference, and SNR). Each block took approximately 9 min to complete and each listener participated in two to three blocks per day over a 2-week period.

B. Results and discussion

The results of the experiment are shown separately for the male and female talkers in the left panels of Figs. 8 and 9. Once again, the results are shown in terms of the percentage of correct identifications of the color and number in the target sentences as a function of the SNR of the stimulus. The results for the unshifted masking voices (open circles in the figures) were once again similar to those from previous experiments, with a rapid decrease in performance from +9 to 0 dB SNR, and a plateau in performance at about 45% correct responses at SNRs less than 0 dB. However, the results for the gender-shifted sentences show a systematic increase in performance up to a substantially higher level of performance than that obtained in either of the first two experiments. At a SNR of 0 dB, overall performance improved approximately 35 percentage points (from 45% to 80%) at the larger gender shifts tested for both the male and female talkers. This brought performance up to a level comparable to that measured with different-sex talkers in the earlier experiment by Brungart (2001).

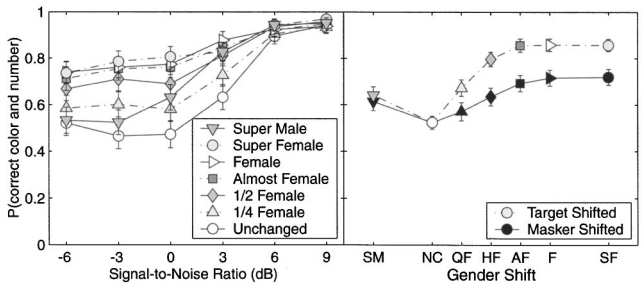


FIG. 8. The left panel shows the probability of a correct identification of the number and color in the target sentence as a function of SNR for the seven different gender shifts introduced into the male voice in experiment 3 (see Table III). The right panel shows performance averaged across the lowest four SNRs (−6, −3, 0, and 3 dB) as a function of the gender shift, with separate curves for trials where the target voice was shifted (open and shaded symbols) and trials where the masking voice was shifted (black symbols). The error bars represent the 95% confidence intervals of the mean.

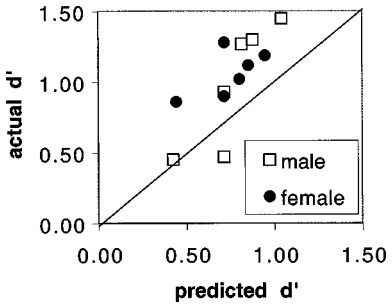


FIG. 10. The actual d' improvement in experiment 3 at 0 dB SNR when one voice changes in both $F0$ and vocal-tract length versus predicted d' from separate changes in either variable in experiments 1 and 2. Published tables (Hacker and Ratcliff, 1979) were used to convert the d' values from the probability of a correct color and number identification in each condition with the assumption that each response represented a choice between four orthogonal alternatives. See text for details.

The right panels of Figs. 8 and 9 show performance averaged across the four lowest SNRs tested in the experiment (−6 to +3 dB) as a function of the gender shift introduced between the two talkers in the stimulus. These data are plotted separately for trials where the target speech was shifted in gender (shaded and open symbols) and where the masker was shifted in gender (black symbols). The data for both the male and female talkers show that performance was generally better when the target was shifted in gender than when the masker was shifted in gender. In part, this may occur because the listener's attention is drawn to the unusual characteristics of the gender-shifted talker. The difference between the target-shifted and masker-shifted configurations was largest when the male target talker was female-shifted, which may reflect the bias in favor of the talker with the shorter vocal-tract length found in experiment 2. The difference between the target-shifted and masker-shifted configurations was substantially smaller for the male-shifted female talkers, which may reflect a conflict between the novelty of the gender-shifted talker and the bias in favor of the talker with the shorter vocal tract. Overall, the results suggest that the performance advantages that occurred with the shorter vocal tract in experiment 2 represent a relatively weak effect that only dominates the results when the vocal-tract length is changed at a fixed value of $F0$.

The results of experiment 3 also reveal other asymmetries between the effects of the gender shifts for the male and female talkers. When the voice characteristics of the shifted voice fell between the normal ranges of male and female voices, there was generally a more rapid improvement in performance with gender change for the female talkers than for the male talkers. This was especially true for the female-$\frac{1}{4}$ male condition, which produced performance about 10 percentage points higher than the corresponding male-$\frac{1}{4}$ female condition (upward-pointing triangles in the right panels of the figures). In both cases, however, nearly 100% of the benefit of separation in gender was achieved in the almost male or almost female condition, which represent the normally occurring differences in vt and $F0$ for a male and female voice (squares in the right panels of the figures). A larger asymmetry between the male and female talkers occurred when the voice characteristics were shifted outside the normally occurring range of vt and $F0$. Specifically, the performance improvement produced by shifting a female voice to a super-female voice (upside-down triangles in the right panel of Fig. 9) was nearly three times as large as the performance improvement produced by shifting a male voice to a super-male voice (upside-down triangles in the right panel of Fig. 8). The reason for this differential improvement is seen more clearly in the left panel of Fig. 8, which shows that performance in the super-male condition falls off substantially more rapidly than performance in any condition except the unshifted condition as the SNR of the stimulus decreased. It appears that the listeners had extreme difficulty segregating a male voice from a super-male voice at low SNRs. At this point we have no explanation for this result.

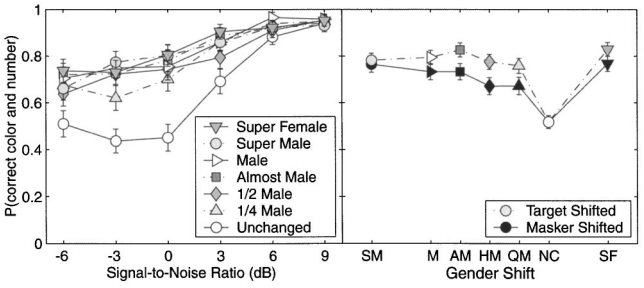


FIG. 9. Identical to Fig. 8 except the data are shown for the female talkers in experiment 3 with the gender shifts outlined in Table IV.

To address the question of whether the effects of a shift in $F0$ and a shift in vocal-tract length are additive, we predicted the improvements in performance in the 0 dB SNR conditions of experiment 3, which varied both $F0$ and vocal-tract length, from the improvements shown for these two parameters separately in experiments 1 and 2. We assumed that the two factors were orthogonal and so the square of the predicted d' was the sum of the squares of the two predicting d's. The results of this analysis in Fig. 10 show that the factors behave super-additively—predicted values for performance when both variables change are, except for 1 point out of 12, lower than the actual values obtained in experiment 3. Listeners therefore gain more benefit from a joint change in

*F*0 and vocal-tract length than is predicted from changes to each of these variables separately.

This super-additive combination of *F*0 and *vt* cues can explain most, but not all, of the improvement in performance that occurs when the target and masking speech signals are spoken by different-sex talkers than when they are spoken by the same-sex talkers. Brungart (2001) found that listeners responded correctly in the CRM task approximately 85% of the time when different-sex competing talkers were presented at a SNR of 0 dB. However, even in the male–super-female and female–super-male configurations of experiment 3, where the *F*0 shift (105%) was slightly larger than the average *F*0 difference between the male and female talkers in the corpus (100%) and the *vt* shift (24%) was twice as large as the average *vt* shift between the male and female talkers in the corpus (12%), the listeners responded correctly only approximately 80% of the time when the SNR was 0 dB. The 5 percentage point performance difference between the best condition in experiment 3 and the ''TD'' condition in the earlier experiment is presumably the result of other variations in the voices of the different talkers in the corpus, such as voice-source characteristics, intonation, and speaking rate.

V. CONCLUSIONS

These three experiments have examined the roles that differences in *F*0 and differences in vocal-tract length have on the segregation of multiple simultaneous speech signals. The imposed *F*0 shifts were additional to the natural differences already present between separate utterances by the same talker. The major results can be summarized as follows:

(1) Differences in *F*0 produce systematic improvements in segregation performance when the difference is 2 semitones or greater, but *F*0 alone cannot improve performance to the level that occurs with different-sex talkers.
(2) Differences in vocal-tract length produce systematic improvements in segregation performance when the ratio of lengths is 1.08 or greater, but differences in vocal-tract length alone cannot improve performance to the level that occurs with different-sex talkers. In general, performance is better when the shorter vocal tract belongs to the target talker.
(3) Systematic changes in both *F*0 and vocal tract that simulate an incremental shift in gender produce substantially larger improvements in performance than differences in *F*0 or *vt* alone. The combined effect is larger than that predicted assuming additivity. In general, shifting one of two utterances from a female voice towards a male voice produces a greater improvement in performance than shifting from male towards female.
(4) The intonation patterns of the individual talkers can play as large a role as *F*0 or *vt* in determining overall segregation performance. The natural variations in intonation in the utterances spoken by talker 5 were so large that no additional improvement was obtained by introducing artificial changes in the *F*0s or vocal-tract lengths of the competing phrases.

ACKNOWLEDGMENTS

The first author was supported by Grant No. GR/M90146 from the UK EPSRC and Grant No. G9801285 from the UK MRC. The second author was supported in part by AFOSR Grant No. LRIR01HE01COR. Peter Assmann and two anonymous reviewers provided helpful comments on the paper.

Assmann, P. F., and Summerfield, A. Q. (**1989**). ''Modelling the perception of concurrent vowels: Vowels with the same fundamental frequency,'' J. Acoust. Soc. Am. **85**, 327–338.

Assmann, P. F., and Summerfield, A. Q. (**1990**). ''Modelling the perception of concurrent vowels: Vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **88**, 680–697.

Atal, B. S., and Hanauer, S. L. (**1971**). ''Speech analysis and synthesis by linear prediction of the acoustic wave,'' J. Acoust. Soc. Am. **50**, 637–655.

Bird, J., and Darwin, C. J. (**1998**). ''Effects of a difference in fundamental frequency in separating two sentences,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 263–269.

Boersma, P., and Weenink, D. (**1996**). ''Praat, a System for doing Phonetics by Computer, version 3.4,'' Institute of Phonetic Sciences, University of Amsterdam, Vol. 132, pp. 1–182, www.praat.org

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). ''A speech corpus for multitalker communications research,'' J. Acoust. Soc. Am. **107**, 1065–1066.

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

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). ''Intonation and the perceptual separation of simultaneous voices,'' J. Phonetics **10**, 23–36.

Bronkhorst, A. W. (**2000**). ''The cocktail party phenomenon: a review of speech intelligibility in multiple-talker conditions,'' Acustica **86**, 117–128.

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

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

Darwin, C. J., and Hukin, R. W. (**2000**). ''Effectiveness of spatial cues, prosody and talker characteristics in selective attention,'' J. Acoust. Soc. Am. **107**, 970–977.

Dirks, D., and Bower, D. (**1969**). ''Masking effects of speech competing messages,'' J. Speech Hear. Res. **12**, 229–245.

Egan, J. P., Carterette, E. C., and Thwing, E. J. (**1954**). ''Some factors affecting multi-channel listening,'' J. Acoust. Soc. Am. **26**, 774–782.

Ericson, M. A., and McKinley, R. L. (**1997**). ''The intelligibility of multiple talkers separated spatially in noise,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum, Mahwah, NJ), pp. 701–724.

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.

Hacker, M. J., and Ratcliff, R. (**1979**). ''A revised table of *d′* for *M*-alternative forced choice,'' Percept. Psychophys. **26**, 168–170.

Kuwabara, H., and Takagi, T. (**1991**). ''Acoustic parameters of voice individuality and voice-quality control by analysis-synthesis method,'' Speech Commun. **10**, 491–495.

Moulines, E., and Charpentier, F. (**1990**). ''Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones,'' Speech Commun. **9**, 453–467.

Peterson, G. H., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.

Scheffers, M. T. (**1983**). ''Sifting vowels: Auditory pitch analysis and sound segregation,'' Ph.D. dissertation, Groningen University, The Netherlands.

Discrimination and identification of vowels by young, hearing-impaired adults[a)]

Carolyn Richie,[b)] Diane Kewley-Port, and Maureen Coughlin
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 27 November 2002; accepted for publication 4 August 2003)

This study examined the effects of mild-to-moderate sensorineural hearing loss on vowel perception abilities of young, hearing-impaired (YHI) adults. Stimuli were presented at a low conversational level with a flat frequency response (approximately 60 dB SPL), and in two gain conditions: (a) high level gain with a flat frequency response (95 dB SPL), and (b) frequency-specific gain shaped according to each listener's hearing loss (designed to simulate the frequency response provided by a linear hearing aid to an input signal of 60 dB SPL). Listeners discriminated changes in the vowels /ɪ e ɛ ʌ æ/ when F1 or F2 varied, and later categorized the vowels. YHI listeners performed better in the two gain conditions than in the conversational level condition. Performances in the two gain conditions were similar, suggesting that upward spread of masking was not seen at these signal levels for these tasks. Results were compared with those from a group of elderly, hearing-impaired (EHI) listeners, reported in Coughlin, Kewley-Port, and Humes [J. Acoust. Soc. Am. **104**, 3597–3607 (1998)]. Comparisons revealed no significant differences between the EHI and YHI groups, suggesting that hearing impairment, not age, is the primary contributor to decreased vowel perception in these listeners. [DOI: 10.1121/1.1612490]

PACS numbers: 43.71.Ky, 43.71.Es, 43.71.Lz [KWG] Pages: 2923–2933

I. INTRODUCTION

It is often recognized that on tests of speech perception the performance of hearing-impaired listeners is poorer than that of normal-hearing listeners. Because hearing loss is associated with aging and many studies include older adults, age is often a confounding factor in studies of speech perception by listeners with sensorineural hearing loss. Information regarding the auditory capabilities of young, hearing-impaired listeners is thus necessary to determine the role that age may play in speech perception. Specifically, comparison of young, hearing-impaired and elderly, hearing-impaired listeners may help distinguish among various auditive and cognitive factors involved in speech perception.

The purpose of this paper is to examine the effects of mild-to-moderate sensorineural hearing loss on two vowel perception tasks, discrimination and identification. While it is commonly recognized that elderly listeners may perform more poorly on auditory tests of speech recognition compared to young listeners, the reasons for differences seen between groups are not entirely clear. Some researchers argue that there are a number of cognitive factors associated with aging that affect test performance (e.g., Era *et al.*, 1986), while others argue that the main reasons for differences between young and elderly listeners are auditive (e.g., Nábělek, 1988). While there is some amount of evidence for both cognitive and auditive factors affecting elderly listeners' performance on auditory tests of speech recognition, converging evidence suggests that auditive factors contribute most to test results, and that age-related cognitive factors may be involved in a secondary manner (e.g., Jerger *et al.*, 1991; van Rooij and Plomp, 1992; Humes *et al.*, 1994). Previous work in vowel discrimination and identification from this laboratory, however, did not examine young hearing-impaired listeners and thus could not separate the age-related cognitive and auditive factors involved in performance (e.g., Coughlin *et al.*, 1998).

A. Peripheral factors affecting performance on speech perception tasks

The ''auditive'' hypothesis argues that general reduction in auditory sensitivity associated with aging is primarily responsible for decreased performance of elderly listeners on tests of speech recognition. Nábělek (1988), for example, tested listeners with varying age and degree of hearing loss on vowel identification in quiet, noise (12-talker babble), and reverberation. The mean four-frequency (500, 1000, 2000, and 4000 Hz) average hearing loss produced the highest correlation with vowel identification scores in all listening conditions, suggesting that hearing impairment is likely the most significant factor related to performance on speech perception tasks.

Sensorineural hearing impairment may distort suprathreshold speech in a number of ways, through abnormal growth of loudness, impaired frequency resolution, and absence of two-tone suppression effects (Van Tasell *et al.*, 1982). Past work has examined the auditory filter characteristics of various listeners as well as the effect of listening with such filters on speech perception. Dubno and Dirks (1989) estimated the auditory filter characteristics for normal-hearing listeners and listeners with sensorineural hearing loss. Auditory filter characteristics of the hearing-impaired listeners were significantly broader than those for the normal-hearing listeners. Peters and Moore (1992) also

[a)]Portions of these data were presented at the 1 March 2001 Meeting of the Central Institute for the Deaf (CID), St. Louis, MO.
[b)]Electronic mail: carodavi@indiana.edu

showed that auditory filter shapes for young and elderly hearing-impaired listeners were substantially broader than those for normal-hearing listeners. While filters tend to broaden with increasing hearing loss, there were no clear differences found between the filter characteristics of YHI and EHI listeners. This suggests that when a loss of frequency selectivity occurs, it is associated with an elevation in absolute threshold. Peters and Moore note, ''Results suggest that presbycusis has similar effects on frequency selectivity to idiopathic sensorineural hearing loss in young subjects, consistent with the idea that both types of hearing loss are primarily cochlear in origin'' (1992, pp. 263–264). Thus, both young and elderly hearing-impaired listeners seem to have broadened auditory filters that may ''distort'' acoustic signals.

B. Cognitive and age factors affecting performance on speech perception tasks

The argument that cognitive factors affect elderly listeners' performance on tests of speech recognition relies on the fact that audiological measures alone cannot account for the relatively poor performance of EHI listeners on tests of speech perception. Some researchers suspect changes in cognitive abilities associated with the normal aging process may be the reason for comparably poor test performance, in addition to hearing loss (e.g., Hargus and Gordon-Salant, 1995). Cognitive factors associated with aging, such as working memory capacity and attentional processes, may contribute to auditory test performance. While pure-tone thresholds can often be correlated with results from speech understanding tests for young listeners, the results from speech understanding tests may be lower than expected for older listeners (Era *et al.*, 1986).

Rakerd *et al.* (1996) present clear evidence for the involvement of cognitive processes in speech perception tasks. They studied the effort required of various listeners, when listening to speech. Participants memorized lists of digits and retained them while listening to either speech or speech-shaped noise. Speech listening was ''abnormally'' effortful for both young and elderly hearing-impaired groups; more digits were forgotten when listening to speech than to speech-shaped noise, and the gap between the noise and speech listening performance was significantly greater for hearing-impaired listeners than for normal-hearing listeners. The authors argue that the ''cognitive cost'' of speech understanding was thus greater for the YHI and EHI groups than for the young, normal-hearing (YNH) groups. However, the ''cognitive cost'' of speech understanding did not seem to be significantly different between the YHI and EHI listeners.

On the other hand, some studies indicate that cognitive factors contribute little or nothing to speech recognition performance. In a study of 200 EHI listeners by Jerger *et al.* (1991), degree of hearing loss showed the strongest relation to speech recognition, while cognitive factors accounted for a negligible amount of variance in any of four speech audiometric scores obtained. Humes *et al.* (1994) also argued that cognitive factors contribute little to the amount of variance accounted for in speech recognition test results. Hearing loss was found to be the single largest factor associated with individual differences in speech recognition performance, while measures of cognitive function accounted for little or no additional variance.

Taken together, these studies present mixed results. Overall, most evidence points to the fact that while cognitive processes are certainly involved in speech perception tests (e.g., Rakerd *et al.*, 1996), cognitive factors contribute little to the amount of variance accounted for in measures of speech recognition by persons with hearing impairment (e.g., Jerger *et al.*, 1991; van Rooij and Plomp, 1992; Humes *et al.*, 1994). A necessary caveat, however, is that some of these studies employed lower-level psychophysical measures, some employed measures of connected speech perception, and some a combination.

C. Studies of vowel perception by hearing-impaired listeners

Many researchers have examined the effects of sensorineural hearing loss on tests of vowel perception. Most studies have shown that vowel identification abilities may remain relatively robust in hearing-impaired listeners (e.g., Owens *et al.*, 1968; Godfrey and Millay, 1980; Dorman *et al.*, 1985). Error analyses typically show that the vowels most often substituted in error are those with first and second formants in close frequency proximity, in terms of F1×F2 space.

Van Tasell *et al.* (1987), for example, tested YNH and YHI listeners on identification of seven synthetic steady-state vowels. Vowel masking patterns (VMPs) of thresholds for sinusoidal probes at vowel masker harmonics were also obtained. VMPs of the three hearing-impaired listeners were characterized by smaller dynamic range, poorer peak resolution, and poorer preservation of vowel formant structure compared to the normal-hearing listener tested. However, these VMP features did not necessarily coincide with inaccurate vowel recognition for hearing-impaired listeners (listeners achieved 68%, 75%, and 93% correct). These authors suggested that a gross estimate of formant frequencies, rather than detailed spectral resolution, suffices for accurate vowel identification for some hearing-impaired listeners.

Turner and Henn (1989) also examined the relation between measures of frequency resolution and vowel recognition for normal-hearing listeners and listeners with moderate sensorineural hearing loss. Input filter patterns at six test frequencies were obtained from the listeners. Using each listener's measured filter characteristics, an ''internal spectrum'' for each test vowel was calculated. The input filter patterns were then used to correlate frequency resolution and vowel recognition. On the vowel identification task, the normal-hearing listeners achieved 95% and 97% correct, while the hearing-impaired listeners achieved 60%–83% correct identification. Correlational analyses demonstrated a relation between impaired frequency resolution and vowel recognition ($r=0.67$, $p<0.01$), suggesting that measures of frequency resolution and the acoustic spectra of vowel stimuli may be useful in predicting vowel recognition performance.

In sum, past work has shown that though broadened auditory filters associated with hearing loss may introduce signal distortion, vowel perception by hearing-impaired listeners may be relatively robust compared to other speech

sounds. That is not to say that vowel perception is necessarily normal for listeners with sensorineural hearing loss, however. Discrepancies do exist between the vowel perception abilities of normal-hearing and hearing-impaired listeners. While vowel identification may seem normal for some vowels in some listening conditions, perception of closely spaced vowels may be adversely affected by hearing loss. Further, measures of other psychophysical abilities of YHI listeners for vowel discrimination and identification are lacking.

D. Previous work

This study represents a continuation of research examining the basic speech perception abilities of normal-hearing and hearing-impaired listeners. It is an extension of the earlier work of Kewley-Port and Watson (1994), which examined formant frequency discrimination for isolated vowels in young normal-hearing listeners, and Coughlin *et al.* (1998), which investigated the relation between vowel discrimination and identification abilities of young normal-hearing, elderly normal-hearing (ENH), and elderly hearing-impaired listeners.

Kewley-Port and Watson (1994) established vowel formant (F1 and F2) frequency discrimination thresholds for well-trained, young normal-hearing listeners under conditions of minimal uncertainty. They showed that thresholds for formant frequency are best described as a piecewise-linear function of frequency, which is about 14 Hz in the F1 frequency region ($<$800 Hz) and increases linearly in the F2 region. This study, together with Kewley-Port (2001), revealed some of the effects that training, psychophysical methods, and level of stimulus uncertainty may have on obtaining reliable estimates of the resolution and discrimination capabilities of the auditory system. This research can serve as a baseline of optimal vowel perception abilities in normal-hearing listeners to determine the detrimental effects of hearing impairment.

As a first step in examining the effects of hearing impairment on vowel perception tasks, Coughlin *et al.* (1998) studied three groups: YNH, ENH, and EHI listeners. Results from vowel F1 and F2 discrimination and identification tasks suggested that hearing impairment and age contributed to decreased vowel perception performance in the EHI group. While these and other studies have confirmed the basic vowel formant frequency discrimination capabilities of normal-hearing listeners, and some of the effects of hearing impairment on vowel perception, few have examined performance of *young hearing-impaired* listeners on such tasks. The addition of a group of YHI listeners will help resolve issues surrounding the impact of age and hearing impairment on vowel perception, which have been confounded in past studies. If significant age-related differences are found between EHI and YHI listeners when audibility is similar, then the cognitive factors involved in vowel discrimination and identification should be examined in greater detail. However, it may be possible that there are only minimal differences between EHI and YHI listeners when tested on vowel perception in conditions of similar audibility.

II. EXPERIMENT 1: DISCRIMINATION OF VOWEL STIMULI UNDER DIFFERENT LISTENING CONDITIONS

A. Purpose

The purpose of this experiment was to examine the effects of mild-to-moderate sloping sensorineural hearing loss on vowel formant frequency discrimination for a group of young, hearing-impaired listeners. A group of well-trained YHI adults was tested in order to establish thresholds for discrimination of vowel qualities represented in the first and second formant frequencies, in three listening conditions referred to here as Soft, Loud, and Gain. These three conditions were designed to simulate everyday listening conditions from low to amplified levels.

The Soft condition simulated an average conversational level. It was intended that under this condition vowels would be at least partially audible (i.e., not completely inaudible) at a typical conversational level. The aim of this Soft condition was to simulate a real-world listening level while at the same time ensuring some audibility so that the task could be performed. The second listening condition, Loud, simulated a more audible level for the speech using a flat gain for a presentation level of 95 dB SPL. The third condition, Gain, was included to address concerns raised in previous research with speech perception by hearing-impaired listeners; specifically, that a high level flat gain might degrade frequency resolution, possibly through the upward spread of masking (cf. Dubno and Dirks, 1989; Florentine *et al.*, 1980; Klein *et al.*, 1990; Summers and Leek, 1997). This Gain listening condition simulated the frequency response provided by a linear gain hearing aid based on a 60-dB signal presentation. Thus, all listeners were tested with vowel stimuli presented at (a) a conversational level (*Soft*, approximately 60 dB SPL); (b) a high level, flat gain (*Loud*, 95 dB SPL); and (c) frequency-specific gain based on the listener's hearing loss (*Gain*).

Results were assessed in order to determine the effects of listening condition on performance. First, performance in the Soft condition was compared with performance in the two gain conditions (Loud and Gain). It was hypothesized that hearing-impaired listeners' performance would be better in the two gain conditions than in the low level condition. Second, results from the Loud and Gain conditions were also compared, in order to determine if there was a difference in performance between the two conditions. It was hypothesized that there would not be any difference in performance between the Loud and Gain conditions. However, if there was a difference between these two conditions, the hearing-impaired listeners were expected to perform better in the Gain condition than the Loud condition, possibly due to upward spread of masking in the Loud condition.

B. Method

1. Subjects

Five young, hearing-impaired adults with sloping, mild-to-moderate sensorineural hearing loss were paid for their participation in this study. Two women and three men, ranging from 21 to 41 years of age, were tested (mean=31 years). All were native speakers of American English. Three

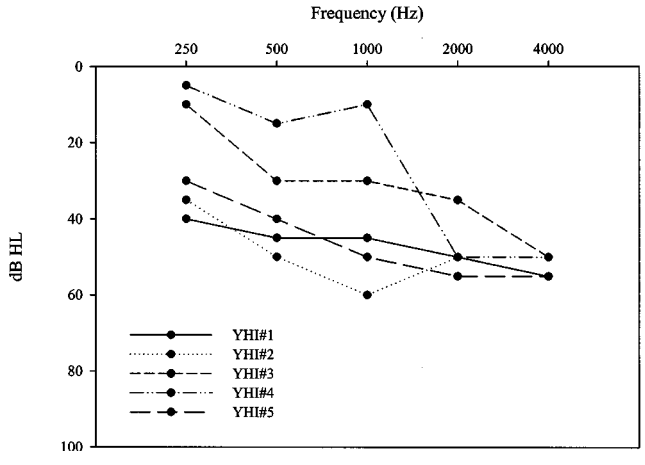


FIG. 1. Pure tone thresholds (dB HL) for the five young, hearing-impaired listeners.

of the hearing-impaired listeners wore binaural hearing aids and reported being consistent hearing aid users, while two of the listeners had never worn hearing aids. None of the listeners used hearing aids during the experiment.

The magnitude and configuration of hearing loss were selected to be similar to those of the EHI listeners tested in Coughlin *et al.* (1998). In order to qualify for this vowel perception study, hearing threshold levels between 30 and 60 dB HL (ANSI, 1996) at 2000 Hz, and otherwise no thresholds greater than 65 dB HL between 250 and 4000 Hz, were required for the test ear.

There was no evidence of middle ear pathology at time of testing, as determined by normal tympanometric results. Air-conduction pure-tone thresholds for octave frequencies between 250 and 8000 Hz were obtained in a sound-treated room using insert (ER-3A) earphones, and can be seen in the audiograms in Fig. 1.

For each listener, only one ear was tested in this study. The ear with the pattern of hearing loss most representative of selection criteria was tested; in some cases this was the listener's better hearing ear and in some cases the opposite was true. Across all listeners, asymmetry at any one audiometric frequency was <40 dB, and therefore masking of the nontest ear was not performed.

While the audiogram of one listener (YHI#4) is distinct from that of the four other listeners, in the lower frequency (F1) range, this listener did match the specific pattern of hearing loss described above. Further, all YHI listeners performed similarly on F1 discrimination in that frequency region, in all listening conditions. Data from YHI# 4 were not notable, and, thus, the inclusion of this listener seems appropriate.

2. Stimuli

The stimuli were five steady-state English vowels, /ɪ e ɛ ʌ æ/. These monophthongal vowels were selected based on previous work indicating that they would maximize response variability as required for successful correlational analyses (e.g., Coughlin *et al.*, 1998; Nábělek *et al.*, 1992), likely due to their acoustic similarity. The /ʌ/ vowel was added to the set of four vowels used in the study reported by Coughlin *et al.* (1998) in order to represent a slightly broader range of formant frequencies.

TABLE I. Frequencies in Hz for formants F1–F3 used in synthesizing vowel stimuli. The range values for the vowels indicate the formant values for the test stimuli, when F1 or F2 was incremented logarithmically from the standard.

	Standard vowels			Range values	
Vowel	F1	F2	F3	F1	F2
ɪ	450	2300	3000	455–522	2315–2513
e	550	2500	3100	555–621	2515–2712
ɛ	600	2200	3000	605–671	2210–2341
ʌ	700	1400	2600	705–771	1410–1542
æ	1000	1950	3000	1007–1107	1960–2091

The stimuli were synthesized based on spectrographic measurements of a female speaker, using the cascade branch of the KLTSYN synthesizer (Klatt, 1980). Stimuli were synthesized at 10 000 Hz. These five vowels were called the ''standard'' vowels used in the discrimination experiment, with formant frequencies shown in Table I. In addition to the five standard vowels, vowel test stimuli were synthesized in sets of 14, for discrimination purposes. In each test set, either the first or second formant (F1 or F2) was incremented from the standard value, according to a log ratio, with ranges seen in Table I. For all stimuli, the fundamental frequency fell in a linear manner from 220 to 180 Hz, over the vowel's duration. Vowel duration was set to 175 ms, the average of the speaker's long and short vowels. Bandwidths for the formants were the same for all vowels, with $BW1 = 70$ Hz, $BW2 = 90$ Hz, and $BW3 = 170$ Hz. In order to minimize inherent intensity differences between the five vowels, their amplitudes were then equated as follows: the rms energy over a 30-ms window for the five standard vowels was measured at the maximum amplitude (approximately 20–50 ms after vowel onset). A median rms value was chosen and all five vowels were then adjusted to be within ± 1 dB of this median value.

3. Listening conditions

Three conditions were designed to simulate everyday listening conditions, from low to amplified levels. The Soft condition simulated an average conversational level. To ensure that vowels would be at least partially audible, vowel levels were checked individually for each listener as follows. On an individual basis, each listener's pure tone average was calculated from thresholds at 500, 750, 1000, 1500, and 2000 Hz and converted to dB SPL. Next, vowel spectra were calculated as follows. F1 of the calibration vowel was measured in dB SPL through TDH-39 headphones and a 6-cm^3 coupler. Samples of each vowel were located near the peak amplitude, multiplied by a hamming window, and a 512-point FFT of the signal was calculated. The values of the first 23 harmonic peaks were then extracted from each of the five vowels and converted to dB SPL (*re*: the calibration vowel) for each harmonic. Next, maxima were selected for each harmonic peak to represent the highest levels across the five vowels. These maxima are shown in Fig. 2 by the solid line.

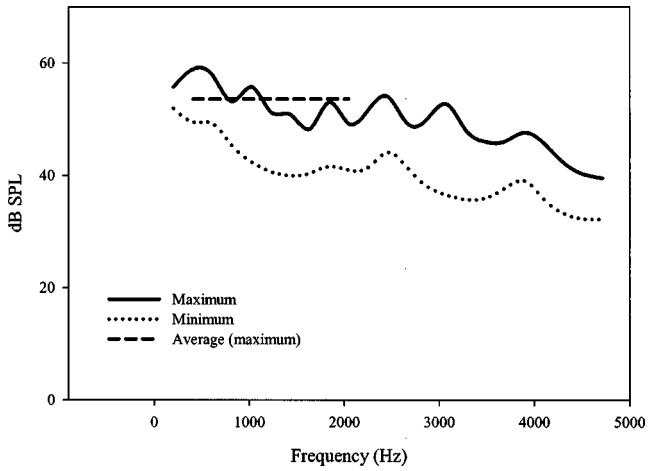


FIG. 2. Spectral peaks of the synthetic vowel stimuli, used to calculate levels for the Soft and Gain conditions. The dashed line representing the average of the maximum peak values (between 500 and 2000 Hz) had to be at least 10 dB supra-threshold for each listener in the Soft condition. The dotted line representing the spectral minima was at least 15 dB supra-threshold in the Gain condition. All dB SPL values are referenced to a 6-cm^3 coupler.

The minimum spectral level of each harmonic peak was also calculated for the five vowels, shown by the dotted curve. (The minimum spectral level is relevant to the Gain condition described below.) The maxima at only harmonics corresponding most closely to 500, 750, 1000, 1500, and 2000 Hz were then averaged. This measurement was termed the average maximum spectral level for the five vowels, and found to be approximately 54 dB SPL (see Fig. 2). All dB SPL values in Fig. 2 are thus referenced to the 6-cm^3 coupler.

Finally, the Soft presentation level was set such that the average maximum spectral level of the vowel stimuli was 10 dB above each listener's pure tone average. This ensured that the vowels would be partially audible. That is, if the listener's pure tone average threshold was less than 10 dB below the maximum spectral level of the vowel stimuli (54 dB SPL), presentation level was increased accordingly. When the maximum spectral level of the vowel stimuli exceeded a listener's pure tone average threshold by at least 10 dB, 60 dB SPL was chosen as the presentation level. Signal levels for the five listeners tested were 60 (YHI#3 and YHI#4), 74 (YHI#1), 75 (YHI#5), and 81 (YHI#2) dB SPL.

The Loud listening condition simulated an audible level for the speech. In this condition the presentation level was 95 dB SPL for each listener. The Gain listening condition was designed to simulate the frequency response provided by a linear gain hearing aid. The amount of gain was determined so that the five-vowel minimum spectral level (dotted line, Fig. 2) was at least 15 dB above the listener's hearing threshold in dB SPL, for each of the frequencies: 250, 500, 750, 1000, 1500, 2000, 3000, and 4000 Hz. This approach was patterned after the desired sensation level (DSL; Cornelisse *et al.*, 1995) approach for the prescription of hearing aid gain. A specific FIR filter was designed to provide the necessary gain at each reference frequency for each listener in this condition, relative to the differences in hearing loss observed in the audiograms in Fig. 1.

4. Calibration

All stimuli were low-pass filtered with a cutoff frequency of 4300 Hz, preserving the first three formant peaks. The 4300-Hz cutoff was combined with the FIR filters designed to achieve the specific flat or shaped signal condition previously described. To achieve the Soft or Loud signal levels, separate flat frequency response, low-pass FIR filters were designed. Stimuli were output through a 16-bit D/A converter (Tucker-Davis Technologies, DA1) followed by a digital FIR filter (Tucker-Davis Technologies, PF1), and a headphone buffer (Tucker-Davis Technologies, HB6). Participants listened to stimuli under TDH-39 headphones, with sounds presented monaurally to the ear previously selected.

For calibration, the previously designed FIR filter for each listener and condition was loaded in the PF1. The calibration /æ/ vowel was then measured in a 6-cm^3 coupler with a Larson-Davis sound-level meter (model 800B) using the linear weighting scale. The desired signal levels were then checked for presentation accuracy within ± 0.5 dB and values were subsequently monitored throughout the experiment.

5. Task

A modified two-alternative, forced-choice procedure was implemented. The standard vowel was always presented first, followed by two more vowel stimuli. The two following vowels included the standard vowel and the test stimulus that differed from the standard in terms of a logarithmic formant frequency increment (see Table I for the range of these values). The listener's task was to determine whether the second or third vowel in the triplet differed from the first and enter the response via computer keyboard. The increment was dictated by an adaptive-track algorithm designed to track 70.7% correct on the psychometric function (Levitt, 1971). Initially, the standard vowel stimulus and its test stimulus differed maximally in terms of formant frequency. As performance improved, the difference between the standard and the test stimulus decreased, via a two-down one-up rule that made them increasingly difficult to discriminate. Listeners were tested individually, seated at a computer in a sound-treated room. Feedback was given in the form of a message on the screen indicating whether the listener's response was correct or incorrect.

6. Training

Other work from our laboratory has illustrated the importance of adequate training for establishing stable threshold estimates (e.g., Kewley-Port, 2001). Thus, all listeners completed two 90-min training sessions. This time-based approach to training was focused on F1 of the /ɪ/ vowel, in the Loud listening condition. Blocks were brief (typically lasting 5–6 min), and included 80 trials. It was thought this training condition would be easiest for listeners, due to the nature of the vowel and listening condition (the first and second formant peaks for /ɪ/ are well spaced in terms of frequency, and the level was supra-threshold for all listeners). All listeners' performance approached asymptote over two 90-min training sessions (approximately 23 blocks).

TABLE II. Means and standard deviations of F1 discrimination thresholds (in Hz) for YHI listeners.

		ɪ F1 (450 Hz)	e F1 (550 Hz)	ɛ F1 (600 Hz)	ʌ F1 (700 Hz)	æ F1 (1000 Hz)
Soft	mean	50.8	31.4	25.9	35.7	51.2
	s.d.	15.6	7.7	21.6	10.8	28.6
Loud	mean	27.9	31.1	22.2	31.6	33.3
	s.d.	9.5	12.0	10.3	20.8	21.3
Gain	mean	32.9	26.5	30.9	26.5	46.8
	s.d.	11.7	8.8	19.8	4.8	27.0

7. Testing

Testing was comprised of five 90-min sessions following training. In each block, listeners were tested on all five vowel stimuli for an increment in either F1 or F2. All ten types of vowel stimuli (5 vowels×2 formant conditions) were randomized within each block of 80 tokens. The three listening conditions varied by block, such that all tokens within a block were presented at the same level. Listeners cycled through the listening conditions (Soft, Loud, and Gain) at their own pace, completing 6, 9, or 12 blocks per test session. A measure of the minimum discriminable formant frequency difference between standard and test stimuli was calculated as ΔF in Hz. ΔF was obtained based on an average of the mean reversals from each listener's best four consecutive blocks within a particular listening condition.

C. Results

Discrimination results were obtained as ΔF in Hertz (for F1 and F2) for the five vowels (/ɪ e ɛ ʌ æ/) in the three level conditions (Soft, Loud, and Gain) for each of the five listeners. Group results for the discrimination experiment are reported in Tables II and III. Formant thresholds for 5 vowels ×2 formants (F1 or F2 varied) averaged over all YHI listeners for the three signal level conditions are shown in Fig. 3. Note that thresholds are nearly constant in the F1 region, and are elevated and increase over the F2 region, for all conditions.

The first goal was to determine whether listeners performed in an appreciably different way in the Soft listening condition compared to the Gain and Loud conditions. It was hypothesized that performance would be substantially better in the Gain and Loud conditions than in the Soft condition, evidenced by smaller ΔF values. Results for the Gain and

TABLE III. Means and standard deviations of F2 discrimination thresholds (in Hz) for YHI listeners.

		ʌ F2 (1400 Hz)	æ F2 (1950 Hz)	ɛ F2 (2200 Hz)	ɪ F2 (2300 Hz)	e F2 (2500 Hz)
Soft	mean	89.9	108.4	119.5	182.6	170.4
	s.d.	41.7	41.6	26.9	53.5	51.9
Loud	mean	111.9	88.0	96.5	139.8	127.9
	s.d.	33.4	66.4	47.3	59.3	80.2
Gain	mean	86.1	96.0	96.1	112.9	118.4
	s.d.	48.1	42.0	39.3	61.7	79.7

TABLE IV. Vowel identification performance for YHI listeners, means and standard deviations in percent correct.

		ɪ	e	ɛ	ʌ	æ
Soft	mean	85.00	42.50	61.88	99.69	99.38
	s.d.	17.14	33.74	23.11	0.63	0.72
Loud	mean	95.00	63.44	74.06	98.13	98.44
	s.d.	4.21	35.67	25.99	1.61	3.13
Gain	mean	90.00	56.88	80.31	97.19	97.81
	s.d.	7.43	49.84	22.39	3.59	4.38

Loud conditions were thus averaged on an individual basis, for each vowel tested, to yield ΔF values in terms of performance on the "non-Soft" conditions. The difference was then calculated between each listener's performance in the Soft condition and the non-Soft conditions, to yield difference scores for each vowel tested, for each listener. Difference scores (between Soft and non-Soft performance on the ten vowels) were then subjected to a repeated measures multivariate analysis of variance (MANOVA) test of significance. The MANOVA statistical method was adopted for the following reasons. The listeners in this study were tested on the same dependent measure (vowel formant difference limen in Hertz) in different listening conditions. However, it is possible that the contrasts involved in testing these effects are not independent of each other, i.e., compound symmetry and sphericity might be violated (see StatSoft, Inc., 2002). If multivariate criteria are used to simultaneously test the significance of two or more repeated measures contrasts, they do not need to be independent of each other (StatSoft, Inc., 2002). The MANOVA, a more stringent statistical test, bypasses the assumptions of the univariate analysis of variance related to compound symmetry and sphericity. This test seemed appropriate for this data, and the results of the MANOVA may be interpreted in basically the same way as an ordinary analysis of variance (Aron and Aron, 1999, p. 535). As expected, this test revealed a significant difference between listeners' performance on the Soft and non-Soft conditions [$F(1,4)=558.283$, $p=0.031$].

Next, the performance in the Gain and Loud conditions

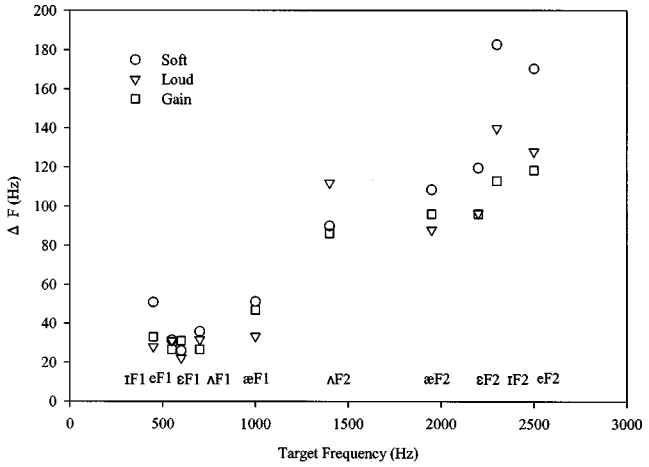


FIG. 3. YHI group discrimination results by listening condition, for F1 and F2 increments in five vowels.

was compared, in order to determine whether the upward spread of masking might have affected performance. Difference scores were calculated for each listener, and each vowel formant tested based on the ΔF values for the Loud and Gain conditions and submitted to a repeated measures MANOVA. This test revealed no significant difference between performance in the Loud and Gain conditions [$F(1,4)=2.423$, $p=0.444$].

Together, these results show that while the hearing-impaired listeners performed better in the non-Soft (i.e., Loud and Gain) conditions than the Soft condition, as expected, there was no significant difference between performance in the Loud and Gain conditions. This suggests that listeners' performance in the Loud condition was not adversely affected by the upward spread of masking.

III. EXPERIMENT 2: IDENTIFICATION OF VOWEL STIMULI UNDER DIFFERENT LISTENING CONDITIONS

A. Purpose

The purpose of this identification experiment was to examine classification abilities for the standard vowel stimuli under the same listening conditions as in Experiment 1. The same predictions for the discrimination experiment were expected to hold for the identification results; listeners were expected to perform better in the Gain and Loud conditions than in the Soft condition, while no difference in performance was expected between the Gain and Loud conditions. If there was a difference seen between the Loud and Gain conditions, poorer performance was expected in the Loud condition, possibly due to the effect of upward spread of masking.

B. Method

Four of the five listeners (YHI#2–YHI#5) who participated in the discrimination experiment also participated in the identification experiment. The five standard vowels used in the discrimination experiment described above were used in the identification experiment. The same listening conditions (Soft, Loud, and Gain) and calibration procedures described for the discrimination experiment were used for the identification experiment.

1. Task

In this experiment, listeners were presented with the vowel stimuli and asked to identify them in a forced-choice classification task. Five key-word responses containing the standard test vowels were displayed, each matched to a single and unique key on the keyboard. Listeners were asked to pick the word from the key-word list with the vowel sound most like the one they heard and then enter the response on a keyboard.

2. Training

Training in this portion of the study consisted of one block of 25 ordered tokens (i.e., five repetitions of each of the five vowels, presented in the order of the key-words on the computer screen), followed by one block of the same tokens randomized, in the Loud listening condition. Feedback was given in both the ordered and randomized training blocks. These two blocks of training were performed at the beginning of each of two test sessions.

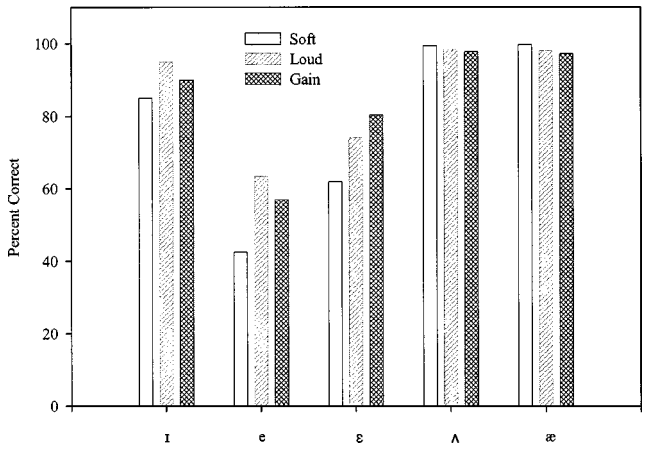


FIG. 4. YHI group identification results for five vowels in three listening conditions.

3. Testing

After training, the remainder of the two 90-min sessions involved testing. The five standard vowels were presented 20 times each (randomized), in blocks of 100 tokens, in each listening condition. No feedback was given during testing. All listeners completed at least four blocks of testing. Performance was thus measured in terms of the percent correctly identified vowels over the first four blocks of testing.

C. Results

Group results for the three listening conditions can be seen in Fig. 4 and Table IV. To determine whether the effects of listening condition seen in the discrimination task (Experiment 1) would also be seen in the identification task, performance in the Soft condition was compared with performance in the non-Soft (i.e., averaged Gain and Loud) conditions. Percent correct values were transformed to rationalized arcsine units (RAUs; Studebaker, 1985). The RAU transform reduces the compressive effect of the percent correct scale for extreme values (at either end of the scale) while percent correct scores between 15% and 85% remain roughly the same. The Soft RAUs were subtracted from those in the non-Soft listening conditions to yield difference scores for each of the four listeners on the five standard vowels. The difference scores were then subjected to a repeated measures MANOVA. However, this test revealed no significant difference between performance in the Soft and non-Soft (Loud and Gain) conditions [$F(1,3)=0.703$, $p=0.681$]. This result is not surprising given that three of the five vowels were identified almost perfectly under all three listening conditions.

Results from the Loud and Gain conditions were also transformed to RAUs, and compared, in order to determine whether there was a significant difference between performances in those two different listening conditions. Difference scores were calculated for each listener and performance on each vowel, in the Loud and Gain conditions. These difference scores were then subjected to a repeated

measures MANOVA. This test also revealed no significant difference between performance in the Loud and Gain conditions [$F(1,3)=0.724$, $p=0.675$]. These results show that the level effect noted in the discrimination task, the significant difference between performance in Soft and non-Soft conditions, was eliminated in the identification task.

In Fig. 4 it may also be noted that performance seems to be less than perfect for two of the five vowels. The Soft listening condition (which showed the poorest performance on the vowels /e/ and /ɛ/) was selected for analysis and a within-subjects one-way ANOVA confirmed the highly significant effect of vowel. Posthoc tests showed that performance on the vowels /e/ and /ɛ/ was significantly different from performance on the other three vowels. However, performance on /e/ and /ɛ/ was not significantly different. For these two poorly identified vowels, analysis of all listening conditions revealed that errors were likely related to the spectral properties of the vowels. That is, /e/ was often confused with the two vowels most similar to it in terms of first formant frequency, /ɪ/ and /ɛ/. The /ɛ/ vowel was often confused with /e/, which had similar F1 and F2 frequencies. Identification results for YHI are thus consistent with those from other studies that report relatively high variability in the vowel identification abilities of listeners with mild-to-moderate sensorineural hearing loss, with some vowels accurately identified and other vowels poorly categorized (e.g., Nábělek *et al.*, 1992). When errors occur, they are likely the result of confusion among vowels with similar spectral properties.

IV. COMPARISON OF RESULTS BETWEEN YOUNG AND ELDERLY HEARING-IMPAIRED LISTENERS

Results from the present experiments were compared with those from previous work in the same lab, reported in Coughlin *et al.* (1998), to see whether age or hearing impairment factors contribute to decreased vowel perception abilities. In Coughlin *et al.* (1998), a group of four elderly hearing-impaired listeners (69–74 years of age) performed discrimination and identification tasks identical to those reported in this study, for four of the same vowels (/ɪ e ɛ æ/; the /ʌ/ vowel was not heard by the EHI listeners). The hearing loss configuration was also similar between studies (see Fig. 5). Elderly listeners in the Coughlin *et al.* study had moderate, sloping sensorineural hearing loss very similar to that measured in four of the five YHI listeners in this study (all but YHI#4). However, signal levels varied slightly between studies. The EHI listeners tested in Coughlin *et al.* listened to signals presented at 70 and 95 dB SPL. For this reason, results will be compared for performance in the Loud condition, in which listeners in both studies heard vowels presented at 95 dB SPL.

Otherwise, discrimination and identification testing for the two groups were identical between the two studies. That is, listeners followed the same training protocol, including the same amount of training on the same stimuli (the /ɪ/ vowel), in the same listening condition (95 dB SPL). Discrimination and identification testing was conducted on the same materials in the same sound booth, with the same calibration techniques. Efforts were made to achieve parity between the EHI group tested previously and the YHI group tested and reported here. Results from the time-based training method used for both YHI and EHI listeners suggested near asymptotic performance in both groups (the average ΔF of the last four blocks of training sessions was 17.9 Hz for the EHI group, and 12.0 Hz for the YHI group).

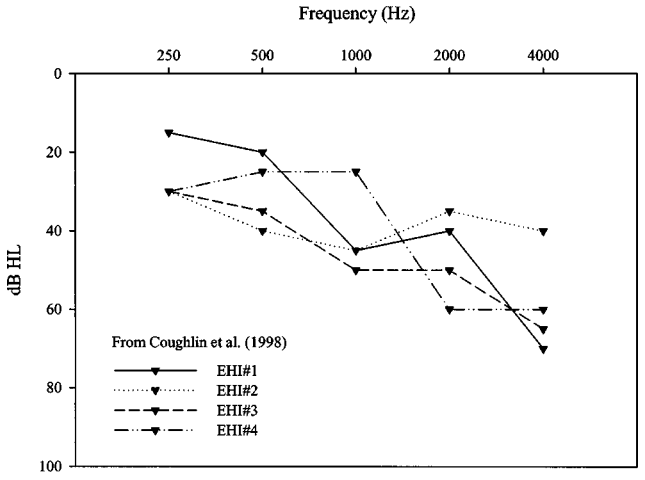


FIG. 5. Pure tone thresholds (dB HL) for the four elderly, hearing-impaired listeners reported in Coughlin *et al.* (1998).

Individual discrimination threshold results for the two groups can be seen in Fig. 6. A between-group, with formant as a repeated measure, analysis of variance was performed on the discrimination results for the two groups. This test showed no main effect of group (YHI vs EHI) to yield [$F(1,7)=1.073$, $p=0.335$]. However, a significant main effect of formant [$F(7,49)=20.689$, $p<0.001$] and a significant group×formant interaction was seen [$F(7,49)=2.549$, $p=0.025$]. Since most of the variability was observed in the F2 region, performance on those formants was then subjected to a repeated-measure, two-way ANOVA (group, formant). No main effect of group [$F(1,7)=1.322$, $p=0.288$] or formant [$F(3,21)=0.284$, $p=0.836$] was obtained. However, a significant group ×formant interaction was obtained [$F(3,21)=3.814$, p

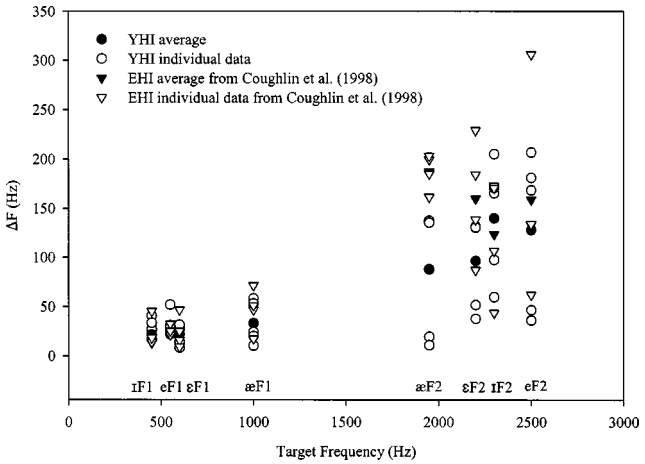


FIG. 6. Individual thresholds and averages for the five YHI listeners in the present study and four EHI listeners reported in Coughlin *et al.* (1998) for F1 and F2 increments in four standard vowels presented in the Loud listening condition (95 dB SPL).

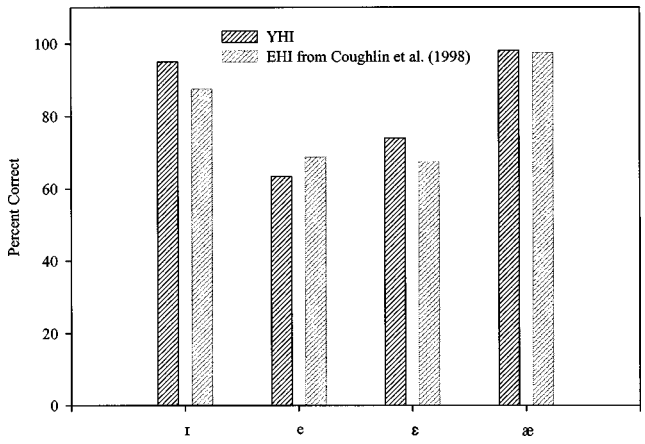


FIG. 7. Identification results for four YHI listeners in the present study and four EHI listeners reported in Coughlin *et al.* (1998) for four standard vowels presented in the Loud listening condition (95 dB SPL).

$=0.025$]. A closer examination of YHI–EHI performance in the F2 region was undertaken, and planned comparisons revealed a significant difference between groups only for discrimination of /æ/ F2 [$t(1)=2.865$, $p=0.023$], as may be expected from Fig. 6.

Coughlin *et al.* (1998) report that formant frequency discrimination for YNH, ENH, and EHI listeners were approximately 30 Hz in the F1 region, with thresholds increasing as a function of frequency in the F2 region. The YHI group average formant frequency threshold in the F1 region obtained in the present study was also approximately 30 Hz. These threshold values are only somewhat higher than those reported for highly trained listeners in a similar task in Kewley-Port (2001), approximately 20 Hz in the F1 region.

However, Coughlin *et al.* (1998) showed that in the F2 region, thresholds for EHI listeners were about 290% higher than YNH listeners under the same conditions. For the YHI listeners tested in this study, a 210% increase in F2 thresholds is seen over the YNH listeners tested in Coughlin *et al.* (1998) under the same conditions. Given the nonsignificant difference seen between EHI listeners reported in Coughlin *et al.* (1998) and YHI discrimination performance in the present experiment, it appears that mild-to-moderate levels of hearing impairment are associated with a greater than 200% decrease in second formant resolution, even when the vowels are fully audible.

Results were also compared between groups for identification performance on the four vowels in the Loud listening condition (95 dB SPL). Results seen in Fig. 7 are essentially identical for the two groups. Scores were converted from percent correct to RAUs and submitted to an ANOVA for the mixed-level factors of group and vowel. As expected, this test showed no significant main effect of group [$F(1)=0.011$, $p=0.921$], and a significant main effect of vowel [$F(3)=5.669$, $p=0.006$]. As noted in Coughlin *et al.* (1998), these results from hearing-impaired listeners differ from those of young normal-hearing listeners, who identify all four vowels with $>97\%$ accuracy.

In sum, comparison of YHI and EHI performance on discrimination and identification of vowels in the Loud condition was not significantly different between groups. This indicates that for listeners with similar audiometric configuration, systematic age-related differences in performance are not seen in these types of speech perception tasks. In contrast to what Coughlin *et al.* (1998) suggested, age was not found to be a factor contributing to decreased vowel perception performance for these hearing-impaired listeners.

V. DISCUSSION

A. Upward spread of masking

Florentine *et al.* (1980) note that for many listeners with cochlear hearing impairment, an increased spread of masking toward higher frequencies is seen. In the case of vowel perception, the increased upward spread of masking for hearing-impaired listeners may have the effect that the higher-level first formant may mask lower-level formants at higher frequencies. Some hearing aid designs take this into account by effectively reducing the amount of gain given to the lower-frequency spectral components of the input signal.

In Coughlin *et al.* (1998) it was possible that the upward spread of masking affected performance of listeners in the Loud listening condition because effects were mostly seen in the F2 region. In order to determine whether the upward spread of masking impaired vowel perception in the Loud listening condition, the present study included a Gain condition which provided a shaped gain function (*re*: 60 dB SPL signal level) on an individual basis. However, as noted in the results sections above, no differences were found between performance in the Loud and Gain conditions for either discrimination or identification performance. This indicates that for these listeners, if upward spread of masking occurred in the Loud listening condition (compared to the Gain condition), it did not seem to affect performance on these vowel discrimination and identification tasks.

B. The relation between discrimination and identification performance

For the YHI listeners, an important question is how vowel discrimination performance relates to identification performance. In this regard, it is interesting to determine whether identification of the vowels in this study was related to their discrimination. Though the level effects obtained in the discrimination experiment were not obtained in the identification experiment, the vowels were not identified uniformly well.

It is hypothesized that the somewhat impaired identification of these vowels may be related to listening in the F1 frequency range, for these listeners. As mentioned above, first formant frequency discrimination abilities for the YHI listeners in this study were similar to those of the YNH listeners reported in Coughlin *et al.* (1998) while discrimination for F2 was degraded by more than 200%. What can then explain the identification results showing three vowels were identified nearly perfectly while two vowels, /e/ and /ɛ/, were correctly identified around 60%? It is argued that the vowels closer together in terms of formant frequency are more difficult for YHI listeners to correctly identify [cf. the error patterns obtained by Owens *et al.* (1968), Dorman *et al.* (1985), Turner and Henn (1989), and Nábělek *et al.*

(1992)]. That is, spectral similarity amongst the vowel stimuli may lead to decreased identification performance.

The two vowels that were misidentified most often in this study, /e/ and /ɛ/, are the vowels closest together in terms of first formant frequency. Kewley-Port and Zheng (1999) have previously argued that the distance between the closest English vowels in F1×F2 space must exceed discrimination thresholds if they are to be reliably identified in normal discourse. A metric was thus designed to determine whether this relation holds or not for the present data for YHI listeners.

For F1 and F2 of each standard vowel, the distance between that formant and its nearest neighbor was calculated, in Hertz. In order to obtain an individualized measure of each listener's frequency resolution in the F1 and F2 regions, ΔF1 and ΔF2 in Hz (averaged over the five vowels) were calculated for each listener, in each listening condition. This measure of frequency resolution was compared with the separability of the vowels in the following manner. For each listener in each condition, the average ΔF1 was subtracted from the distance between that vowel's F1 and its nearest neighbor. The same calculations were performed for the F2 region. This yielded a measure of whether the vowel formant should have been discriminable from its nearest neighbor or not (positive results indicated discriminable neighbors, negative numbers indicated indiscriminable neighbors). These results were then correlated with identification performance achieved for that listening condition, in RAUs.

The results of a Pearson product moment correlation between the described metric in the F1 region and correct identification performance in RAUs yielded a moderate level of correlation, $r=0.487$ ($p<0.05$). Results of a Pearson product moment correlation between the described metric in the F2 region and correct identification performance in RAUs yielded a similar correlation, $r=0.492$ ($p<0.05$). This indicates that to the limited extent that variability exists in identification performance, it may be partially accounted for by the relation between frequency resolution (of ΔF1 or ΔF2) and the proximity of the vowels to each other in terms of frequency.

Supporting the argument that the YHI listeners may be using F1 frequency information to identify the vowels is the fact that negative values (indicating indiscriminable vowel neighbors) were only obtained for the metric in the F2 region. This follows from the fact that in some cases, the average ΔF2 exceeded the distance between two vowels. While this is not conclusive proof of a uniform listening strategy being employed by all YHI listeners in the study, it does suggest that identification performance may be related to the frequency resolution capabilities of the YHI listeners in combination with the relative distance (i.e., confusability) between the vowels. Thus, while normal-hearing listeners use both F1 and F2 information to correctly identify vowels, it appears hearing-impaired listeners use primarily F1 cues due to vowel confusability and reduced abilities in the F2 region.

C. Effects of age and cognitive factors on performance

This study examined vowel perception by YHI with a configuration of sensorineural hearing loss similar to that of EHI listeners tested in a previous study (Coughlin *et al.*, 1998). Age as a factor affecting performance on tests of speech perception is of particular interest, since it has been argued that cognitive and auditive factors associated with aging may confound tests of speech perception. Two different perceptual tests were administered here. For the vowel identification task performance by young and elderly listeners was identical. This task seems simpler and more related to everyday listening than the second task, a laboratory discrimination task. In the discrimination task performance was not significantly different between young and elderly listeners. The clear interpretation of these data is that similar hearing impairment resulted in nearly identical performance, regardless of age. The elevated F2 threshold for the EHI listeners for the /æ/ vowel does leave open the possibility that in a more cognitively challenging perceptual task, some age differences may be observed.

As noted by other researchers, hearing-impaired listeners have broadened auditory filters that may distort acoustic signals (e.g., Peters and Moore, 1992; Turner and Henn, 1989). While the effects of broadened auditory filters on speech perception are often difficult to discern, it is possible that they may be a contributor to the elevated F2 thresholds observed in the present study. Based on the results offered in this experiment, other factors such as age and upward spread of masking have been ruled out. However, it seems that the effects of broadened auditory filters on vowel identification may be slight at least in cases of mild-to-moderate sensorineural hearing impairment (cf. Van Tasell *et al.*, 1987). The effects observed in the present data were restricted to only two of the five vowels tested, but the complete overlap of the young and elderly results gives strong support to the primacy of auditive factors.

While changes in cognitive factors may be associated with aging, and thus affect performance on auditory tests, converging evidence including that presented in this study indicates that the audibility of the acoustic signal available to the hearing-impaired listener is more predictive of performance than age alone. Continued development and refinement of testing techniques has revealed a strong relation between the audibility of the speech signal and the performance abilities of the listener, for these tasks, regardless of age. Nonetheless, further exploration of the cognitive processing abilities involved over a wide range of tests of speech perception, particularly in adverse listening conditions, is still warranted.

VI. CONCLUSIONS

This study examined the effects of mild-to-moderate sensorineural hearing loss on vowel discrimination and identification abilities in YHI adults. Expected level effects were seen in the discrimination experiment, such that listeners performed better in the Loud and Gain conditions than in the Soft condition, though there was no significant difference in performance between the Loud and Gain conditions. These level effects seemed to lessen somewhat in the vowel identification experiment. There was no evidence that the upward spread of masking affected listeners' performance on either vowel perception task. The effects of hearing impairment

were negligible for discrimination in the F1 region, but apparent in the F2 region. In the vowel identification task, categorization of two of the five vowels showed effects of sensorineural hearing loss. A modest correlation was found between formant discrimination abilities and vowel identification.

Comparison of results with those from an earlier study (Coughlin *et al.*, 1998) failed to show significant differences in either discrimination or identification abilities between EHI and YHI listeners with a similar pattern of hearing loss, in the Loud listening condition. This finding suggests that hearing impairment, not age, is the primary contributor to vowel perception in elderly listeners with presbycusic hearing loss and young listeners with idiopathic sensorineural hearing loss, for these vowel perception tasks.

ACKNOWLEDGMENTS

The authors wish to thank Larry Humes for contributions to experimental design and insightful comments on an earlier version of this paper. This research was supported by NIHDCD-02229.

ANSI (**1996**). ANSI S3.6-1996, ''Specification for Audiometers'' (American National Standards Institute, New York).

Aron, A., and Aron, E. (**1999**). *Statistics for Psychology* (Prentice–Hall, Upper Saddle River, NJ).

Cornelisse, L., Seewald, R., and Jamieson, D. (**1995**). ''The input/output formula: A theoretical approach to the fitting of personal amplification devices,'' J. Acoust. Soc. Am. **97**, 1854–1864.

Coughlin, M., Kewley-Port, D., and Humes, L. (**1998**). ''The relation between identification and discrimination of vowels in young and elderly listeners,'' J. Acoust. Soc. Am. **104**, 3597–3607.

Dorman, M., Marton, K., Hannley, M., and Lindholm, J. (**1985**). ''Phonetic identification by elderly normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **77**, 664–670.

Dubno, J., and Dirks, D. (**1989**). ''Auditory filter characteristics and consonant recognition for hearing-impaired listeners,'' J. Acoust. Soc. Am. **85**, 1666–1675.

Era, P., Jokela, J., Qvarnberg, Y., and Heikkinen, E. (**1986**). ''Pure-tone thresholds, speech understanding, and their correlates in samples of men of different ages,'' Audiology **25**, 338–352.

Florentine, M., Buus, S., Scharf, B., and Zwicker, E. (**1980**). ''Frequency selectivity in normally-hearing and hearing-impaired observers,'' J. Speech Hear. Res. **23**(3), 646–669.

Godfrey, J., and Millay, K. (**1980**). ''Perception of synthetic speech sounds by hearing-impaired listeners,'' J. Aud. Res. **20**, 187–203.

Hargus, S., and Gordon-Salant, S. (**1995**). ''Accuracy of speech intelligibility index predictions for noise-masked young listeners with normal hearing and for elderly listeners with hearing impairment,'' J. Speech Hear. Res. **38**, 234–243.

Humes, L., Watson, B., Christensen, L., Cokely, C., Halling, D., and Lee, L. (**1994**). ''Factors associated with individual differences in clinical measures of speech recognition among the elderly,'' J. Speech Hear. Res. **37**, 465–474.

Jerger, J., Jerger, S., and Pirozzolo, F. (**1991**). ''Correlational analysis of speech audiometric scores, hearing loss, age, and cognitive abilities in the elderly,'' Ear Hear. **12**, 103–109.

Kewley-Port, D. (**2001**). ''Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context, and training,'' J. Acoust. Soc. Am. **110**, 2141–2155.

Kewley-Port, D., and Watson, C. (**1994**). ''Formant frequency discrimination for isolated English vowels,'' J. Acoust. Soc. Am. **95**, 485–496.

Kewley-Port, D., and Zheng, Y. (**1999**). ''Vowel formant discrimination: Towards more ordinary listening conditions,'' J. Acoust. Soc. Am. **106**, 2945–2958.

Klatt, D. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.

Klein, A., Mills, J., and Adkins, W. (**1990**). ''Upward spread of masking, hearing loss, and speech recognition in young and elderly listeners,'' J. Acoust. Soc. Am. **87**, 1266–1271.

Levitt, H. (**1971**). ''Transformed up-down methods in psychophysics,'' J. Acoust. Soc. Am. **49**, 467–477.

Náběl̆ek, A. (**1988**). ''Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss,'' J. Acoust. Soc. Am. **84**, 476–484.

Náběl̆ek, A., Czyzewski, Z., and Krishnan, L. (**1992**). ''The influence of talker differences on vowel identification by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **92**, 1228–1246.

Owens, E., Talbott, C., and Schubert, E. (**1968**). ''Vowel discrimination of hearing-impaired listeners,'' J. Speech Hear. Res. **11**, 648–655.

Peters, R., and Moore, B. (**1992**). ''Auditory filter shapes at low centre frequencies in young and elderly hearing-impaired subjects,'' J. Acoust. Soc. Am. **91**, 256–266.

Rakerd, B., Seitz, P., and Whearty, M. (**1996**). ''Assessing the cognitive demands of speech listening for people with hearing losses,'' Ear Hear. **17**, 97–106.

StatSoft, Inc. (**2002**). *Electronic Statistics Textbook*, Tulsa, OK: StatSoft. WEB: http://www.statSoft.com/textbook/stathome.html

Studebaker, G. (**1985**). ''A 'rationalized' arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.

Summers, V., and Leek, M. (**1997**). ''Intraspeech spread of masking in normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **101**, 2866–2876.

Turner, C., and Henn, C. (**1989**). ''The relation between vowel recognition and measures of frequency resolution,'' J. Speech Hear. Res. **32**, 49–58.

van Rooij, J., and Plomp, R. (**1992**). ''Auditive and cognitive factors in speech perception by elderly listeners. III. Additional data and final discussion,'' J. Acoust. Soc. Am. **91**, 1028–1033.

Van Tasell, D., Fabry, D., and Thibodeau, L. (**1987**). ''Vowel identification and vowel masking patterns of hearing-impaired subjects,'' J. Acoust. Soc. Am. **81**, 1586–1597.

Van Tasell, D., Hagen, L., Koblas, L., and Penner, S. (**1982**). ''Perception of short-term spectral cues for stop consonant place by normal and hearing-impaired subjects,'' J. Acoust. Soc. Am. **72**, 1771–1780.

The correlogram: A visual display of periodicity

Svante Granqvist[a)]
Department of Speech, Music and Hearing, KTH, Stockholm, Sweden

Britta Hammarberg[b)]
Department of Speech, Music and Hearing, KTH, Stockholm, Sweden and Department of Logopedics and Phoniatrics, Karolinska Institute, Huddinge University Hospital, Sweden

(Received 27 October 2000; accepted for publication 6 May 2003)

Fundamental frequency (F_0) extraction is often used in voice quality analysis. In pathological voices with a high degree of instability in F_0, it is common for F_0 extraction algorithms to fail. In such cases, the faulty F_0 values might spoil the possibilities for further data analysis. This paper presents the correlogram, a new method of displaying periodicity. The correlogram is based on the waveform-matching techniques often used in F_0 extraction programs, but with no mechanism to select an actual F_0 value. Instead, several candidates for F_0 are shown as dark bands. The result is presented as a 3D plot with time on the *x* axis, correlation delay inverted to frequency on the *y* axis, and correlation on the *z* axis. The *z* axis is represented in a gray scale as in a spectrogram. Delays corresponding to integer multiples of the period time will receive high correlation, thus resulting in candidates at F_0, $F_0/2$, $F_0/3$, etc. While the correlogram adds little to F_0 analysis of normal voices, it is useful for analysis of pathological voices since it illustrates the full complexity of the periodicity in the voice signal. Also, in combination with manual tracing, the correlogram can be used for semimanual F_0 extraction. If so, F_0 extraction can be performed on many voices that cause problems for conventional F_0 extractors. To demonstrate the properties of the method it is applied to synthetic and natural voices, among them six pathological voices, which are characterized by roughness, vocal fry, gratings/scrape, hypofunctional breathiness and voice breaks, or combinations of these. [DOI: 10.1121/1.1590972]

PACS numbers: 43.72.Ar [DOS] Pages: 2934–2945

I. INTRODUCTION

Fundamental frequency (F_0) is a commonly used parameter, being the main acoustic correlate to perceived pitch. In the field of voice quality research, F_0 extraction is particularly relevant, for example for evaluation of pre- and post-treatment of voice disorders and for measuring F_0 perturbation. Fundamental frequency extraction has received a great deal of attention in speech and voice research. Several different algorithms have been invented (e.g., Hess, 1983; Titze and Liang, 1993; Hess, 1995), and the algorithms have been applied both to acoustic waveforms and to electroglottographic (EGG) signals (Rothenberg, 1973; Fourcin, 1986). Examples of such methods are peak picking and methods based on spectral or cepstral properties of the signal or on waveform matching by means of autocorrelation or autodifference (e.g., Hess, 1983).

The waveform-matching technique has many important advantages. For example, it is independent of determination of the instant of excitation and has a low sensitivity to noise (Titze and Liang, 1993). Also, it can offer several estimates of fundamental frequency per period. The basic idea of the waveform-matching technique is to compare the signal in two time windows separated by a variable time delay. Certain lengths of this time delay will achieve a high correlation. These delays correspond to multiples of the period time. For example, the comparison can be realized in terms of a correlation function, which is a straightforward procedure since few variables are involved. If the waveform matching technique is to be used for F_0 extraction, the F_0 extraction algorithm must select which peak in the correlation function corresponds to the fundamental period time. Normal voices rarely cause selection problems. However, for dysphonic voices with an unstable period time, different F_0 extraction algorithms will give different results (Rabinov *et al.*, 1995; Karnell *et al.*, 1991). Phonation containing *bicyclic* segments (equivalent to *period doubling*, Titze, 1995) is a typical example; most F_0 extractors will select a fundamental frequency of $F_0/2$ for the bicyclic segments. This is in some sense correct since the period actually is doubled. However, different algorithms require different magnitudes of bicyclicity in order to arrive at this result. Hence, different F_0 extraction programs yield different results. This is problematic when the extracted F_0 data are used for deriving perturbation measures, such as jitter. For example, in a study of vocal fry, Blomgren *et al.* (1998) reverted to a semimanual extraction method instead of using the automatic methods, since their voice samples were characterized by a high amount of variability.

The problems outlined above do not originate from the waveform-matching algorithm but rather from the selection mechanism. Because fundamental frequency is defined as the inverse of period time, no F_0 exists if the signal is not perfectly repetitive, strictly speaking. Therefore, especially for pathological voices, there will be cases when the task of

[a)]Electronic mail: svante.granqvist@speech.kth.se
[b)]Electronic mail: britta.hammarberg@klinvet.ki.se

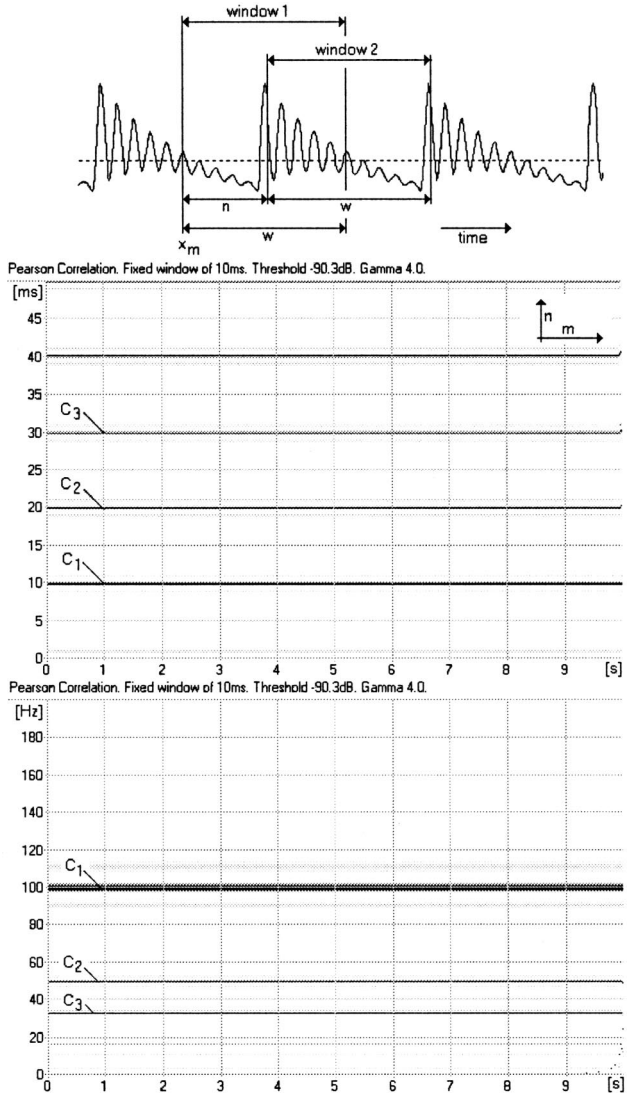


FIG. 1. The upper panel shows how the two correlation windows are applied to a perfectly periodic signal (100-Hz saw-tooth wave, formant filter at 1000 Hz, bandwidth 100 Hz). Middle and lower panels show the resulting time and frequency correlogram.

extracting a single F_0 is ill-defined or unrealistic. In such cases, an improved description of the perturbation itself may be more relevant.

In this paper, we present a display showing the raw correlation functions in a three-dimensional graph. We propose the term *correlogram* for these displays. The result is a picture reflecting periodicity characteristics of a voice rather than an extracted F_0 curve. The correlogram is free from a selection mechanism, leaving to the user to select the F_0 value or to what extent F_0 extraction is at all appropriate. This type of display should be particularly useful for voices where F_0 selections are difficult, that is, all voices with a high amount of F_0 perturbation. The method is tested on synthetic signals and natural voices and compared with other methods.

A. Descriptions of selected perceptual voice terms

In the case of pathological voices, information about periodicity or lack of periodicity is particularly relevant. Many attempts have been made to correlate F_0 perturbation characteristics with perceptual features. Such correlates are interesting, since a complete understanding of the relationship between perception and acoustics would allow objective measurements of voice qualities. In the following, some perceptual terms frequently used for pathological voices are reviewed together with the typically associated F_0 extraction problems.

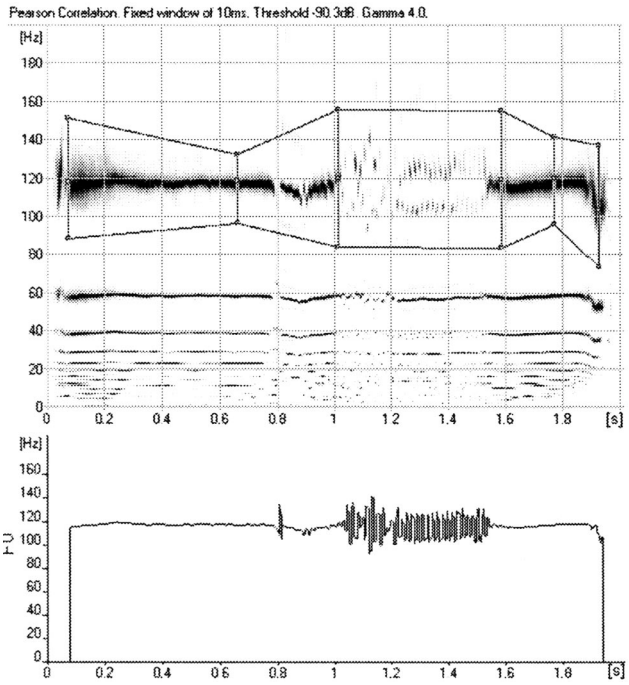


FIG. 2. The correlogram used for semimanual F_0 extraction. The first candidate was traced manually (upper panel) and the computer program then extracted the F_0 value within the traces that corresponded to the highest correlation (lower panel). Note the absence of octave leaps during the bicyclic segment at 1–1.5 s.

1. Creaky voice, vocal fry, and pulsed phonation

These terms appear to be associated either with low pitch and a prolonged glottal closed phase or by a complex pattern of glottal excitations, giving rise to subharmonics (Titze, 1995; Laver, 1980; Ladefoged, 1988; Hammarberg and Gauffin, 1995).

2. Roughness

This term also appears to be associated with period time perturbation. However, the term appears mostly, but not always, to be linked to a more random perturbation than what is commonly associated with the multicyclic type of vocal fry. The term is also sometimes associated with low-frequency noise (Hammarberg and Gauffin, 1995; De Krom, 1995; Titze, 1995; Ishiki *et al.*, 1969; Hillenbrand, 1988; Omori *et al.*, 1997; Imaizumi, 1986).

3. Gratings/scrape

This is a term mainly used in Sweden (Swedish: *skrap*). The term is often translated to ''high-frequency roughness'' (Hammarberg and Gauffin, 1995).

4. Breathiness

This is caused by soft or incomplete closure of the glottis and is often associated with high-frequency noise.

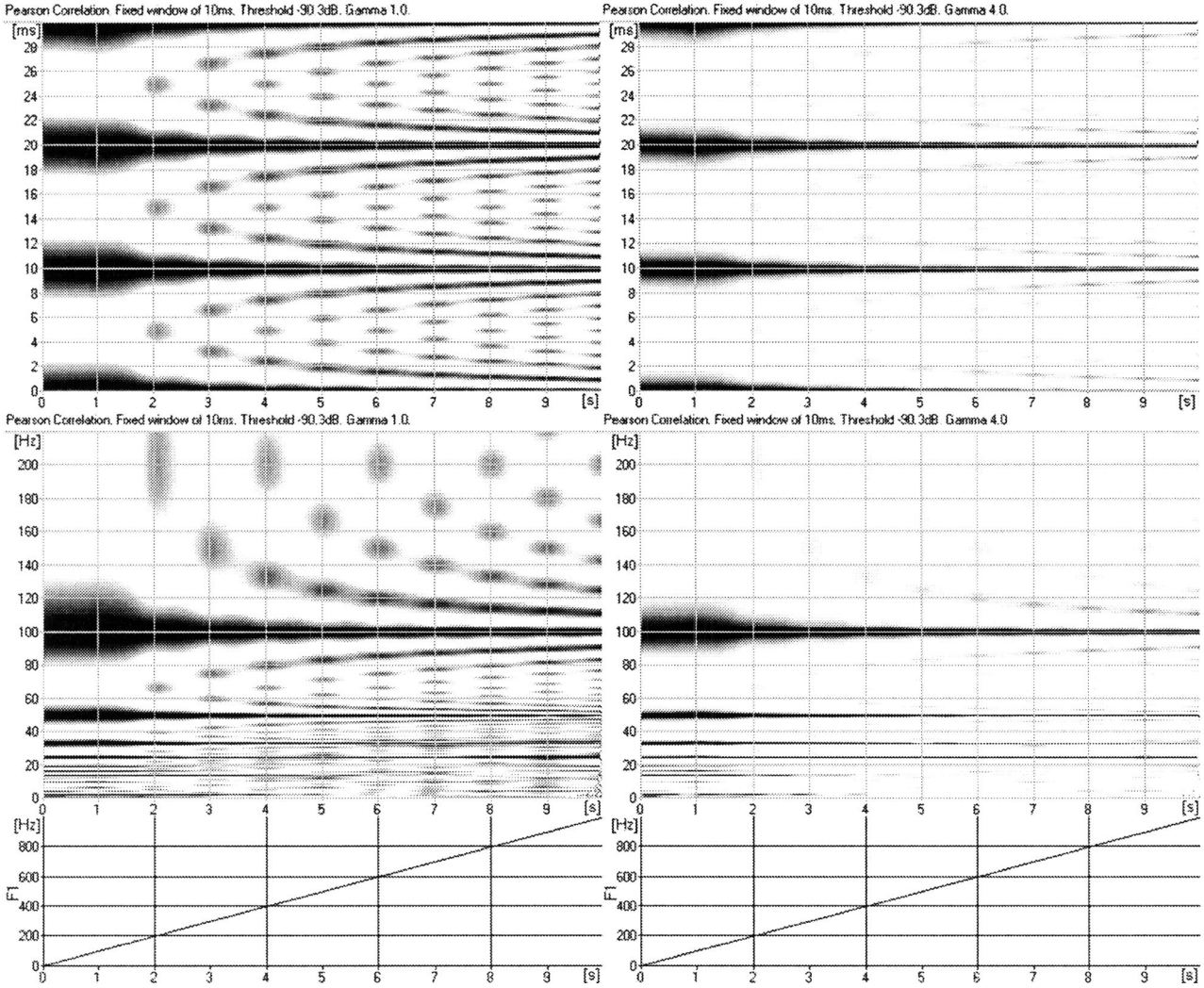


FIG. 3. Illustration of sidebands. The signal was a 100-Hz saw-tooth wave passed through a formant filter in which the formant frequency was increased from 0 to 1000 Hz (bandwidth=100 Hz). The sidebands are prominent with $\gamma=1$ (left panels) but become considerably suppressed by $\gamma=4$ (right panels).

Breathy voice can be produced in both hypo- and hyperfunctional laryngeal settings, which give rise to two different types of breathiness. These modes of phonation correlate strongly to a high or low relative level of the fundamental, respectively (Titze, 1995; Hammarberg and Gauffin, 1995; Hammarberg, 1986).

5. Voice breaks, vocal breaks, or register breaks

These occur when the vocal fold suddenly switches from one mode of vibration to another, for example between modal and falsetto register (Sundberg, 1987; Hammarberg, 1986; Švec and Pešák, 1994).

Most of the above voice qualities present problems for F_0 extraction. The complex patterns of glottal excitation that often are associated with vocal fry or gratings/scrape typically cause octave leaps in the F_0 curve, while pulsed phonation in principle can produce a smooth, continuous curve. Roughness, when associated with a random distribution of period time, mostly generates an unstable F_0 curve, but also, different F_0 algorithms tend to yield different F_0 values. Most programs generally handle breathy voices successfully since they contain little F_0 perturbation, but, for certain algorithms, particularly if event based, a high noise level can cause errors. F_0 extractors generally succeed in tracking the F_0 in voice breaks. All these problems with F_0 extraction in pathological voices are unfortunate since the F_0 perturbation appears to represent an important characteristic of such voices (Hillenbrand, 1988; Gauffin *et al.*, 1995). Hence, alternative methods for displaying periodicity variation should be useful in the analysis of pathological voices.

II. METHOD

The correlogram is based on the correlation between two time windows of the signal (Fig. 1). It displays the correlation in a novel manner in terms of a graph showing several such correlation functions, displayed in a gray scale similar to the Fourier transforms in a spectrogram. The method has been implemented by the first author (S.G.) as a program module of the SOUNDSWELL SIGNAL WORKSTATION software (Hitech Development AB, Sweden).

Different waveform-matching functions can be used. In this paper we have selected the Pearson correlation coefficient

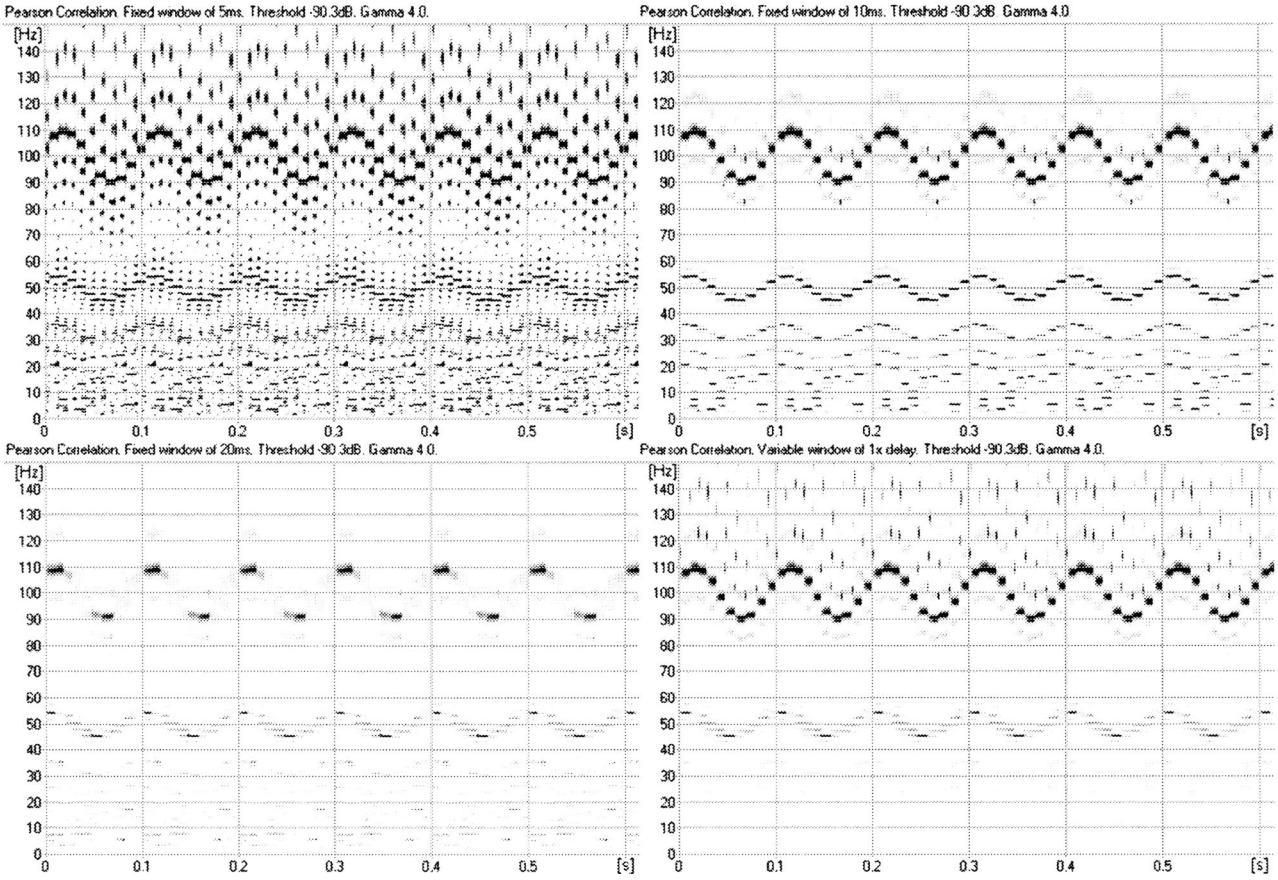


FIG. 4. The effects of window length: 5, 10, 20 ms, and variable (upper left, upper right, lower left, and lower right panel, respectively). The signal was a 100-Hz saw-tooth wave, frequency modulated at 10 Hz, ±10% (formant filter at 1000 Hz, bandwidth 100 Hz).

$$r_{m,n}=\frac{\sum_{k=m}^{m+w-1}(x_k-\overline{x_m})\cdot(x_{k+n}-\overline{x_{m+n}})}{\sqrt{\sum_{k=m}^{m+w-1}(x_k-\overline{x_m})^2\cdot\sum_{k=m}^{m+w-1}(x_{k+n}-\overline{x_{m+n}})^2}} \tag{1}$$

where

$$\overline{x_m}=\frac{\sum_{i=m}^{m+w-1}x_i}{w},$$

or, in computational form

$$r_{m,n}=\frac{w\sum_{k=m}^{m+w-1}x_kx_{k+n}-\sum_{k=m}^{m+w-1}x_k\cdot\sum_{k=m}^{m+w-1}x_{k+n}}{\sqrt{\left(w\sum_{k=m}^{m+w-1}x_k^2-\left(\sum_{k=m}^{m+w-1}x_k\right)^2\right)\cdot\left(w\sum_{k=m}^{m+w-1}x_{k+n}^2-\left(\sum_{k=m}^{m+w-1}x_{k+n}\right)^2\right)}}, \tag{2}$$

where x_j is the jth sample, m is the starting sample number, n is the delay separating the starting points of the windows, and w is the window width. This function is normalized, so the result will be restricted to the range $-1\leqslant r_{m,n}\leqslant 1$. When the delay n corresponds to one or several fundamental periods, a maximum will occur in the correlation coefficient. This is true, regardless of where in the fundamental period the starting point of the window is located, so there is no need to determine the point of excitation.

Note that the use of the Pearson correlation coefficient rather than a simple cross correlation is advantageous since only the Pearson alternative is insensitive to a dc component of the signal. A dc component increases the cross-correlation coefficient, in particular if it is large compared to the amplitude of the voice signal. Higher correlations observed due to a dc offset could be misleading in the case of periodicity analysis.

For each time value along the x axis a correlation coef-

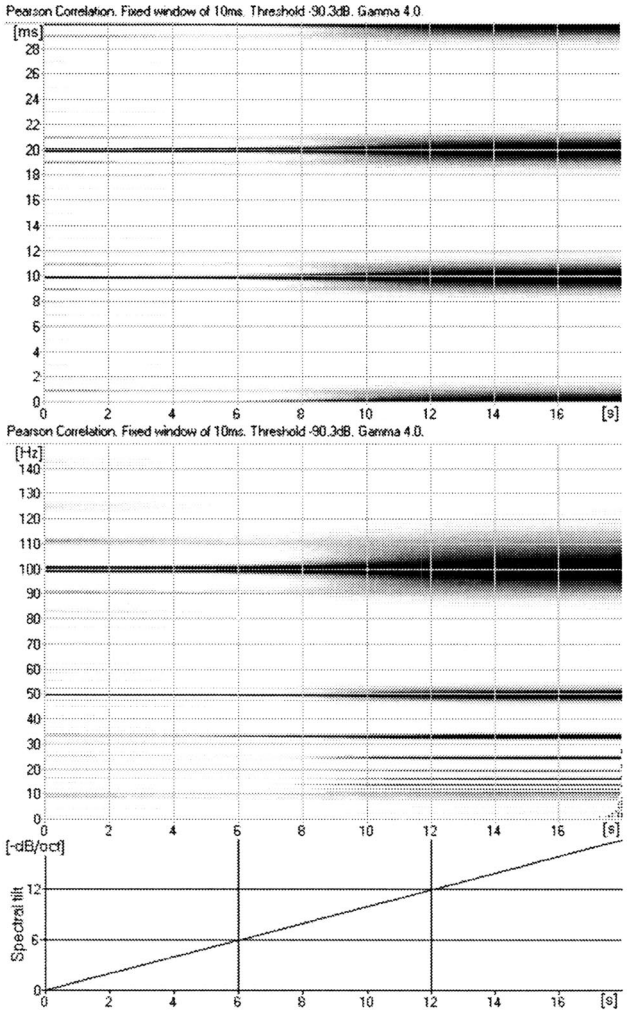


FIG. 5. The effect of spectral tilt on sidebands. The source signal was created by additive synthesis. The spectral tilt was varied from 0 dB/octave at 0 s, to −18 dB/octave at 18 s. The source signal was fed through a formant filter at 1000 Hz, bandwidth 100 Hz.

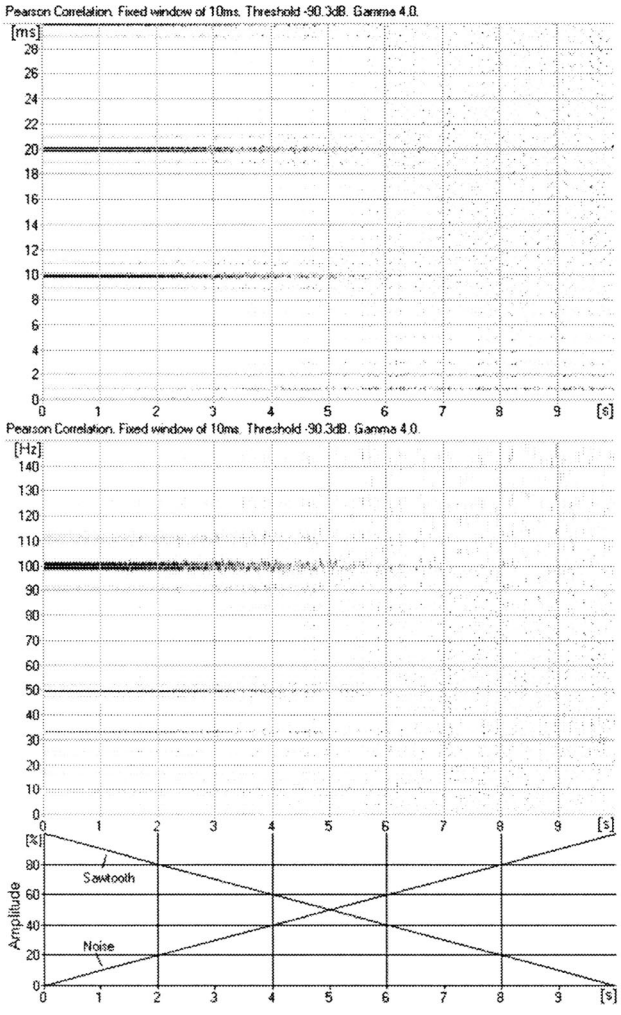


FIG. 6. The effect of adding white noise. At 0 s, the source signal consists of a saw-tooth wave only, and at 10 s of white noise only. The source signal was passed through a formant filter at 1000 Hz, bandwidth 100 Hz.

ficient is calculated with a starting sample *m* corresponding to the time coordinate of the *x* axis. This correlation coefficient is calculated for different delays *n*, along the *y* axis. The correlation, $r_{m,n}$ in this point, is displayed along a gray scale, with black corresponding to $r_{m,n}=1$ and white for $r_{m,n}\leqslant 0$. If the signal is perfectly periodic with a period time T_0, the correlogram will show a set of horizontal black bands, representing different *candidates* for the fundamental period time $C_1, C_2, C_3, \ldots$, at delays *n* corresponding to $T_0, 2T_0, 3T_0, \ldots$ (Fig. 1, middle panel).

Correlograms can also be presented with an inverted *y* axis, thus showing frequency rather than time. In this case F_0 candidates, $C_1, C_2, C_3, \ldots$ will appear at F_0, $F_0/2$, $F_0/3$, and so on (Fig. 1, lower panel). Both these representations have certain advantages. In a time correlogram, the candidates C_n appear as horizontal stripes that are equidistant. It also shows more salient high-order candidates. In a frequency correlogram, the strip density is greater at lower than at higher frequencies, but on the other hand, it is more clearly related to pitch. The selection of delay representation is a matter of choice, but since periodicity is mostly expressed as fundamental frequency, rather than fundamental period time, the frequency representation seems intuitively more appropriate.

The length of the correlation window, *w*, will affect the appearance of the correlogram. A shorter window gives better resolution in time, but may also show the first formant in terms of *sidebands* surrounding each candidate. Normally, a frame length of about one fundamental period is appropriate. An interesting possibility is to let the window length vary as the correlation delay varies, such that the window length is equal to the delay. With this procedure the window length is automatically adjusted to an appropriate value as fundamental frequency varies. The procedure is computationally less efficient, however, especially if correlations are calculated for frequencies close to 0 Hz.

For practical reasons, the correlograms show *r* raised to the power of γ for values of $r>0$. Higher values of γ make the middle levels of gray brighter, and vice versa, see below; $\gamma=4$ has been found to be an appropriate value.

The correlogram allows semimanual extraction of F_0. In this case the user restricts the range of allowed F_0 values to the range around a candidate (Fig. 2). The software then extracts the frequency corresponding to the highest correlation within this range. The manual control allows the user to select the appropriate candidate, thus eliminating the risk that

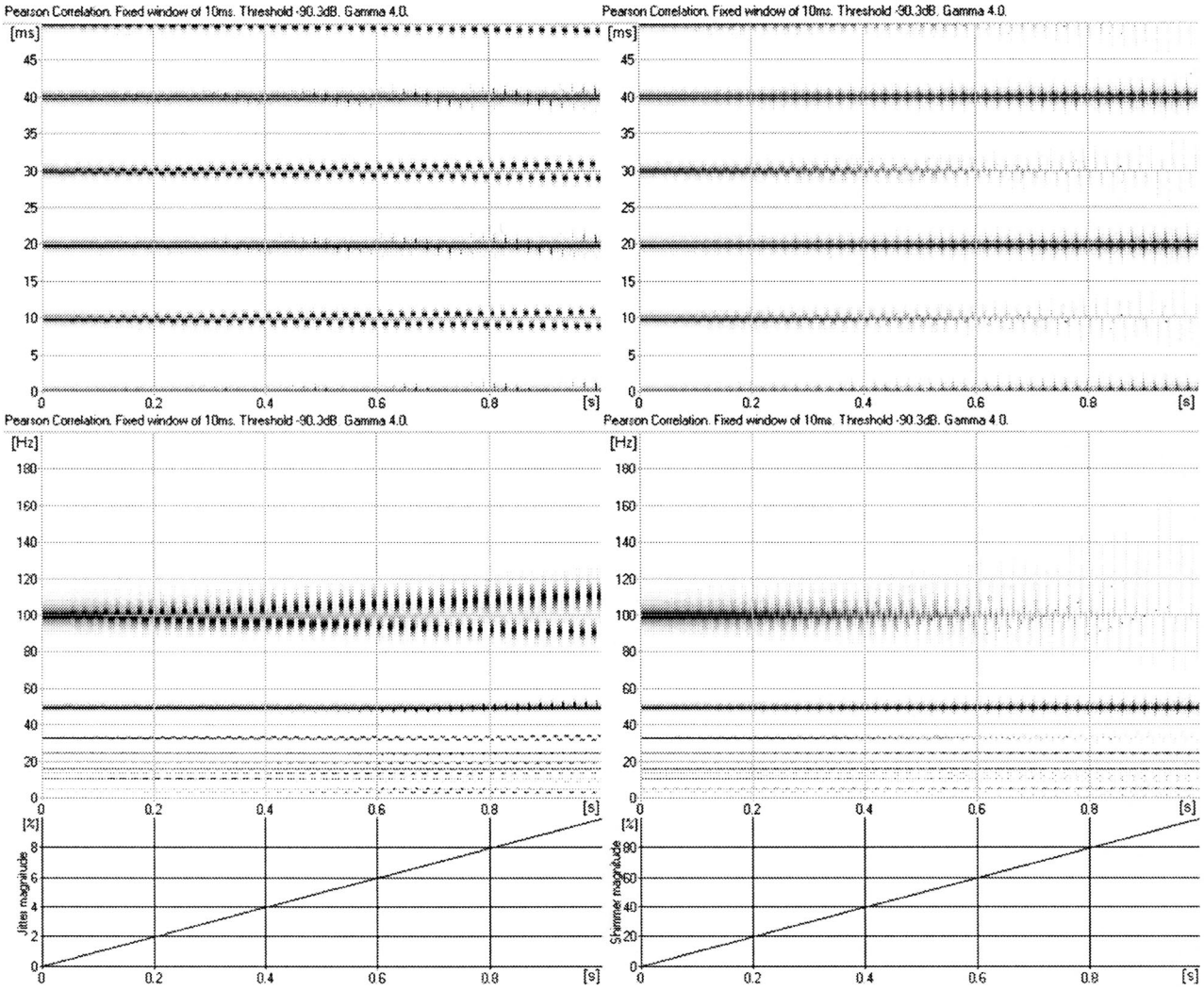


FIG. 7. Examples of bicyclic F_0 (left panels) and amplitude variation (right panels) in saw-tooth waveforms. F_0 and amplitude variations increased from 0% to 10% and from 0% to 100%, respectively.

an automatic algorithm selects a faulty candidate and placing the responsibility on the user.

A. Applications

1. Synthesized signals

The properties of the correlogram analysis method can be efficiently demonstrated by applying it to synthesized sound, since the properties of such sounds are well-defined while natural voices mostly contain combinations of acoustic properties in unknown magnitudes. All synthesized sound files were generated using a sampling rate of 16 kHz.

Figure 3 illustrates sidebands and the effect of the chosen value of γ. The signal was created using a 100-Hz saw-tooth waveform that was fed through a formant filter with a fixed bandwidth of 100 Hz and a resonance frequency, F_1, increasing from 0 to 1000 Hz during the 10-s-long sound file. In the left panels ($\gamma=1$) a segment of a secondary periodicity of 5 ms/200 Hz appears at 2 s, when F_1 equals 200 Hz. As F_1 is increased, this secondary periodicity shifts and new periodicities appear. The amplitude of the sidebands increase when F_1 coincides with a partial. Eventually, the secondary periodicities form continuous sidebands around the candidates. If γ is increased to 4 (right panels) the sidebands are suppressed such that the candidates can more easily be differentiated from the sidebands. The candidates also appear narrower.

Figure 4 illustrates the effects of window length, w, on the frequency correlogram. The signal was a saw-tooth wave of 100 Hz, frequency modulated $\pm 10\%$ at 10 Hz and fed through a formant filter of 1000 Hz, bandwidth 100 Hz. For $w=5$ ms (upper left panel) candidates are present at all times, while the sidebands are intermittent. Time resolution is good. For $w=10$ ms (upper right panel) the sidebands are attenuated and time resolution is still good due to the similarity between T_0 and window length. For $w=20$ ms (lower left panel) the candidates fade at rapid frequency transitions due to poor time resolution, while the sidebands are well suppressed. A window length equal to correlation delay n (e.g., 10 ms at n corresponding to 100 Hz and 20 ms at n corresponding to 50 Hz, etc.) yields a high time resolution, visible, intermittent sidebands at high frequencies, but low time resolution and no sidebands at low frequencies (bottom right panel). This is illustrated in terms of the steps in C_2 overlap and extend over longer time than those pertaining to

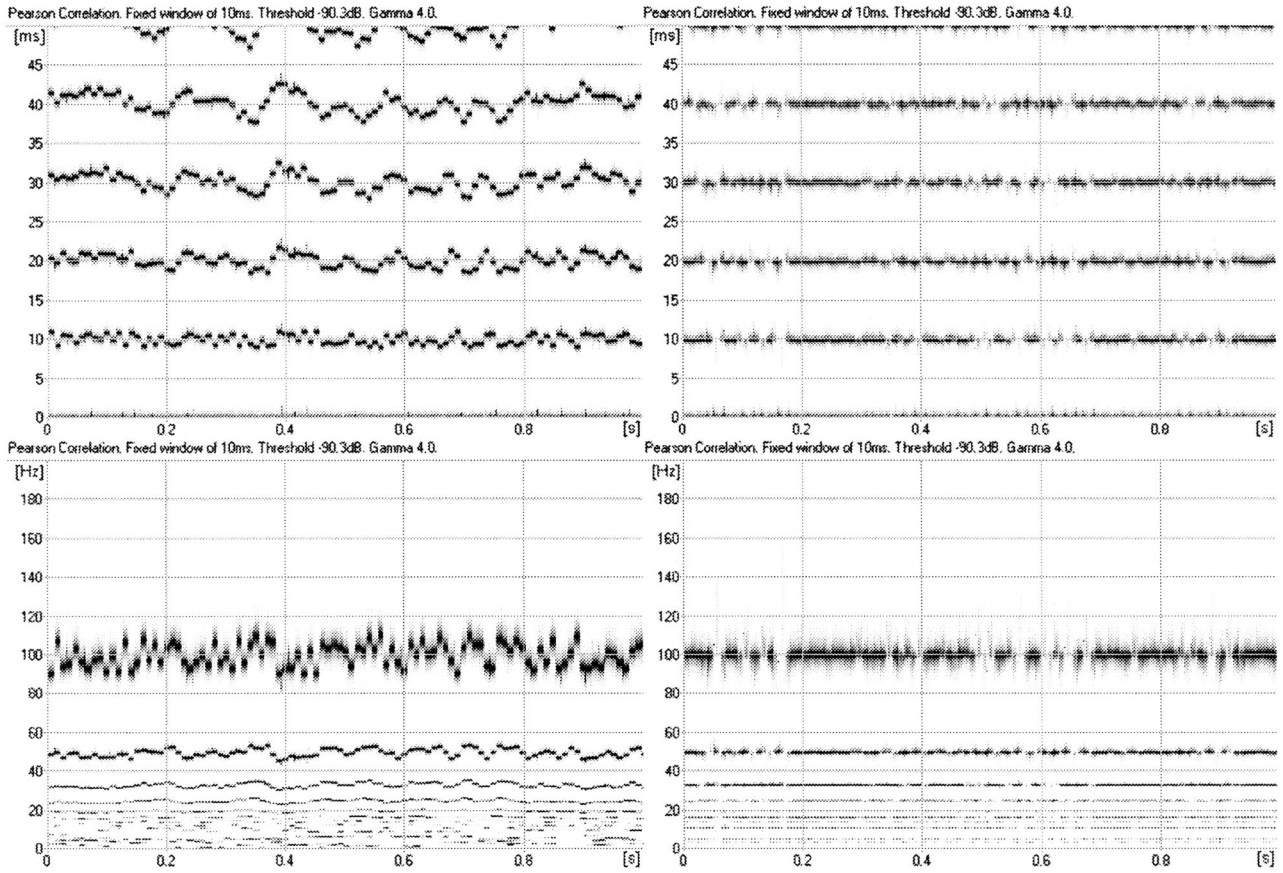


FIG. 8. Examples of random F_0 (left panels) and random amplitude variation (right panels) in saw-tooth waveforms. F_0 and amplitude variations was restricted to $\pm 10\%$ and $\pm 100\%$, respectively.

C_1. As seen in the figure, a window length approximating T_0 appears appropriate. In cases of completely unknown T_0, however, a window length equal to correlation delay might be preferable. In this case, C_1 will always be analyzed with a window of length T_0.

Figure 5 illustrates the effect of spectral tilt on sidebands. The signal, $F_0 = 100$ Hz, was created by adding sinusoids according to a spectral tilt, which was continuously varied at a constant rate from 0 to -18 dB/octave over an 18-s-long sound file. This signal was fed through a formant filter at 1000 Hz, bandwidth 100 Hz. It can be expected that strong harmonics in the formant region will add extra periodicities that are visible in the correlogram. Correspondingly, if the signal only contains the fundamental, this is the only periodicity that will be visible. These effects can be seen in the figure, and we will describe the two cases that occur at 0 and 18 s, in which the partials at F_1 are strong and weak, respectively. A tilt of 0 dB/octave ($t = 0$ s) produces pronounced sidebands. The *absolute* width of the candidates is largely determined by the distance between the candidate and the sideband. This distance is a direct function of F_1, and thus the candidate width is largely determined by F_1. In other terms, the *relative* width is determined by the frequency ratio F_1/F_0. On the other hand, at -18 dB/octave ($t = 18$ s) when the signal is dominated by the fundamental, the absolute width is mainly determined by F_0, and thus the relative width is nearly independent of both F_0 and F_1. The width's dependence on F_1 can also be seen in Fig. 3, where the candidates become narrower for higher values of F_1. It should be noted that it is not the increasing spectral tilt *per se* that makes the sidebands fade away, but rather the fact that the level of the partial at the formant is reduced. As a rule of thumb, sidebands appear if the spectral level at the first formant is near or above the level of the fundamental. However, this is also slightly dependent on the chosen value of γ.

Figure 6 illustrates the effect of adding noise to a periodic signal. At 0 s, the signal, $F_0 = 100$ Hz, is a saw-tooth waveform, and at 10 s it consists of white noise only. This signal was fed through a 1000-Hz formant filter, bandwidth 100 Hz. The levels of the saw-tooth waveform and the noise were matched, so that the output level of the formant filter was equal at the start and end of the sound file. This means that the harmonics-to-noise ratio (HNR) was infinite at 0 s, 1 (0 dB) at 5 s, and 0 at 10 s. The effect of the noise on the correlogram is noticeable at about 2 s, corresponding to an HNR of 4 (12 dB). At about 5 s, or HNR$=1$ (0 dB), the candidates more or less disappear, the only visible periodicity appearing at 1 ms and created by the F_1 at 1000 Hz.

Figure 7 compares time and frequency correlograms of synthesized saw-tooth waveforms that contain bicyclic F_0 or amplitude variation; this can be seen as a special case of jitter or shimmer. In the jitter case, the period time (mean 10 ms) was varied every other period, starting at 0% and ending at $\pm 10\%$. The jitter (left panels) can be seen as an F_0 fluctuation, the first candidate C_1 splitting into two "stripes" that reach 91 to 111 Hz at the end of the frequency correlogram

and 11 to 9 ms at the end of the time correlogram. In the shimmer case (right panels), the amplitude of the periods was varied every other period. The magnitude of the shimmer varies from 0% at the start to ±100% at the end, that is, at the end, every second period has an amplitude of zero, while the intermediate periods have an amplitude twice the original. In this case, the first candidate also shows an oscillating pattern, distinct however, from that characterizing jitter. The odd-order candidates gradually fade as the shimmer quantity increases. The shimmer magnitude is seen less clearly than the jitter magnitude since the former is reflected in terms of the gray scale while the latter is represented by the position along the *y* axis.

Figure 8 compares time and frequency correlograms of synthesized saw-tooth waveforms that contain random F_0 or amplitude variation; these represent other types of jitter and shimmer. In the jitter case, the period time (mean 10 ms) was randomly distributed within ±10%. The jitter (left panels) can be seen as a random F_0 fluctuation, the first candidate C_1 fluctuating in the range 91 to 111 Hz in the frequency correlogram and 11 to 9 ms in the time correlogram. In the shimmer case (right panels), the amplitude of each period was varied randomly. The magnitude of changes was 100%; in other words, the amplitude was randomly distributed between zero and full scale. In this case, the first candidate also shows a fluctuating pattern, again distinct from that characterizing jitter.

Although the time correlogram is directly linked to the waveform-matching function, we shall henceforth focus on frequency correlograms.

2. Natural voices

Some examples of correlograms and narrow-band spectrograms of pathological voices are presented in Figs. 9–15. All these figures concern examples of voices that may cause difficulties for F_0 tracking programs. The difficulties are due to ambiguity about whether F_0 is represented by C_1 or by C_2 (Figs. 9, 10, and 11), due to a high noise level and unstable C_1 (Figs. 12, 13, and 14), or due to the well-excited first formant, which makes the sidebands hard to differentiate from C_1 (Fig. 15).

For describing the voices, the terminology proposed by Hammarberg and Gauffin (1995) was used; all voice examples except one (Fig. 15) were taken from Hammarberg's library of pathological voices. The voice samples had been rated by groups of 6 to 14 voice clinicians (speech language pathologists and phoniatricians) using the Stockholm Voice Evaluation Approach (SVEA) assessment protocols (Hammarberg, 1986, 2000). These examples have been found useful in teaching as archetypes of various perceived voice properties, as in each example a particular voice quality is dominant over others. However, since all natural voices are perceptually multidimensional (Kreiman *et al.*, 1994, 1996), each example still represents more than one single perceptual feature. This well-known fact makes direct mapping of acoustic to perceptual features difficult.

Figure 9 presents the voice of a man, age 41, who was diagnosed with chronic laryngitis and whose voice quality was characterized as rough. This example shows short (about

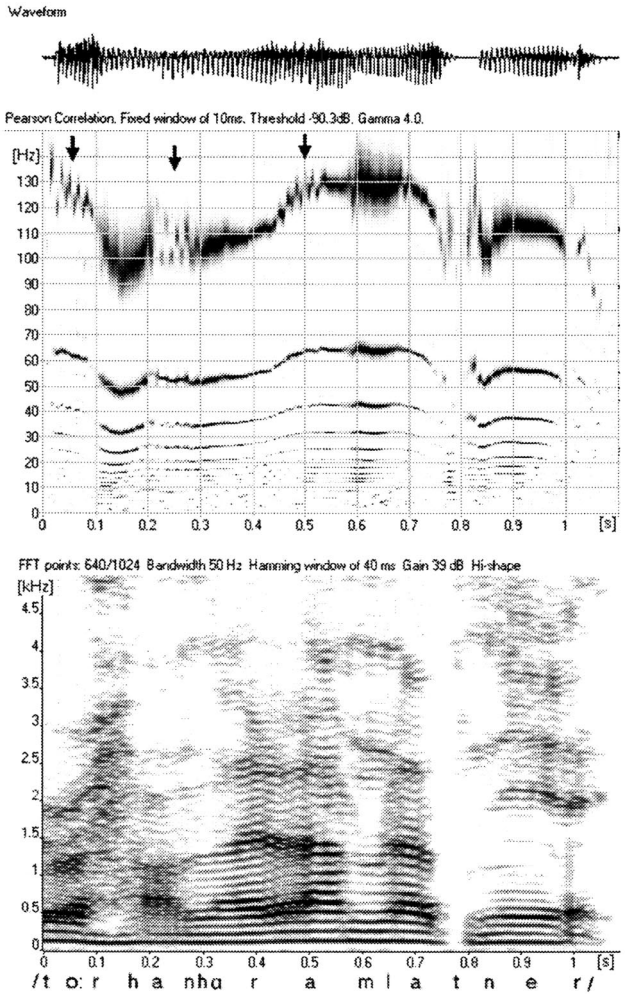


FIG. 9. Waveform (top), correlogram (middle), and narrow-band spectrogram (bottom) of speech. The voice was characterized by roughness. Bicyclic segments appear around 0.05, 0.25, and 0.5 s. At about 0.8 s, there is a lack of periodicity due to the voiceless consonant /t/.

100 ms) bursts of bicyclicity starting at 0, 0.2, and 0.45 s. Widening of the candidates can be seen at 0.1–0.2 s, 0.6–0.7 s, and 0.85–1 s. The two former cases of widening are probably due to the low amplitude of the overtones in the signal, and the last is probably due to a low first formant. Figure 10 presents the voice of a man, age 32, who was diagnosed with incomplete voice mutation and whose voice quality was characterized as a mixture between vocal fry and gratings/scrape. This example contains bicyclicity throughout most of its duration. Figure 11 presents the voice of a man, age 29, who was diagnosed with a benign tumor, perceived as having gratings/scrape only.

These examples all show short (100 ms) bursts of bicyclicity, and the only obvious difference among them is the F_0 at which they occur. In Fig. 10 there also is a longer (about 250 ms) bicyclic sequence. Careful inspection of the spectrograms (lower panel) reveals subharmonics coinciding with the bicyclic segments in the correlograms but with poorer time resolution. The subharmonics could be visualized more clearly in the spectrogram if a narrower bandwidth had been chosen. This would, however, further deteriorate the time resolution.

Figure 12 presents the voice of a man, age 50, who was

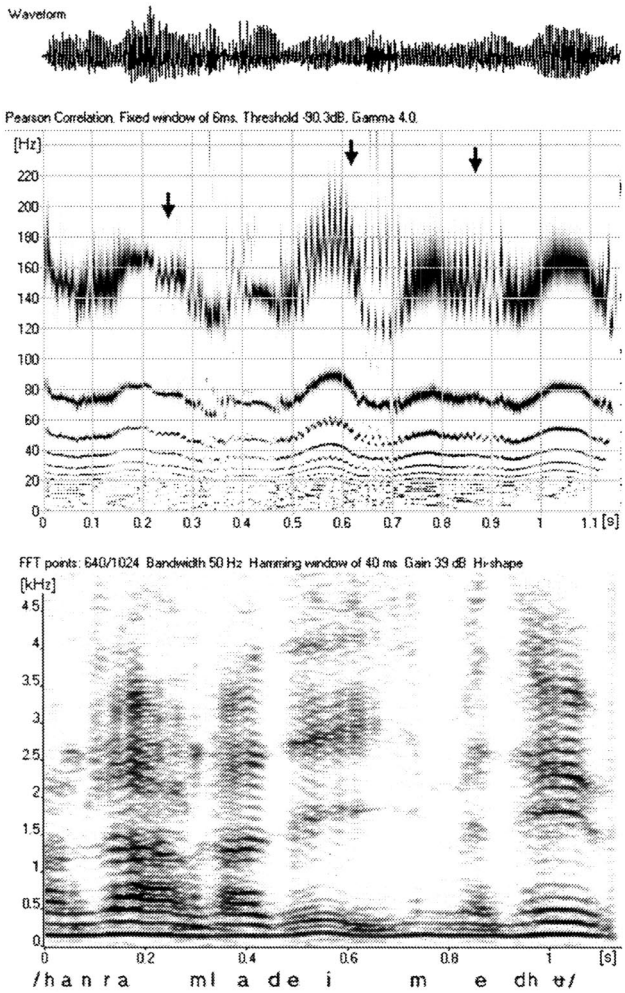


FIG. 10. Waveform (top), correlogram (middle), and narrow-band spectrogram (bottom) of speech. The voice was characterized by vocal fry and gratings/scrape. Bicyclic segments appear around 0.25 s, at 0.5–0.75 s, and at 0.8–0.95 s.

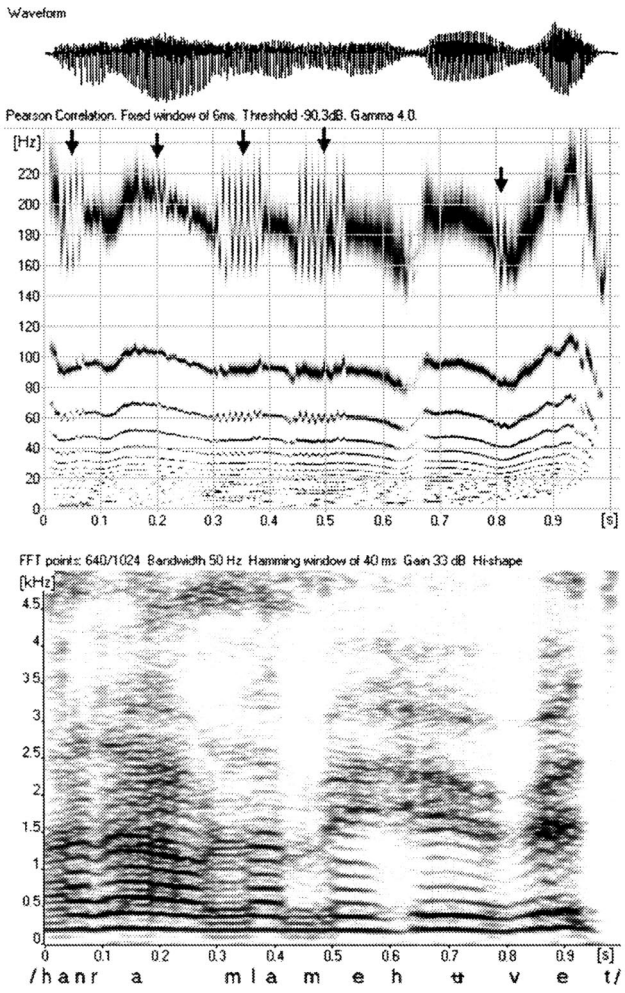


FIG. 11. Waveform (top), correlogram (middle), and narrow-band spectrogram (bottom) of speech. The voice was characterized by gratings/scrape. Bicyclic segments appear around 0.05, 0.35, and 0.5 s and also with less magnitude around 0.2 and 0.8 s. Note the abnormally high F_0.

diagnosed with paralytic dysphonia, and whose voice quality was characterized by hypofunctional breathiness with roughness. The voice produces wide and unstable candidates. As sometimes also found in rough voices, an instant of bicyclicity occurs, near $t=0.7$ s. However, the corresponding subharmonics in the spectrogram (lower panel) are not easily spotted, probably due to the short duration of the bicyclicity.

Figure 13 presents the voice of a man, age 40, who was diagnosed with paralytic dysphonia; the voice was characterized by hypofunctional breathiness. In the narrow-band spectrogram few harmonics except the fundamental are visible. The correlogram shows wide candidates and no bicyclicity.

Figure 14 presents the voice of a woman, age 75, who was diagnosed with paralytic dysphonia which shows repeated voice breaks between falsetto and modal register with a high degree of instability (Hammarberg, 1986). C_1 suddenly disappears at $t=0.5$ and 1.35 s as the voice switches from falsetto to modal.

Finally, Fig. 15 presents the voice of an opera singer. Sidebands are prominent, indicating a well-excited first formant. As also can be seen in the spectrogram, the singer apparently tuned F_1 to either two or three times F_0, such that either the second or third partial coincides with F_1. This strategy increases the sound-pressure level, which is an important ability in operatic singing.

III. DISCUSSION

Since F_0 fluctuation plays an important role in many different pathological voice qualities, F_0 extraction would be one way to study such voice qualities. Unfortunately, F_0 extraction applied to voices with a high degree of F_0 perturbation presents problems that are not easily solved. The most typical example is bicyclic voice, where F_0 extractors tend to yield $F_0/2$. Since the transition from normal phonation to bicyclicity can be gradual, although without a pitch glide, an F_0 extraction algorithm must determine when to switch from displaying F_0 to displaying $F_0/2$. Such switching results in an octave leap. In the correlogram, this problem is circumvented by eliminating the selection mechanism and displaying raw correlation functions in a three-dimensional graph. Hence, the correlogram can describe highly perturbed voices, even when the value of F_0 is far from obvious.

The appropriateness of extracting F_0 from pathological voices can sometimes be questioned. Pathological voices often show large period-to-period variation, and since the sig-

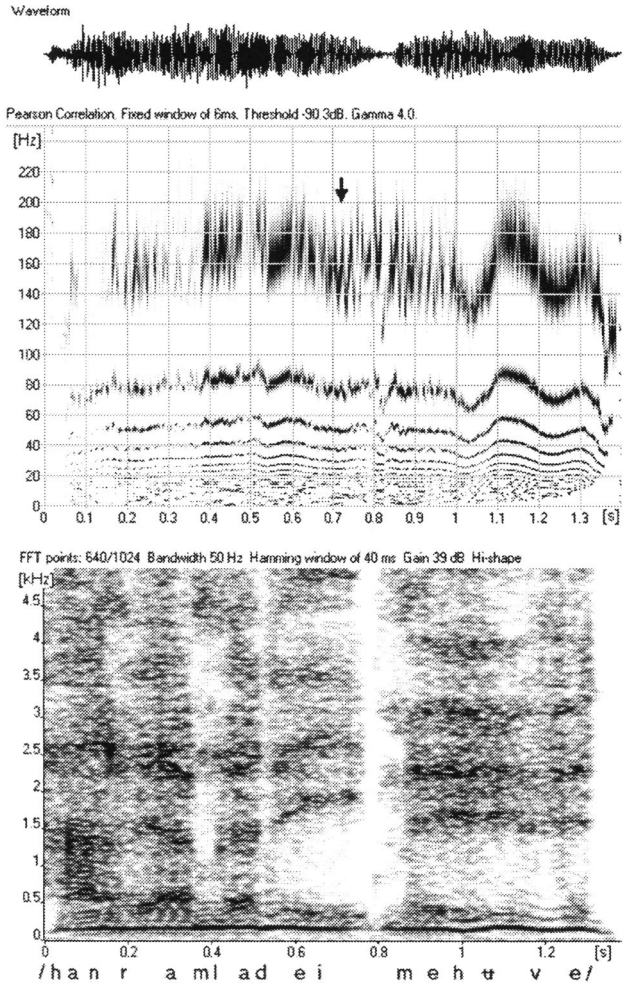


FIG. 12. Waveform (top), correlogram (middle), and narrow-band spectrogram (bottom) of speech. The voice was characterized by hypofunctional breathiness with roughness. The candidates all are unstable and wide, and a short instance of bicyclicity can be seen at around 0.7 s.

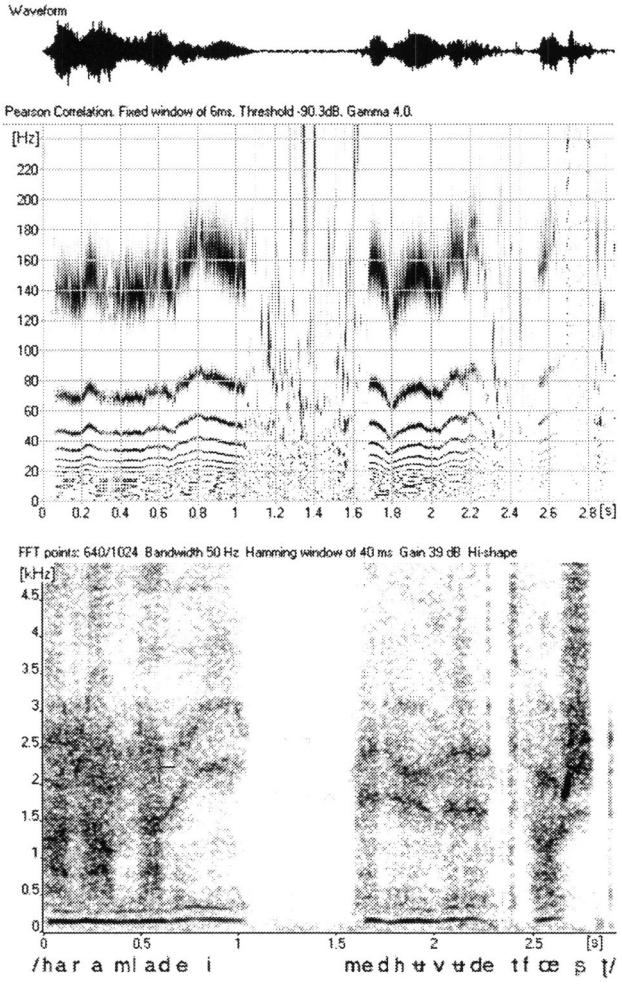


FIG. 13. Waveform (top), correlogram (middle), and narrow-band spectrogram (bottom) of speech. The voice was characterized by hypofunctional breathiness. All candidates are wide due to the dominant fundamental. The segment between 1.0 and 1.7 s represents silence.

nal is not exactly repetitive, no strict period time exists; or, an ambiguity exists with regard to F_0. For such voices, it appears appropriate not to enforce an F_0 selection, but rather to display the correlation functions, as in a correlogram. In some cases a perceptual evaluation of pitch may also be a worthwhile alternative.

The interpretation of a correlogram requires some care, however. While the F_0 candidates appear as dark horizontal bands, there can also be sidebands, which originate from formant ringings of high amplitudes and are thus not to be considered as candidates. Sidebands appear when an overtone coincides with a formant, and can mostly be identified from their relatively less dark appearance (i.e., low correlation). In some cases, such as when F_1 is twice F_0, the distinction between sidebands and candidates can be less clear.

The presence of sidebands can also be used as an indication of a high positive level difference between the first formant (L_1) and the fundamental (L_0). Hence, the presence of sidebands may indicate a sonorous or pressed voice with a well-excited first formant. On the other hand, a low or negative L_1-L_0 difference has been shown to be related to hypofunctional breathiness (Hammarberg, 1986). In the correlogram this would correspond to a wide candidate. It must be kept in mind, however, that the candidate width also depends on the formant frequencies.

A correlogram is a time-domain analysis tool. This means that it does not directly display spectral properties, such as harmonics, which would require a Fourier transform. It should be noted that the candidates have no direct connection to the harmonics of the signal. It is true that C_1 corresponds to the first harmonic, H_1, but the presence of C_2 does not necessarily indicate the presence of a subharmonic. However, the combined occurrence of a constant C_2 and a varying C_1 would indicate bicyclicity, and a constant C_3 and a varying C_2 and C_1 tricyclicity, etc. These characteristics indicate the presence of subharmonics, although the subharmonics *per se* are not visualized in a correlogram.

Compared to the narrow-band spectrogram, the correlogram shows a better time resolution, due to the shorter time windows needed. For instance, to display a narrow-band spectrogram with visible subharmonics, the length of the time window must correspond to several fundamental periods, whereas in the correlogram, the time windows typically are as short as one fundamental period. The short time windows have the effect that the correlogram can visualize short

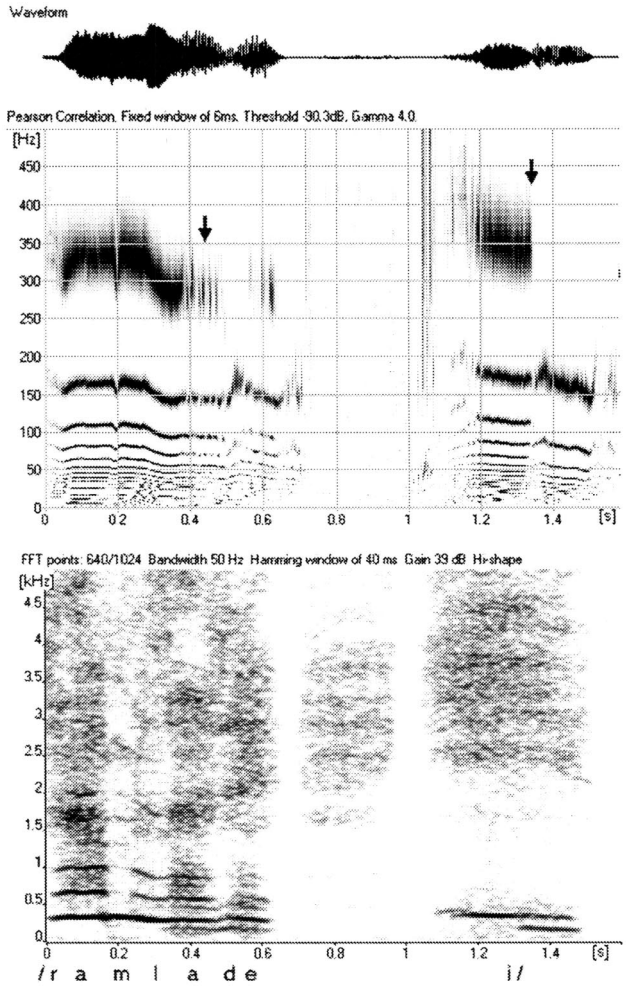


FIG. 14. Waveform (top), correlogram (middle), and narrow-band spectrogram (bottom) of speech. The voice was characterized by repeated register breaks. At 0.4 and 1.35 s the voice switched from falsetto to modal. The noise between 0.7 and 0.9 s is inhalatory stridor.

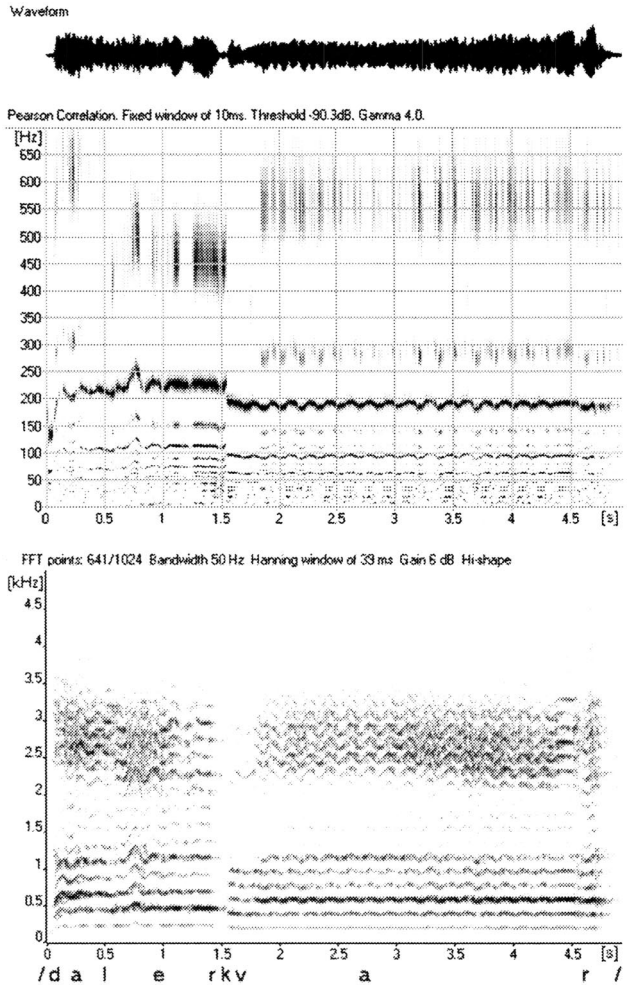


FIG. 15. Waveform (top), correlogram (middle), and narrow-band spectrogram (bottom) of singing, operatic style. Prominent sidebands can be seen either at two or three times F_0 due to a well-excited first formant.

bursts of bicyclicity that would not be easily seen in a narrow-band spectrogram.

The correlogram has also been used for extraction of F_0 from violin playing (Gleiser *et al.*, 1998) by means of manual tracing. In these experiments, the violin player was accompanied by piano playing, which, however, was suppressed by placing the microphone on the violin bridge. Violin sound typically presents difficulties in F_0 extraction. However, the correlogram method was surprisingly successful and showed a remarkable insensitivity to the piano sound. In this study, the vibrato rate was also extracted in a second step, by performing correlogram analysis and tracing on the extracted oscillating F_0 curve.

The computation of a correlogram is generally more computationally intensive than the computation of a spectrogram. However, with the increasing power of computers, the computation speed is less of a problem. For example, every correlogram presented above required less than 3 s computing time on a 1700-MHz Pentium 4 system running WINDOWS 2000.

These initial applications suggest that the correlogram is useful for future work for refining, revising, and standardizing the relations between acoustical voice characteristics and perceived voice quality parameters. Correlograms should be useful also for the training of an analytic listening to voice qualities. Presenting images of the perturbation of voices together with the sounds seems a valuable opportunity that may pave the way to a better agreement on the meaning of voice terms across the voice community.

The correlogram illustrates the periodicity of the waveform in a robust way, since it lacks the selection mechanism of F_0 extractors. It illustrates differences between periodic and random period-to-period variations. The robustness of the correlogram should make it a particularly valuable tool for periodicity analysis in such cases of pathologic speech where standard F_0 extraction methods fail or where they present ambiguous results.

IV. CONCLUSIONS

Correlation functions have previously been used to extract F_0 information from voice signals, automatically selecting a single value to represent F_0, sometimes even in ambiguous cases. The correlogram method presented here shows the raw correlation functions. In cases of periodic or quasiperiodic phonation, such as in some pathological voices, it displays several F_0 candidates, and leaves the user

to select one by tracing, if appropriate. In some cases of quasiperiodic phonation, the correlogram illustrates the type of aperiodicity, differentiating signal characteristics such as multicyclic or random perturbations, typically associated with vocal fry or roughness. It should be worthwhile to test the correlogram in cases of quasiperiodic signals where traditional F_0 tracking methods fail.

ACKNOWLEDGMENTS

This work was supported by research grants from the Bank of Sweden Tercentenary Foundation and the Swedish Council for Work Life Research. We would also like to thank Jan Gauffin and Stellan Hertegård for valuable advice and Johan Sundberg for discussions and editorial assistance.

Blomgren, M., Chen, Y., Ng, M., and Gilbert, H. (**1998**). ''Acoustic, aerodynamic, physiologic and perceptual properties of modal and vocal fry registers,'' J. Acoust. Soc. Am. **103**, 2649–2658.

DeKrom, G. (**1995**). ''Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments,'' J. Speech Hear. Res. **38**, 794–811.

Fourcin, A. (**1986**). ''Electrolaryngographic assessment of vocal fold vibration,'' J. Phonetics **14**, 435–442.

Gauffin, J., Granqvist, S., Hammarberg, B., Hertegård, S., and Håkansson, A. (**1995**). ''Irregularities in the voice: Some perceptual experiments using synthetic voices,'' Proc. ICPhS-95, Vol. 2, pp. 242–245.

Gleiser, J., Friberg, A., and Granqvist, S. (**1998**). ''A method for extracting vibrato parameters applied to violin performance,'' TMH-QPSR 4/1998, 39–44.

Hammarberg, B. (**1986**). ''Perceptual and acoustic analysis of dysphonia,'' published doctoral thesis, Dept. of Logopedics and Phoniatrics, Karolinska Institute, Stockholm.

Hammarberg, B. (**2000**). ''Voice research and clinical needs,'' Folia Phoniatr. Logop. **52**, 93–102.

Hammarberg, B., and Gauffin, J. (**1995**). ''Perceptual and acoustical characteristics of quality differences in pathological voices as related to physiological aspects,'' in *Vocal Fold Physiology, Voice Quality Control*, edited by O. Fujimura and M. Hirano (Singular, San Diego), pp. 283–303.

Hess, W. (**1983**). *Pitch Determination of Speech Signals* (Springer, Berlin).

Hess, W. (**1995**). ''Determination of glottal excitation cycles in running speech,'' Phonetica **52**, 196–204.

Hillenbrand, J. (**1988**). ''Perception of aperiodicities in synthetically generated vowels,'' J. Acoust. Soc. Am. **83**, 2361–2371.

Imaizumi, S. (**1986**). ''Acoustic measures of roughness in pathological voice,'' J. Phonetics **14**, 457–462.

Ishiki, N., Okamura, H., Tanabe, M., and Morimoto, M. (**1969**). ''Differential diagnosis of hoarseness,'' Folia Phoniatr. **21**, 9–19.

Karnell, M., Scherer, R., and Fischer, L. (**1991**). ''Comparison of acoustic voice perturbation measures among three independent voice laboratories,'' J. Speech Hear. Res. **34**, 781–790.

Kreimann, J., Gerrat, B., and Berke, G. (**1994**). ''The multidimensional nature of pathologic voice quality,'' J. Acoust. Soc. Am. **96**, 1291–1302.

Kreimann, J., and Gerratt, B. R. (**1996**). ''The perceptual structure of pathologic voice quality,'' J. Acoust. Soc. Am. **100**, 1787–1795.

Ladefoged, P. (**1988**). ''Discussion of phonetics: A note on some terms for phonation types,'' in *Vocal Physiology: Voice Production, Mechanisms and Functions*, edited by O. Fujimura (Raven, New York), pp. 373–375.

Laver, J. (**1980**). *The Phonetic Description of Voice Quality* (Cambridge University Press, Cambridge).

Omori, K., Kojima, H., Kakani, R., Slavit, D., and Blaugrund, S. (**1997**). ''Acoustic characteristics of rough voice: Subharmonics,'' J. Voice **11**, 40–47.

Rabinov, R., Kreiman, J., Gerratt, B., and Bielamowicz, S. (**1995**). ''Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter,'' J. Speech Hear. Res. **38**, 26–32.

Rothenberg, M. (**1973**). ''A new inverse filtering technique for deriving the glottal airflow waveform during voicing,'' J. Acoust. Soc. Am. **53**, 1632–1645.

Švec, J., and Pešák, J. (**1994**). ''Vocal breaks from the modal to falsetto register,'' Folia Phoniatr. Logop. **46**, 97–103.

Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois University Press, Dekalb, IL).

Titze, I. (**1995**). ''Definitions and nomenclature related to voice quality,'' in *Vocal Fold Physiology, Voice Quality Control*, edited by O. Fujimura and M. Hirano (Singular, San Diego), pp. 335–342.

Titze, I., and Liang, H. (**1993**). ''Comparison of F_0 extraction methods for high-precision voice perturbation measurements,'' J. Speech Hear. Res. **36**, 1120–1133.

The dependency of timbre on fundamental frequency[a)]

Jeremy Marozeau,[b)] Alain de Cheveigné,[c)] Stephen McAdams, and Suzanne Winsberg
Institut de Recherche et Coordination Acoustique/Musique (Ircam-CNRS), 1, place Igor Stravinsky, F-75004 Paris, France

(Received 15 August 2002; revised 25 June 2003; accepted 25 August 2003)

The dependency of the timbre of musical sounds on their fundamental frequency (F_0) was examined in three experiments. In experiment I subjects compared the timbres of stimuli produced by a set of 12 musical instruments with equal F_0, duration, and loudness. There were three sessions, each at a different F_0. In experiment II the same stimuli were rearranged in pairs, each with the same difference in F_0, and subjects had to ignore the constant difference in pitch. In experiment III, instruments were paired both with and without an F_0 difference within the same session, and subjects had to ignore the variable differences in pitch. Experiment I yielded dissimilarity matrices that were similar at different F_0's, suggesting that instruments kept their relative positions within timbre space. Experiment II found that subjects were able to ignore the salient pitch difference while rating timbre dissimilarity. Dissimilarity matrices were symmetrical, suggesting further that the absolute displacement of the set of instruments within timbre space was small. Experiment III extended this result to the case where the pitch difference varied from trial to trial. Multidimensional scaling (MDS) of dissimilarity scores produced solutions (timbre spaces) that varied little across conditions and experiments. MDS solutions were used to test the validity of signal-based predictors of timbre, and in particular their stability as a function of F_0. Taken together, the results suggest that timbre differences are perceived independently from differences of pitch, at least for F_0 differences smaller than an octave. Timbre differences can be measured between stimuli with different F_0's.
 [DOI: 10.1121/1.1618239]

PACS numbers: 43.75.Cd, 43.66.Jh, 43.66.Hg [NFV] Pages: 2946–2957

I. INTRODUCTION

The word ''timbre'' has several meanings. In a musical context it designates aspects of sound that allow an instrument to be identified and distinguished from others. In the context of psychoacoustic experiments, it designates an elementary sound quality akin to pitch or loudness (the ''Klangfarbe'' of Helmholtz, 1885). In the next paragraph we shall use the words ''identity'' and ''quality'' to distinguish these two meanings, respectively. The identity of a musical instrument obviously depends in some way on the quality of the sounds it produces (their ''timbre'' in a psychoacoustic sense). However, this dependency may be complex.

For certain instruments, quality varies as a function of the note played, the intensity at which it is played, and time. This is obvious from casual listening, and corroborated by measurements or calculations that show variations of signal properties that are known to affect sound quality (spectral centroid, harmonicity, etc.) (Martin, 1999). For example, notes of the trumpet become brighter with increased intensity (Luce and Clark, 1967), while those of the violin are subject to complex interactions between body resonances and the harmonic spectrum of string vibration (Fletcher and Rossing, 1998). The latter varies with fundamental frequency (F_0) and thus with the note played. The timbre of a wind instrument may change abruptly between the low register (with the register hole closed) and the high register (with the hole open), a characteristic revealed only if the instrument is played over a range of notes (Risset and Wessel, 1999). Sound qualities produced by a particular instrument follow a particular ''trajectory,'' and indeed we could formulate the hypothesis that this in part determines its identity. In other words, timbre (identity) might depend on the *pattern of variation* of timbre (quality) specific to an instrument. To test such a hypothesis experimentally requires comparing timbre (quality) across time, intensity, or F_0. The purpose of the present study was to characterize variations of timbre (quality) as a function of F_0.

The standard methodology for studying timbre is multidimensional scaling (MDS) (Grey, 1977). Typically, subjects are presented with pairs of sounds and asked to rate their dissimilarity on a continuous scale. Dissimilarity scores are processed by an MDS algorithm to produce models of ''timbre space'' that give insight into the nature of the timbre percept. It is usually found that the timbre space involved in a task is of small dimensionality (two to four dimensions), that different subjects may weight dimensions differently, and that these dimensions can usually be predicted by signal-based ''descriptors.'' The relevant dimensions (and corresponding descriptors) tend to vary between experiments, no doubt as a function of the set of sounds included in each experiment. Nevertheless certain dimensions (e.g., ''brightness,'' predicted by a ''spectral centroid'' descriptor) tend to recur in all. MDS seems the appropriate tool to study variations of timbre with F_0.

There are potential problems however. A difference in

[a)]Portions of these results were presented at the 141st meeting of the Acoustical Society of America.
[b)]Electronic mail: jeremy.marozeau@ircam.fr
[c)]Electronic mail: alain.de.cheveigne@ircam.fr

F_0 produces a difference in *pitch* that adds to the dissimilarity between sounds. Even if this extra term is constant, its contribution sets a lower limit to every dissimilarity score, and so the method could be insensitive to small variations in timbre. Worse, if the F_0-dependent term varies, these variations would be confounded with variations of timbre-related dissimilarity and affect the validity of cross-F_0 timbre comparisons. Past studies that allowed F_0 to vary generally found that pitch dominated dissimilarity at the expense of timbre (Miller and Carterette, 1975), with the result that MDS solutions were relatively insensitive to F_0-induced variations of timbre.

One solution is to instruct subjects to *ignore* pitch when making timbre judgments. Unfortunately we are not *a priori* certain that they can do so. Pitch and timbre might not be *separable*, that is, timbre comparisons may be possible between sounds with the same pitch, but not between sounds with different pitch. This worry is reinforced by the scarcity of F_0-dependent timbre studies in the past. A first aim of our study was to determine whether subjects can reliably make cross-F_0 comparisons of timbre while ignoring differences in pitch.

If they can, we may hope to bring an empirical answer to a question such as: how does timbre change with F_0? Two sorts of change are to be expected: first, instrument-specific changes such as evoked earlier, for example due to changes of resonator geometry as a function of the note played, and second, hypothetical changes of a more basic nature, due to a perceptual interaction between pitch and timbre, or the presence of F_0 as a cofactor in the relation between signal descriptors and psychophysical dimensions of timbre. A second aim of our study was to measure F_0-dependent timbre changes, in particular of a basic, noninstrument-specific nature.

There are reasons to expect interactions between pitch and timbre. While pitch is defined as ''that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high'' (ANSI, 1960), more complex structures have been proposed such as a spiral involving both a linear dimension of tone height and a circular dimension of chroma (Shepard, 1964; Ueda and Nimmo-Smith, 1987). Chroma is related to fundamental periodicity, while tone height depends more on the spectral envelope (Patterson *et al.*, 1993). The envelope also determines timbre, so it could be that timbre and pitch are not entirely distinct. This might result in nonseparability (if a pitch difference degrades comparisons between timbre) or a systematic shift (if pitch and timbre are partly colinear).

To a first approximation, the spectral envelope of a vowel does not change with variations of F_0, and vowel identity (another usage of ''timbre'') is likewise relatively invariant. Small systematic variations have nevertheless been observed (see de Cheveigné and Kawahara 1999, for a review). Slawson (1968) asked subjects to adjust the formant frequencies of different-F_0 vowels so that they had the same timbre. The best match was obtained for a 10% increase of formant frequencies for a one-octave increase in F_0. This suggests that envelope-related dimensions of timbre might depend on F_0 in addition to their dependency on envelope characteristics. In other words, F_0 might need to be included as a cofactor in the formulas of signal-based descriptors that predict those dimensions.

A third aim of our study was to test the validity of signal-based descriptors across F_0. Signal-based measures that correlate well with perceptual dimensions revealed in MDS studies (such as spectral centroid, log attack time or spectral flux) have been proposed as ''descriptors'' for applications such as the retrieval of multimedia data (Misdariis *et al.*, 1998; Peeters *et al.*, 2000). Such applications involve data at a wide range of F_0's, yet these descriptors have been tested only with a restricted set of F_0's (often only one). There is clearly a need to verify their generality, and if necessary to modify them to improve their generality. This might entail adjustment of the formulas to remove a spurious F_0 dependency, or inclusion of an F_0-dependent corrective term or, in the extreme, establishment of an array of F_0-dependent formulas.

It is worth discussing the forms of dependency of timbre on F_0 that we expect to find. Supposing a ''timbre space'' such as revealed in MDS studies, three hypotheses can be distinguished: (1) invariance of instrument positions with changes in F_0, (2) isometric displacement keeping relative positions invariant, and (3) non-isometric displacement.

According to hypothesis (1), variations of timbre with F_0 are negligible compared for example to between-instrument differences. Hypothesis (2) allows for a rotation or drift in timbre space common to all instruments. Hypothesis (3) allows that timbres of individual instruments change in arbitrary ways. The experiments were designed to decide between these hypotheses.

We used recordings of natural musical instrument sounds as stimuli. By doing so we confounded two sorts of F_0-dependent timbre changes: those specific to instruments, and those of a non-instrument-specific nature. We reasoned that natural instrumental sounds would guarantee the musical relevance of our sampling, while instrument-specific effects could be interpreted by a *posthoc* analysis of the waveforms of the stimuli.

II. EXPERIMENTS

A. Experiment I

Experiment I consisted of three sessions labeled a, b, and c. In each, subjects rated the dissimilarity between stimuli with the same F_0. This F_0 varied from session to session.

1. Methods

a. Stimuli. Ten natural and two synthetic instruments were used. Each instrument was played at three notes: B3 (247 Hz), C♯4 (277 Hz) and Bb4 (466 Hz), chosen to explore the effects of a small difference (two semitones: B3–C♯4) and a moderate difference (11 semitones: B3–Bb4) of F_0. Natural instrument samples were extracted from the Studio On Line (SOL) database of Ircam (IRCAM, 2000): *guitar* (B3 was played on the E string, C♯4 on the A string, and Bb4 on the D string), *harp, violin pizzicato* (B3 and C♯4 on the G string, Bb4 on the D string), *bowed violin* (strings were the same as for the violin pizzicato), *bowed double bass* (all

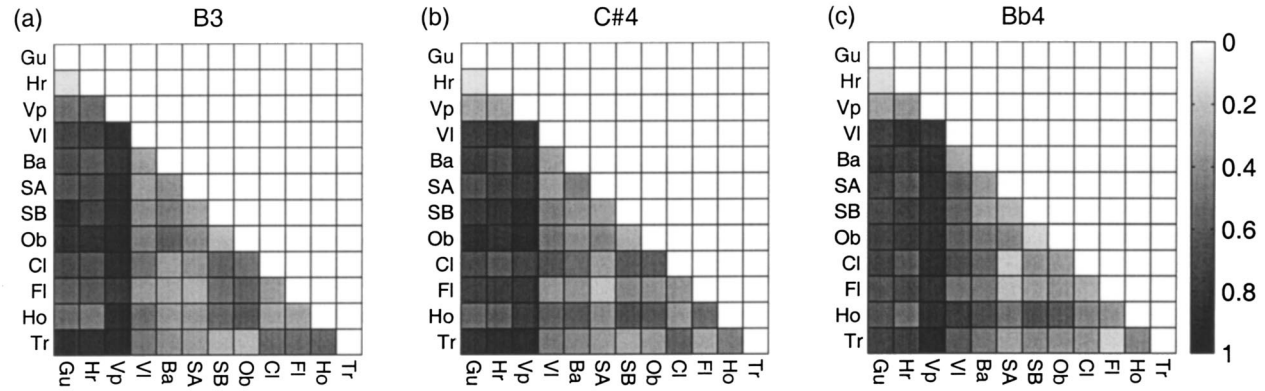


FIG. 1. Dissimilarity matrices for the three sessions of experiment I. Within each matrix, each square represents the dissimilarity between two instruments. Darker means greater dissimilarity. The first three columns of each matrix correspond to impulsive instruments (Gu, Hr, Vp). Dissimilarities are greater for pairs that associate an impulsive instrument with a sustained instrument than for pairs of instruments within either group. Patterns of dissimilarity are similar at each F_0.

notes were played on the G string), *oboe, clarinet, flute, horn in F, trumpet in C*. In the following, these instruments will be abbreviated as Gu, Hr, Vp, Vl, Ba, Ob, Cl, Fl, Ho, and Tr, respectively. In addition to these natural instruments, synthetic instruments SA and SB were created using fixed spectral envelopes derived from that of the saxophone.

Stimuli were clipped to a duration of 1.5 s by applying a 200-ms cosinusoidal offset ramp. Amplitudes were determined by asking six subjects (who did not participate in the main experiments) to adjust levels of stimuli presented at approximately 60 dB SPL for equal loudness. Stimuli were sampled at a rate of 44 100 Hz with 16-bit resolution, and presented diotically over Sennheiser 520 II headphones.

b. Subjects. Twenty-seven subjects aged 22 to 30 (14 men and 13 women, 15 musicians and 12 nonmusicians), participated in the experiment. Musicians were defined as having played an instrument for at least 3 years.

c. Procedure. Before the experiment, the subjects were informed that the goal of the experiment was to estimate the similarity of timbre between sounds. Timbre was defined as ''the fourth component of sound quality, the first three being pitch, loudness and duration.'' For each pair, they were instructed to judge whether the timbres were similar or different, using the entire scale of the cursor. Eventual differences of pitch, loudness, duration or ''recording noise'' were to be ignored. The identity of the instrument, if recognized, was also to be ignored. Subjects sat inside an audiometric booth. Presentation software was based on the PsiExp environment (Smith, 1995). The screen comprised a mouse-controlled cursor labeled from ''similar'' (coded 0) to ''different'' (coded 1), and two buttons (one to listen to the pair again, the other to validate the response). The experiment consisted of three sessions that were performed on the same day, separated by 5-min breaks. Before each session the subjects were presented with each of the 12 stimuli in random order to acquaint them with the range of timbre differences in the set of instruments. They were then presented with the full set of 66 pairs of different stimuli. The order within pairs and the order of pairs were random (a different randomization was used for each session and subject). Data for this and the following experiments are available at http://www.ircam.fr/pcm/archive/timbref0.

2. Results

a. Outliers, effect of musical experience. Correlation coefficients between dissimilarity scores were calculated for all pairs of subjects. These scores were submitted to a hierarchical cluster analysis based on the nearest-neighbor (complete linkage) algorithm (Kaufman and Rousseeuw, 1990). On the basis of this and a similar analysis for experiment II, three subjects were discarded for both experiments. Analysis was performed on data of the remaining 24 subjects.

To reveal an eventual effect of musical experience, an analysis of variance (ANOVA) was performed for each session with between-subjects factor musical experience (2) and within-subjects factor instrument pair (66), taking into account the fact that experience levels were represented by variable numbers of subjects (Abdi, 1987; Wonnacott and Wonnacott, 1990). No effect of musical experience was found, either as a main effect $[F(1,22)<1]$ or as an interaction $[F(65,1430)<1]$. Data for both groups are subsequently combined.

b. Dissimilarity matrices. Dissimilarity scores for each subject and session were placed in a matrix of dimension nXn, where n is the number of stimuli and the ijth entry $(i>j)$ is the dissimilarity between stimuli i and j. Since order was not distinguished, only the lower triangle was filled. Matrices averaged over subjects are plotted in Fig. 1 for the three sessions. Averaged over F_0's and subjects, dissimilarities ranged from 0.146 between the guitar and the harp to 0.872 between the trumpet and the violin pizzicato. One can distinguish two groups of instruments: impulsive (Gu,Hr,Vp) and sustained (Vl,Ba,SA,SB,Ob,Cl,Fl,Ho,Tr). Dissimilarities tended to be small within each group (upper left and lower right triangles) and large between groups (lower left rectangle), a pattern that was stable across F_0's.

To quantify the effects of F_0, a repeated-measures ANOVA was performed with factors $F_0(3) \times \text{instrument pair}(66)$. Results are shown in Table I. The

TABLE I. ANOVA table for experiment I. S: subjects, F_0: fundamental frequency, I: instrument pairs, SS: sum of squares, MS: mean square, F: F-values, ϵ: Greenhouse–Geisser correction factor applied to the degrees of freedom, p: corrected p-value, R^2: percentage of total variance accounted for by each effect. Adding a total of 48.5% due to intersubject differences, variance scores sum to 100%.

Source	df	SS	MS	F	ϵ	p	R^2
S	23	32.02	1.39				
F_0	2	0.93	0.46	5.72	0.90	0.008	**0.3**
F_0*S	46	3.73	0.08				
I	65	167.65	2.58	53.97	0.10	0.0001	**46.6**
I*S	1495	71.45	0.05				
F_0*I	130	16.53	0.13	5.65	0.122	0.0001	**4.6**
F_0*I*S	2990	67.29	0.02				

effects of both main factors were significant, as was their interaction. It is instructive to consider effect sizes. The percentage of total variance accounted for by each effect is indicated by the R^2 coefficient (last column in Table I) (Wonnacott and Wonnacott, 1990). The main effect of instrument pair represents the part of interinstrument dissimilarity that is constant across F_0. It accounts for about 47% of the variance. The interaction and main effect of F_0 together represent the part of dissimilarity that varies across F_0. They account for only about 5%. In agreement with the relatively small interaction, correlation coefficients between matrices (averaged over subjects, considering only the lower triangular parts) are relatively large: 0.88 between "a" and "b," 0.81 between "a" and "c," and 0.89 between "b" and "c" (df=64, $p<0.001$ for all three coefficients).

It could be argued that F_0-related effects are dwarfed by the contrast between impulsive and sustained instruments. Table II shows the percentage of variance accounted for by each effect for the full data set (column 2), or when dissimilarity scores are restricted to pairs of impulsive, sustained, or impulsive and sustained instruments (columns 3–5). After removing this major source of F_0-independent variance, as expected, F_0-independent effects represent a smaller proportion of total variance. However they still are larger than F_0-related effects.

To summarize the results of experiment I, interinstrument *timbre dissimilarities* varied significantly with F_0, but the variation was relatively small. It would be nice to conclude that *timbres* themselves were stable to the same degree [hypothesis (1) of the Introduction]. Unfortunately the results of experiment I do not allow us to draw that conclusion. As

TABLE II. Percentage of variance accounted for by each effect in experiment I, for the complete data set (All) or for data restricted to pairs of impulsive or sustained instruments, or to mixed pairs (impulsive and sustained). The first line (F_0-related) represents the sum of the F_0 effect and its interactions. The last line (other) represents variance due to disagreement between subjects. Each column sums to 100.

	R^2(%)			
Source	All	Impulsive	Sustained	Mixed
F_0-related	4.9	10.2	9.5	3.8
I	46.6	26.2	16.4	17.1
Other	48.5	63.6	74.1	79.1

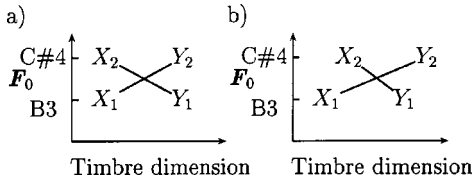


FIG. 2. Experiment II. Illustration of the hypothetical effect of F_0 on a dimension of timbre space (abscissa) along which two instruments X and Y differ. The instruments are represented at two different F_0's as X_1, Y_1 and X_2, Y_2 respectively, and it is supposed that $\overline{X_1Y_1}=\overline{X_2Y_2}$, as found in experiment I. The left plot illustrates hypothesis (1) (invariance) and the right hypothesis (2) (isometric shift). The latter implies $\overline{X_1Y_2}\neq\overline{X_2Y_1}$.

comparisons were made only at constant F_0, an eventual shift or rotation of the entire set of instruments in timbre space [hypothesis (2)] could not be detected. Furthermore, as subjects were instructed to use the full scale of dissimilarities in each session, an eventual compression or expansion also could not be detected. The next experiment allows for a shift, rotation, compression, or expansion to be detected.

B. Experiment II

In experiment II subjects rated the dissimilarity between stimuli with a constant *difference* of F_0 (ΔF_0) of either 2 semitones or 11 semitones. In contrast to experiment I, the response matrices were full, as each instrument pair was compared using both F_0 orders, and same-instrument pairs were included. Subjects were instructed to ignore differences in pitch which, contrary to experiment I, were salient. Supposing they can do so, this experiment allows us to refine the conclusions of experiment I, and in particular to decide between hypotheses (1) (invariance) and (2) or (3) (isometric or non-isometric deformation).

If hypothesis (1) is true, dissimilarity matrices should show three features. First, values on the diagonal should be zero. Second, the matrix should be symmetric: the lower triangular part should be the mirror image of the upper triangular part. Third, the lower triangular part should be identical to that observed at each F_0 in experiment I. To understand why the matrix should be symmetric, consider two instruments (X and Y) that differ along some dimension of timbre space (abscissa of Fig. 2). The positions of X and Y along this dimension at two F_0's are represented by X_1, Y_1 and X_2, Y_2, respectively. From experiment I we know that distances $\overline{X_1Y_1}$ and $\overline{X_2Y_2}$ are approximately equal. If additionally the timbres themselves are stable along this dimension, then we must have $\overline{X_1Y_2}=\overline{X_2Y_1}$ [Fig. 2(a)]. If instead they shift along this dimension, then $\overline{X_1Y_2}\neq\overline{X_2Y_1}$ [Fig. 2(b)]. Equality thus means either that timbres of X and Y did not shift with F_0, or that the shift was in a direction *orthogonal* to the dimension along which X and Y differ. Supposing that this holds for all instrument pairs, it follows that timbres did not move in the timbre space that spans the instrument set. Symmetry of the dissimilarity matrix, if observed, implies timbre invariance with respect to F_0 [hypothesis (1)].

1. Methods

Stimuli were those of experiment I, paired with a constant ΔF_0 of 2 semitones (B3–C♯4, session "a") or 11

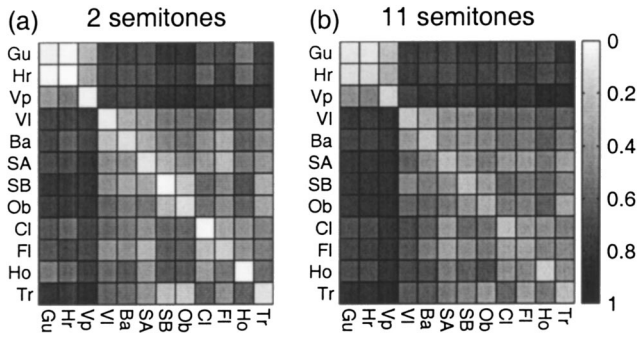


FIG. 3. Dissimilarity matrices for the two sessions of experiment II, each corresponding to a different F_0 pair. The lower triangular part corresponds to pairs for which the instrument on the abscissa was on the lower F_0 and the instrument on the ordinate on the higher. The upper triangular part corresponds to the opposite order. The diagonal represents instruments compared to themselves with an F_0 difference.

semitones (B3–Bb4, session ''b''). All pairs were included, resulting in 144 stimulus pairs per session. Within a session, the order of F_0's (low-high or high-low) was always the same, so as to make it easier for subjects to ignore the difference in pitch. Subjects were the same 27 that participated in experiment I. The three subjects that were eliminated from experiment I were also eliminated here. The remaining subjects were divided into four groups of approximately the same size (six to eight subjects) that differed in the order of presentation of sessions (''ab'' *versus* ''ba''), and in the order of F_0's within each session (low-first versus high-first). The proportion of musicians and nonmusicians was approximately the same in each group. Subjects performed both sessions on the same day (approximately one week after experiment I), separated by a 10-min pause.

2. Results

a. Dissimilarity matrices. Matrices averaged over subjects are plotted in Fig. 3 for the two sessions. Several features are obvious. First, ratings along the diagonals are relatively small. Second, the matrices appear fairly symmetrical. Third, the two matrices resemble each other. Fourth, the lower triangular parts of each matrix resemble the three matrices of experiment I.

To quantify these effects, the upper and lower triangular parts of both matrices were excised, ignoring the diagonals. The upper triangular parts were reflected with respect to the diagonal so as to have the same shape as the lower triangular parts, and the data were submitted to a repeated-measures ANOVA with factors instrument pairs $(66)\times\Delta F_0(2)\times F_0$ orders (2). Results are shown in Table III. The F_0 order factor (upper versus lower triangular parts) is not interpretable in itself as it depends on the arbitrary way in which instrument pairs were combined with F_0 pairs. It is included so as to allow F_0-dependent variance to be quantified.

Main effects and two-way interactions were significant, the three-way interaction was not. Effect sizes are quantified by R^2 scores (last column of Table III). The pair effect represents 43.8% of total variance, whereas F_0-related effects together sum up to a total of only 2.8%. As in experiment I, it appears that timbre dissimilarity depends on F_0 differences

TABLE III. ANOVA table for experiment II. S: Subjects, I: Instrument pairs, ΔF_0: F_0 difference, O: F_0 order, SS: sum of squares, MS: mean square, F: F-values, ϵ: Greenhouse–Geisser correction factor applied to the degrees of freedom, p: corrected p-Value, R^2: percentage of total variance accounted for by each effect (intersubject differences amounted to 53.4%).

Source	df	SS	MS	F	ϵ	p	R^2
S	23	54.5	2.37				
ΔF_0	1	1.01	1.01	5.31	1	0.03	**0.2**
ΔF_0*S	23	4.37	0.19				
O	1	0.69	0.69	10.48	1	0.003	**0.1**
O*S	23	1.52	0.07				
I	65	221.4	3.41	51.5	0.08	0.0001	**43.8**
I*S	1495	98.9	0.07				
ΔF_0*O	1	0.23	0.23	7.54	1	0.012	**0.04**
ΔF_0*O*S	23	0.68	0.03				
ΔF_0*I	65	5.58	0.09	3.62	0.21	0.0001	**1.1**
ΔF_0*I*S	1495	35.45	0.02				
O*I	65	4.14	0.06	2.3	0.2	0.006	**0.8**
O*I*S	1495	41.33	0.03				
ΔF_0*O*I	65	2.47	0.04	1.71	0.19	0.06	**0.5**
ΔF_0*O*I*S	1495	33.28	0.02				

to a limited degree. Taking the average over lower and (reflected) upper parts of the matrix for each session, the correlation coefficient between sessions is 0.95 (df=64, $p<0.001$). Averaging over sessions within experiment II and within experiment I, the correlation coefficient between experiments is 0.98 (df=64, $p<0.001$).

Table IV shows the percentage of variance accounted for by each effect for the full data set (column 2), or when dissimilarity scores are restricted to pairs of impulsive, sustained, or impulsive and sustained instruments (columns 3–5). The ratio of F_0-invariant effects (I) to F_0-dependent effects (ΔF_0, etc.) is smaller for restricted sets (particularly pairs of sustained instruments) than for the full set. Nevertheless, for each subset F_0-invariant effects remain larger than F_0-dependent effects.

When the diagonals of the matrices (not included in the previous analysis) were averaged over instruments, dissimilarity was 0.11 at 2 semitones and 0.20 at 11 semitones. Single sample t-tests show that the mean is significantly different from zero [$t(287)=10.7$, $p<0.0001$ for 2 ST; $t(287)=15.4$, $p<0.0001$ for 11 ST]. Further, in a repeated-measures ANOVA with factors Instrument $(12)\times\Delta F_0(2)$, the main factors were significant [$F(11,253)=5.3$, $\epsilon=0.43$, $p=0.0001$ and $F(1,23)=16.1$, $p=0.0005$, respectively], but their interaction was not. These results suggest that the pitch difference affected the dissimilarity judgments and that the

TABLE IV. Percentage of variance between dissimilarity scores accounted for by each effect of experiment II, for the entire data set (All) or for data restricted to pairs of impulsive, sustained, or impulsive and sustained (mixed) instruments. The last line (Other) represents variance due to disagreement between subjects. Each column sums to 100.

	R^2(%)			
Source	All	Impulsive	Sustained	Mixed
ΔF_0-dependent	2.8	6.3	4.6	2.8
I	43.8	29.2	15.2	9.1
Other	53.4	64.5	80.2	88.1

effect increased with increasing pitch difference. The effect was independent of instrument, however.

Supposing timbre invariance, we expected the diagonals to be zero. To some degree, the nonzero values observed can be attributed to an edge effect due to the fact that the response range had a lower bound of zero (variability of responses then necessarily results in a nonzero mean). However, given the significant effects of instrument and ΔF_0 this explanation is at best incomplete: we must admit a shift of timbre with F_0 (or a contamination of dissimilarity responses with pitch dissimilarity). The values on the diagonal are nevertheless small. Averaged over ΔF_0's, same-instrument dissimilarities were smaller (mean: 0.16) than different-instrument dissimilarities (mean: 0.59). The largest same-instrument dissimilarity (0.25 for the flute) was smaller than every different-instrument dissimilarity score except one (0.1 for Gu/Hr).

To summarize the results of experiment II, a first outcome is that subjects can compare timbre across F_0 despite salient pitch differences. Subjects apparently performed the tasks of experiments I and II in similar fashion. As a second outcome, we can rule out the hypothesis of a large global shift of timbre space with F_0, as dissimilarity matrices were symmetrical and their diagonals small. This extends the conclusion of experiment I that instruments retain their relative positions as F_0 changes: they also do not shift *as a group*. However, beyond these conclusions valid in the first approximation, both experiments revealed effects that were significant, albeit small. It would be nice to infer from these effects the nature of shifts of individual intruments. Unfortunately, each score reflects the timbre of *two* instruments, and it is not obvious which of the two determined a change in dissimilarity. Experiment III introduces a new form of analysis that reveals timbre changes of individual instruments with F_0.

C. Experiment III

In experiment III subjects rated timbre dissimilarity between pairs of instruments with and without a difference in F_0. The aim was to extend and generalize the results of experiments I and II, and in particular to see whether subjects could make reliable timbre dissimilarity judgments between sounds that differed by a variable amount along the pitch dimension.

1. Methods

To keep the stimulus set size within reasonable limits, 9 instruments were selected among the 12 used in experiments I and II. These were Gu, Hr, Vp, Vl, SA, Ob, Cl, Ho, and Tr. Each was played at two F_0's, resulting in a set of 18 sounds that were paired (excluding comparisons between the same instrument at the same F_0) to produce 153 pairs that were presented in a single session with a 5 min pause half-way. There were two sessions: ''a'' with notes B3 and C♯4 (2 semitones), ''b'' with notes B3 and Bb4 (11 semitones). Within a session, different-F_0 pairs were presented in the same order, low-high or high-low (depending on the subject). Otherwise, presentation conditions and instruments were the same as for experiment II.

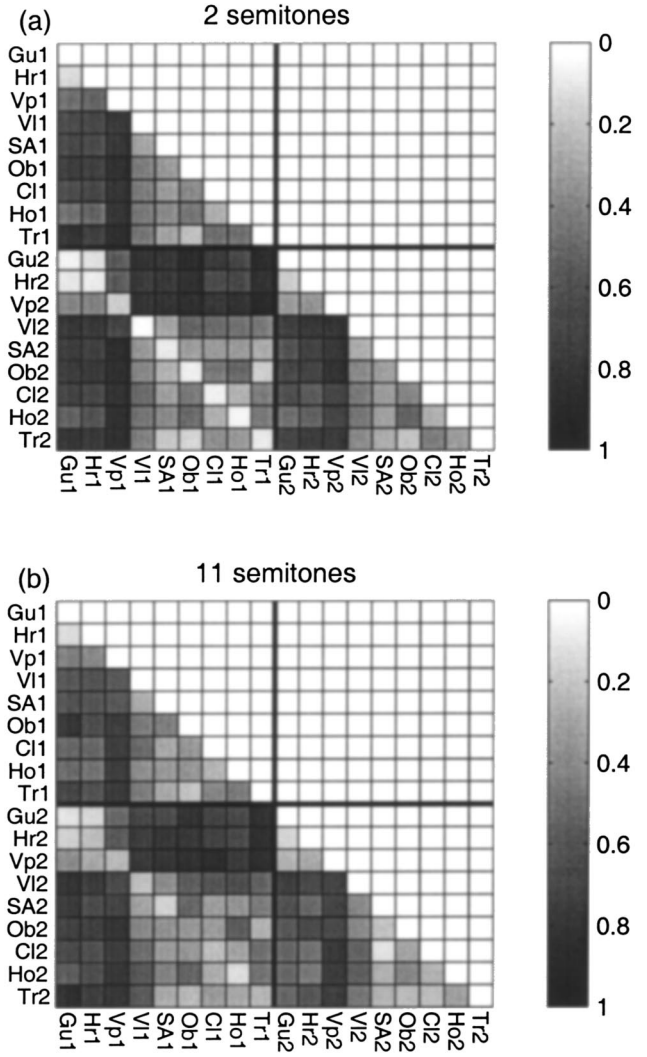


FIG. 4. Dissimilarity matrices for experiments IIIa and IIIb. For the axis labels, indices 1 and 2 mean that instruments were played at B3 and C♯4, respectively (Bb4 in experiment IIIb).

Session ''a'' involved 25 subjects aged 19 to 30, 15 men and 10 women, 13 musicians and 12 nonmusicians. None had participated in experiments I or II. Session ''b'' involved 18 subjects (11 of which had taken part in session ''a''), aged 19 to 30, eight women and ten men, nine musicians and nine nonmusicians.

2. Results

a. Outliers. Among the 25 subjects of session ''a,'' three gave answers that were poorly correlated with the rest [$r < 0.33$] and were excluded from the analysis. None were excluded from session ''b.''

b. Dissimilarity matrices. Dissimilarity scores averaged over subjects were placed in the lower triangular part of a matrix as shown in Fig. 4(a) for session ''a.'' This matrix has three parts: an upper-left triangle (instruments compared at B3), a lower-right triangle (instruments compared at C♯4), and a 9×9 square (instruments compared across F_0's). The two triangles are analogous to the matrices of experiments Ia and Ib, the square to that of experiment IIa.

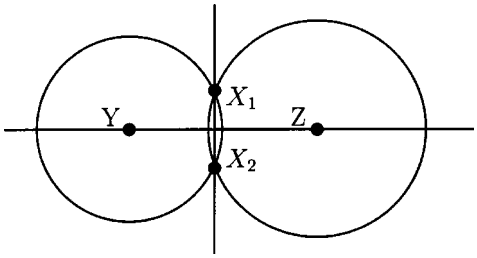


FIG. 5. Experiment III. Schema illustrating the anchor method of analysis of timbre change. The plane represents a hypothetical two-dimensional timbre space, *Y* and *Z* are ''anchor'' instruments, and *X* is an instrument whose displacement with F_0 is being considered. If the dissimilarity of *X* with respect to *Y* does not change when the F_0 of *X* is changed, X_1 and X_2 must be on a circle centered on *Y*. Similarly, if dissimilarity with respect to *Z* does not change, X_1 and X_2 must also belong to a circle centered on *Z*. The displacement of *X* is therefore on a line orthogonal to *YZ*. If the same is true for all anchor pairs, the displacement of *X* is orthogonal to the space they span (or else it is zero).

TABLE V. Effect size (R^2) of factors F_0 and $F_0\times$anchor for each instrument at both ΔF_0's. Only effects significant at the $p<0.05$ level are shown. These figures quantify the magnitude of displacement of each instrument in timbre space as a function of F_0.

	R^2(%)			
	2 semitones		11 semitones	
Instrument	F_0	$F_0\times$anchor	F_0	$F_0\times$anchor
Gu	⋯	⋯	⋯	⋯
Hr	⋯	⋯	⋯	⋯
Vp	2.79	2.51	⋯	4.28
Vl	⋯	⋯	1.17	⋯
SA	⋯	⋯	0.55	1.52
Ob	⋯	⋯	1.23	1.92
Cl	⋯	⋯	0.59	⋯
Ho	⋯	⋯	⋯	⋯
Tr	0.40	⋯	1.66	1.92

To compare results with those of experiments I and II, a triangular matrix similar to the one just described was populated with scores from corresponding conditions of experiment Ia (upper left triangle), Ib (lower right triangle) and IIa (square). The correlation between this composite matrix and that obtained from experiment IIIa was 0.95 (df=151, $p<0.0001$). Similarly, a triangular matrix was populated with scores of experiments Ia, Ic, and IIb. The correlation between this composite matrix and that obtained from experiment IIIb was 0.92 (df=151, $p<0.0001$). This indicates a high degree of similarity between data sets despite the difference in task and subjects. Overall ANOVAs are not reported here (they support conclusions similar to experiments I and II). Instead, a different analysis is presented that assigns effects to timbre changes of individual instruments.

c. Instrument-specific ANOVAs. Each of the nine instruments was analyzed in turn. For each, the eight other instruments were used as ''anchors'' with respect to which to measure its timbre changes.

To illustrate the principle, take an instrument *X* and denote its timbre at two different F_0's as X_1 and X_2, assimilated to two points in timbre space. We wish to know if X_1 and X_2 are distinct, and for this we use a second instrument *Y* as anchor. We ignore eventual shifts of *Y* itself for the moment. The displacement of *X* towards or away from *Y* can be estimated by comparing $\overline{X_1Y}$ and $\overline{X_2Y}$. In geometric terms, the equality

$$\overline{X_1Y}=\overline{X_2Y} \tag{1}$$

implies that *X* has followed a hypersphere centered on *Y* (illustrated as a circle in Fig. 5). If a similar equality holds for another anchor instrument *Z*, then X_1 and X_2 belong to the intersection of two hyperspheres. In the plane (Fig. 5), the intersection consists of two points on a line perpendicular to *YZ*. In three dimensions it would be a circle in a plane orthogonal to *YZ*, and in higher dimensions a sphere or hypersphere in a hyperplane orthogonal to *YZ*. In every case the displacement is *orthogonal* to the timbre dimension along which *Y* and *Z* differ. If Eq. (1) holds for every anchor, taking them two by two, it follows that the displacement of *X* is orthogonal to the subspace that contains the anchors. Supposing that the anchors together span the whole of timbre space, *X* did not move in this space.

Actually, each instrument has two positions, e.g., Y_1 and Y_2 for *Y*. Either could be used as the anchor, but there is a difficulty. Testing for $\overline{X_1Y_1}=\overline{X_2Y_1}$, instruments on the left have the same F_0 but those on the right differ. The comparison is thus sensitive to eventual effects of an F_0 difference *per se* (for example, if subjects failed to completely ignore pitch). Using Y_2 instead as the anchor we have a similar problem in the other direction. However, by adding term to term,

$$\overline{X_1Y_1}+\overline{X_1Y_2}=\overline{X_2Y_1}+\overline{X_2Y_2}, \tag{2}$$

F_0-related effects apply equally to both sides and thus balance out. Equation (2) can be used in place of Eq. (1) for the previous analysis. To summarize, if Eq. (2) holds when an instrument is compared to each of the eight others, we may assume that that instrument's timbre did not change with F_0.

d. Two semitones. For each instrument *X*, terms of Eq. (2) were compared using a repeated-measures ANOVA with factors anchor (8)$\times F_0$ (2). To be precise: the F_0 factor contrasted timbres X_1 and X_2 at two different F_0's by comparing $\overline{X_1Y_1}+\overline{X_1Y_2}$ to $\overline{X_2Y_1}+\overline{X_2Y_2}$. The anchor factor contrasted the various anchor instruments *Y*. Nine such ANOVAs were performed. The main effect of anchor was, as expected, highly significant for all instruments, and will not be considered further. For seven instruments (Gu,Hr,Vl,SA,Ob,Cl,Ho), the effect of F_0 and its interaction with anchor were not significant. For the trumpet, the main effect of F_0 was significant but tiny ($R^2=0.4\%$, as compared to 65.5% for anchor). For the violin pizzicato, both the main effect of F_0 and its interaction with anchor were significant and relatively large. The other instruments remained essentially stable when F_0 changed from B3 to C♯4. These results are summarized in columns 2 and 3 of Table V.

e. Eleven semitones. Similar ANOVAs were performed for session ''b.'' Effect sizes are summarized in the last two columns of Table V. These effects were nonsignificant for Gu, Hr, and Ho, and very small for Cl. They were significant and larger for Vp, Vl, SA, Ob, and Tr.

To summarize the results of experiment III, subjects succeeded in making timbre dissimilarity judgments while largely ignoring a difference in pitch that was present on some trials and not on others. Experiments I and II had found *timbre dissimilarity* to be fairly stable with F_0 changes. Experiment III refined this conclusion: *timbre* itself was stable for some instruments (eight for 2 semitones, four for 11 semitones, out of nine intruments). Timbres of others appeared to change slightly. The next section presents a MDS analysis that allows these changes to be interpreted in terms of displacement within a model of perceptual timbre space.

III. MDS ANALYSIS

For each session of experiment III, the dissimilarity matrices for all subjects were processed by the EXSCAL MDS program (Winsberg and Carroll, 1989). We chose a two-way MDS model without individual differences, as this model is rotationally invariant, allowing solutions to be rotated and their dimensions compared to physical descriptors, as well as compared across experiments. The two-way EXSCAL model postulates that the distance, d_{ij}, between the ith and jth stimuli, is given by

$$d_{ij}=\left[\sum_{r=1}^{R}(X_{ir}-X_{jr})^2+(S_i+S_j)\right]^{1/2}, \tag{3}$$

where X_{ir} is the coordinate of the ith stimulus on the rth dimension and R is the number of dimensions. In this model, in addition to R common dimensions, the stimuli have unique dimensions not shared by other stimuli. The specificity or uniqueness of the ith stimulus is denoted S_i. Since a maximum likelihood criterion is used to estimate the fit of the model to the data, BIC statistics (Schwarz, 1978) can be used to choose the dimensionality R and decide whether additional unique dimensions should be included.

The BIC criterion suggested three- and two-dimensional models without specificities for experiments IIIa and IIIb, respectively. For nonlinear models like MDS, BIC statistics have a heuristic value and do not preclude consideration of other models. We therefore also examined two-, three-, and four-dimensional models in search of a model interpretable in terms of dimensions related to signal descriptors. In each case, the solution was rotated with a procrustean procedure to a target matrix of signal descriptors described in Sec. IV. Only the four-dimensional solutions will be described in detail.

Solutions for sessions ''a'' (2 semitones) and ''b'' (11 semitones) are illustrated in the upper and lower parts of Fig. 6, respectively. For each instrument, the position at B3 is represented by the symbol and that at the other F_0 (C♯4 or Bb4) by the extremity of the line. The first three dimensions of the spaces are well correlated between sessions (0.99, 0.95, and 0.94, respectively). A large score for dimension 1 is to be expected because the salient contrast between impulsive and sustained instruments is unlikely to depend on F_0. However, the good scores for dimensions 2 and 3 suggest that additional dimensions of timbre are stable across F_0. The fourth dimension was poorly correlated between sessions (0.31).

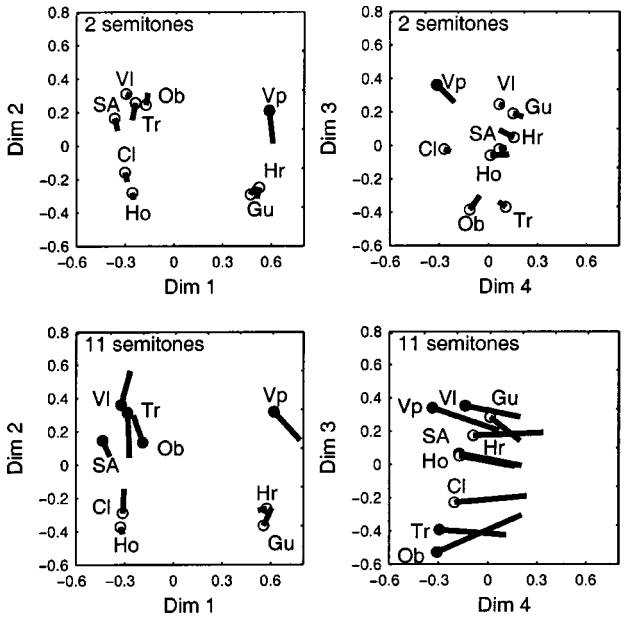


FIG. 6. Timbre spaces for experiment IIIa (top) and IIIb (bottom). The symbol represents the position of the instrument at the lower F_0 (B3), the end of the line represents the position of the same instrument at the other F_0 (C♯4 or Bb4). The symbol is filled for instruments for which a significant timbre change was found in Sec. II C and open for others.

Filled symbols in Fig. 6 are instruments for which we know (on the basis of the ANOVAs of the previous section) that their timbre changed. We would expect the lines to be of nonzero length (in at least one projection) for them and of zero length for open symbols. Such is not always the case. A possible explanation for this discrepancy is that the data used for ANOVAs excluded dissimilarities between the same instrument at different F_0's, whereas the MDS included them. Even if the timbre of an instrument did not change across F_0, the measured dissimilarity was likely to take a nonzero value as a result of an edge effect (Sec. II B 3 a) or a residual pitch dissimilarity that the subjects failed to ignore. This has the effect of ''pushing apart'' the corresponding points of the MDS solution. Whatever the explanation, this discrepancy weakens the usefulness of interpreting the detailed pattern of F_0-induced shifts we observe in Fig. 6.

IV. COMPARISON WITH SIGNAL DESCRIPTORS

In the spirit of previous studies on timbre, this section attempts to relate perceptual dimensions revealed by MDS to descriptors of the signal (sometimes called ''physical dimensions''). A feature of the present study is that this relation is tested over several fundamental frequencies. On the basis of our data we can formulate three constraints for a signal-based descriptor: (1) at each F_0, the descriptor should predict the corresponding perceptual dimension; (2) for instruments whose timbre did not vary across F_0, descriptor values should not vary; and (3) for instruments whose timbre did vary across F_0, and to the degree that this variation is reliably described in terms of change along a perceptual dimension, we should observe a corresponding change of the descriptor. We consider only data for experiment IIIb (11 semitones).

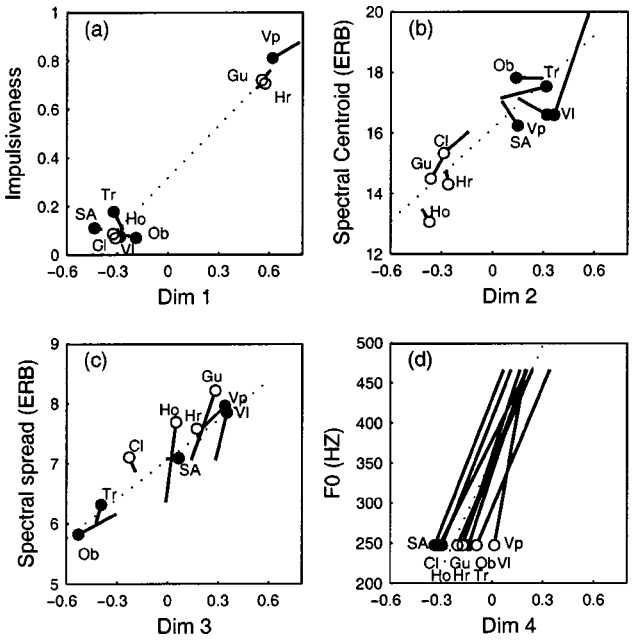


FIG. 7. Experiment IIIb (11 semitones). Scatter plots relating each signal descriptor to the MDS dimension that it explains best. (a) Impulsiveness versus dimension 1. (b) Spectral centroid versus dimension 2. (c) Spectral spread versus dimension 3. (d) F_0 versus dimension 4. For each instrument, the symbol represents its position at note B3, and the opposite end of the line represents its position at note Bb4. Filled symbols indicate instruments for which the timbre changed significantly according to the analysis of C. Open symbols represent instruments for which it did not. Dotted lines represent regression lines.

A. Dimension 1

To predict the first dimension we use a measure of impulsiveness proposed by Susini (1996), defined as follows. The instantaneous power s_n^2 is smoothed by convolution with an 8-ms square window. The duration during which the smoothed power is above 40% of its maximum value is then divided by the duration for which it is above 10%, and one minus this ratio is taken as the measure of impulsiveness. It is close to one for impulsive sounds and to zero for sustained sounds.

Figure 7(a) plots this descriptor as a function of dimension 1 for experiment IIIb. The descriptor does a good job of predicting the clustering of impulsive and sustained instruments into well-separated groups. Correlation coefficients are 0.98 for experiment IIIa and 0.99 for experiment IIIb (df=16, $p<0.01$ in both cases). As a comparison, previous studies (e.g., Krimphoff *et al.*, 1994) suggested the log of attack time as a descriptor for impulsiveness. That descriptor gave correlation scores of 0.95 and 0.94 (df=16, $p<0.01$ in both cases) for experiments IIIa and IIIb, respectively, when the MDS solutions were rotated towards values determined by it.

B. Dimension 2

To predict dimension 2 we used a spectral centroid descriptor similar in spirit to the definition of sharpness (Zwicker and Fastl, 1990; Hartmann, 1997). The waveform was first filtered to model the drop in sensitivity at low and high frequencies due mainly to outer and middle ear filtering (Killion, 1977). Then it was filtered by a gammatone filterbank (Patterson *et al.*, 1995; Slaney, 1993) with channels spaced at half-ERB intervals on an ERB-rate scale (z) calculated according to the formula $z=21.3\log(0.00437f+1)$ between 25 Hz and 19 kHz (Hartmann, 1997). Instantaneous power was calculated within each channel and smoothed by delaying it by $1/4f_c$ (where f_c is the characteristic frequency of the channel), adding it to the undelayed power, and convolving the sum with an 8-ms window corresponding roughly to the equivalent rectangular duration of power integration measured by Plack and Moore (1990). Smoothed power was then raised to the power 0.3 to obtain a rough measure of ''partial loudness'' for each channel. The partial-loudness-weighted average of ERB rate was taken over channels, the result being an ''instantaneous spectral centroid'' function of time according to

$$\overline{z}(t)=\sum_z z\psi_z(t) \Big/ \sum_z \psi_z(t), \tag{4}$$

where $\psi_z(t)$ is the ''partial loudness'' of the channel z at instant t. Finally, the instantaneous centroid $\overline{z}(t)$ was weighted by ''instantaneous loudness'' (sum over channels of partial loudness) and averaged over time to obtain a single descriptor value, $\overline{z}$, to characterize the entire signal.

Figure 7(b) shows the value of spectral centroid as a function of dimension 2. Data points are relatively well aligned. We expect the displacements of those instruments that significantly changed in timbre (filled symbols) to follow this trend. Such is roughly the case for Vl and Tr, but not for SA, Vp, or Ob. The descriptor thus predicts the overall trend but not all details. The correlation between descriptor values and projections along dimension 2 is 0.93 and 0.90 for experiment IIIa and IIIb, respectively (df=16, $p<0.01$ in both cases).

Our definition of spectral centroid is one of many that have been proposed. A common definition is the following:

$$\overline{k}=\sum_k ka_k \Big/ \sum_k a_k, \tag{5}$$

where k is the rank of a partial and a_k is its amplitude (on a linear, power, or log scale). If the spectral envelope remained constant when F_0 varies (approximately the case for most of our instruments), this definition would lead to an *inverse* dependency of k with F_0, a variation of a factor 1.9 between B3 and Bb4. Since timbre was instead rather stable, this definition can be ruled out, as concluded earlier by Slawson (1968) or Plomp (1976). A better definition defines the centroid as a weighted sum of frequencies (e.g., Kendall *et al.*, 1999), for example:

$$\overline{f}=\sum_k f_k a_k \Big/ \sum_k a_k, \tag{6}$$

where k is the rank of a partial or discrete Fourier transform coefficient, f_k is its frequency and a_k is its amplitude (on a linear, power or log scale). For a constant spectral envelope this definition leads to values of $\overline{f}$ that are approximately constant as F_0 varies. However, there are several ways of implementing this definition according to whether a_k desig-

nates the linear, power, or log amplitude, whether k designates the rank of a partial, a DFT coefficient, or a filter band, whether the frequency scale is linear or warped (log or ERB-rate scale), whether a nonlinearity is applied after summing coefficients within channels, etc. Our definition of spectral centroid was chosen to make all operations and parameters explicit in a psychoacoustically reasonable way, and avoid hidden parameters such as window size or sampling rate, or the implicit assumption of a line spectrum needed to apply Eq. (6).

As a comparison, the definition of Eq. (6) implemented according to Peeters *et al.* (2000) gave correlation coefficients of 0.95 and 0.85 for experiments IIIa and IIIb, respectively, when the MDS solutions were rotated using that definition (df=16, $p<0.01$ in both cases).

C. Dimension 3

Dimension 3 was found to be relatively well correlated with a measure $\tilde{z}$ of spectral spread defined as

$$\tilde{z}=\sqrt{\sum_z (z-\overline{z})^2\psi_z \Big/ \sum_z \psi_z}. \tag{7}$$

Figure 7(c) shows the value of spectral spread as a function of dimension 3 for experiment IIIb. Data points are roughly distributed along a line. Two instruments that significantly changed in timbre (Ob and Vp) move roughly along this line, as expected. However, two instruments that did not change timbre (Ho and Gu) also show relatively large changes in descriptor value. Such is also the case for Vl, which did change timbre but (according to the MDS analysis) not along this dimension. The descriptor would be better if such changes could be avoided. Overall, the correlation between descriptor values and projections along dimension 3 was 0.94 and 0.87 for experiments IIIa and IIIb, respectively (df=16, $p<0.01$ in both cases).

As a comparison, the definition of spectral spread of Peeters *et al.* (2000), analogous to the spectral centroid definition of Eq. (6), gave correlation scores of 0.83 and 0.65, respectively, when the MDS solutions were rotated to that descriptor. Our descriptor was also applied to the stimuli used by McAdams *et al.* (1995) and compared to the coordinates along the third dimension of their MDS space. The correlation found was 0.87, as opposed to 0.54 for the spectral flux descriptor used in that study.

D. Dimension 4

Dimension 4 was found to be well correlated with F_0 (0.90) for experiment IIIb. For experiment IIIa the correlation with F_0 was poor and no better descriptor was found. Figure 7(d) plots F_0 as a function of dimension 4 for experiment IIIb. Displacements of all instruments are roughly parallel with the regression line, consistent with the good correlation. Subjects thus based their timbre dissimilarity judgments in part upon a dimension related to F_0. This is the only evidence we found of a pitchlike dimension.

To summarize, the signal descriptors reviewed roughly satisfy the constraint of F_0-invariance for instruments that did not change timbre. For instruments that did change timbre, the minor changes in descriptor value with F_0 were in some cases consistent with the minor changes in timbre, in other cases not. Overall, the descriptors did a very good job of predicting perceptual dimensions. They compared favorably with previously proposed descriptors, but variability of data is such that we cannot reliably conclude on this basis alone that one given descriptor is superior to another.

V. DISCUSSION

A first outcome of this study, not obvious from the start, is that timbres of instruments played at different notes can be compared. Classic techniques such as MDS can be applied, and this opens the perspective for more detailed and extensive studies of timbre variations of specific instruments across their register. Subjects performed the task in a very similar fashion with or without F_0 differences between stimuli, and had little difficulty ignoring the very salient pitch differences that accompanied them. Timbre behaved as if it were separable from pitch, and there was only slight evidence of a small perceptual interaction between pitch and timbre dimensions.

Cross-F_0 timbre comparison being possible, a second outcome is the relative stability of timbre with respect to F_0 changes. For several instruments there was no measurable change in timbre, so we can exclude the hypothesis of a basic, non-instrument-specific dependency of timbre upon F_0. The hypothesis that such a dependency does exist, but was balanced by opposite changes of instrument characteristics, is unlikely to be simultaneously true for four out of nine instruments across all F_0's, and eight between B3 and C♯4. Lack of measurable change is not due to lack of sensitivity of our methods: for other instruments we demonstrated significant timbre changes of relatively small size.

The ''anchor-based'' analysis technique introduced in Sec. II C 2 revealed small but significant timbre changes for certain instruments. The MDS analyses provided an interpretation of the changes in terms of displacement along particular dimensions of timbre space. However, relatively large displacements were also observed for instruments known *not* to have changed timbre significantly, so we must not give too much weight to such detailed features, as argued in Sec. III. MDS solutions were generally stable across experiments and conditions, and the correlations between their dimensions and physical descriptors was high, as found in previous studies.

Stability of timbre as a function of F_0 for certain instruments puts a strong constraint on signal descriptors for predicting timbre: they too must demonstrate the same degree of stability. Such was the case for the descriptors we used, but other methods proposed in the literature may not be so stable. These conclusions are very important for applications that use signal descriptors for content-based indexing of audio and multimedia data. So far, such descriptors had been validated only at particular F_0's. Our results demonstrate that they generalize well to the F_0's we tested, although the question remains open for the wider range of F_0's.

We used relatively small F_0 steps because we expected the task of comparing timbre while ignoring pitch to be difficult (Miller and Carterette, 1975). The pitch differences are

nevertheless quite salient. The smaller step (two semitones, a major second) is one-third the maximum distance along the chroma circle. The larger step (11 semitones, a major seventh) is both larger in terms of tone height and smaller in terms of chroma, and thus offered the opportunity to tease apart the eventual contributions of each. It is also about one-third of the range of typical instruments such as the violin, and thus probes instrument-specific variations to some extent. Obviously, a wider range of notes is needed for a more complete study of instrument-specific timbre variations. The present study showed that such a study is in principle possible. There is, however, evidence that instrument identification performance degrades beyond an octave (Handel and Erickson, 2001).

The generality of our results is also limited by our choice of instruments. Previous studies found that criteria vary according to the stimulus set, leading to MDS solutions that correlate with rather different physical dimensions. One or both of our first two MDS dimensions were usually also salient in those studies, but one cannot exclude that for certain stimulus sets, other dimensions might be relevant that are more sensitive to F_0 changes.

VI. CONCLUSIONS

(1) Subjects made timbre dissimilarity judgments between natural musical instrument sounds that differed in F_0 by 0, 2 or 11 semitones. Results were orderly, even when the stimulus set comprised both same-F_0 and different-F_0 sound pairs. The salient pitch difference produced by the F_0 differences did not prevent timbre comparisons.

(2) As a first approximation, timbre dissimilarities depended little on F_0. Dissimilarity scores varied more between different-instrument pairs than across F_0's. Experiment I (same F_0) showed that instruments kept their relative positions in timbre space at different F_0's, and experiments II and III showed further that they did not shift as a group.

(3) As a second approximation, small but significant timbre changes were observed. Instrument-specific ANOVAs in experiment III found that the changes affected certain instruments and not others. It is likely that these timbre changes were due to instrument-specific changes in, for example, resonator geometry.

(4) The lack of significant effects for certain instruments in the instrument-specific ANOVAs of experiment III, together with the symmetry of dissimilarity matrices in experiment II, suggest the absence of any basic, non-instrument-specific change of timbre with F_0.

(5) Multidimensional scaling yielded low-dimensional linear models of perceptual timbre spaces (four-dimensional without specificities). After an appropriate rotation, dimensions were found to be well correlated with a set of signal-based descriptors. Projections on each of the first three dimensions were relatively stable with F_0. The projection on the fourth was correlated with F_0 for an 11-semitone (but not 2-semitone) step. This is the only evidence we found for a pitchlike dimension in a timbre space.

(6) Signal-based descriptors ''impulsiveness,'' ''spectral centroid,'' ''spectral spread,'' and F_0 were used. The first describes the temporal envelope. The second two describe the spectral envelope in terms of the first two moments of a ''partial loudness'' spectrum (cubic root of power within channels of a cochlear filter bank). These three descriptors appeared to be good predictors of the first three timbre dimensions over the range of F_0's used, while the fourth (F_0) is known as a good predictor of pitch.

This study opens the way for more extensive studies of timbre change with F_0, such as instrument-specific timbre changes across their register. The anchor method applied in experiment III seems particularly promising to distinguish timbre changes from fluctuations due to experimental noise.

ACKNOWLEDGMENTS

The authors thank Stephen Handel for useful comments on a previous version of this paper. This work is part of the first author's Ph.D project which is funded by the Swiss National Science Foundation. The research project is funded in part by the European Union Project CUIDADO and was conducted within the Music Perception and Cognition team at IRCAM, and at CNMAT, University of California, Berkeley.

Abdi, H. (**1987**). *Introduction au Traitement Statistique des Données Expérimentales (Introduction to Statistical Processing of Experimental Data)* (PUG, Grenoble).

ANSI (**1960**). ''USA Standard Acoustical Terminology'' (American National Standards Institute, New York).

de Cheveigné, A. and Kawahara, H. (**1999**). ''Missing data model of vowel perception,'' J. Acoust. Soc. Am. **105**, 3497–3508.

Fletcher, N. H., and Rossing, T. D. (**1998**). *The Physics of Musical Intruments*, 2nd ed. (Springer-Verlag, New York).

Grey, J. M. (**1977**). ''Multidimensional perceptual scaling of musical timbres,'' J. Acoust. Soc. Am. **61**, 1270–1277.

Handel, S., and Erickson, M. L. (**2001**). ''A rule of thumb: The bandwith for timbre invariance is one octave,'' Music Percept. **19**, 121–126.

Hartmann, W. (**1997**). *Signals, Sound, and Sensation* (AIP, New York).

Helmholtz, H. (**1885**). *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (from 1877 trans. by A. J. Ellis of 4th German ed., republ. 1954 by Dover, New York).

IRCAM (**2000**). ''Studio On Line'' http://www.ircam.fr/.

Kaufman, L., and Rousseeuw, P. J. (**1990**). *Finding Groups in Data. An Introduction to Cluster Analysis* (Wiley-Interscience, Brussel).

Kendall, R., Carterette, E., and Hajda, J. (**1999**). ''Perceptual and acoustical features of natural and synthetic orchestral instrument tones,'' Music Percept. **16**, 265–294.

Killion, M. C. (**1977**). ''Revised estimate of minimum audible pressure: Where is the 'missing 6 dB'?'' J. Acoust. Soc. Am. **63**, 1501–1508.

Krimphoff, J., McAdams, S., and Winsberg, S. (**1994**). ''Caractérisation du timbre des sons complexes. Analyses acoustiques et quantification psychophysique'' (Characterization of the timbre of complex sounds. Acoustical analyses and psychophysical quantification), J. Phys. I **4**, 625–628.

Luce, D., and Clark, M. (**1967**). ''Physical correlates of brass-instrument tones,'' J. Acoust. Soc. Am. **42**, 1232–1243.

Martin, K. D. (**1999**). ''Sound-Source Recognition: A Theory and Computation Model,'' Massachusetts Institute of Technology, unpublished doctoral dissertation.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (**1995**). ''Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes,'' Psychol. Res. **58**, 177–192.

Miller, J. R., and Carterette, C. (**1975**). ''Perceptual space for musical structures,'' J. Acoust. Soc. Am. **58**, 711–720.

Misdariis, N., Smith, B., Pressnitzer, D., Susini, P., and McAdams, S. (**1998**). ''Validation of a multidimensional distance model for perceptual dissimilarities among musical timbres,'' J. Acoust. Soc. Am. **103**, 2812.
Patterson, R. D., Milroy, R., and Allerhand, M. (**1993**). ''What is the octave of a harmonically rich note?,'' Contemp. Music Rev. **9**, 69–81.
Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1995**). ''Complex sounds and auditory images,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 429–446.
Peeters, G., McAdams, S., and Herrera, P. (**2000**). ''Instrument sound description in the context of MPEG-7,'' Proc. Int. Conf. Comput. Music, Berlin, pp. 166–169.
Plack, C. J., and Moore, B. (**1990**). ''Temporal window shape as a function of frequency and level,'' J. Acoust. Soc. Am. **87**, 2178–2187.
Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, London).
Risset, J., and Wessel, D. (**1999**). ''Exploration of timbre by analysis and synthesis,'' in *The Psychology of Music*, 2nd ed., edited by D. Deutsch (Academic, New York), pp. 113–169.
Schwarz, G. (**1978**). ''Estimating the dimensions of a model,'' Ann. Stat. **6**, 461–464.
Shepard, R. N. (**1964**). ''Circularity in judgments of relative pitch,'' J. Acoust. Soc. Am. **36**, 2346–2353.
Slaney, M. (**1993**). ''An efficient implementation of the Patterson-Holdsworth auditory filter bank,'' Apple Computer Technical Report 35.
Slawson, A. (**1968**). ''Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency,'' J. Acoust. Soc. Am. **43**, 87–101.
Smith, B. (**1995**). ''PsiExp: an environment for psychoacoustic experimentation using the IRCAM Musical Workstation,'' in *Society for Music Perception and Cognition Conference '95*, edited by D. Wessel (Univ. of California, Berkeley).
Susini, P. (**1996**). ''Analyses acoustiques des sons'' (Acoustical analyses of sounds), IRCAM Technical Report.
Ueda, R. D., and Nimmo-Smith, I. (**1987**). ''Perceptual components of pitch: Spatial representation using a multidimensional scaling technique,'' J. Acoust. Soc. Am. **82**, 1193–1200.
Winsberg, S., and Carroll, J. D. (**1989**). ''A quasi-nonmetric method for multidimensional scaling via an extended euclidean model,'' Psychometrika **54**, 217–229.
Wonnacott, T. H., and Wonnacott, R. J. (**1990**). *Introductory Statistics for Business and Economics*, 4th ed. (Wiley, New York).
Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics: Facts and Models* (Springer-Verlag, New York).

The structural dynamics of the American five-string banjo

Joe Dickey[a)]
Center for Nondestructive Evaluation, The Johns Hopkins University, 810 Wyman Park Drive, Suite G010, Baltimore, Maryland 21211

(Received 24 January 2003; revised 5 September 2003; accepted 8 September 2003)

The American five-string banjo is unique among musical instruments in that many significant parameters that effect tone are easily adjusted. This is probably why so many banjo players fiddle with their banjo. The instrument is a combination of canonical vibrating systems (strings, and a circular membrane) and therefore more amenable to analysis and modeling than most other musical instruments (e.g., the violin). Such an analysis is presented here. The model is a harmonically driven string which excites the other strings and a membrane under tension, causing the membrane to radiate sound. Three figures-of-merit, FOMs, are assumed. They are *loudness*, *brightness*, and *decay* of the sound. The effects of a number of parameters on the proposed FOMs are investigated. Among these are the loss factor and tension of the membrane, the mass of the bridge, and the location on the string of the excitation. It is noted that the calculated effects of the changes agree with generally accepted setup practices. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1621863]

PACS numbers: 43.75.De, 43.40.At [NHF] Pages: 2958–2966

I. INTRODUCTION

The banjo is a simple, if not subtle, instrument. It is a structure comprised of canonical systems, i.e., strings, and a circular membrane, and is therefore relatively straightforward to model. The strings, when plucked or strummed, vibrate and excite the membrane, causing it to radiate sound. The dynamics of strings and circular membranes are relatively well known, relative, that is, to heterogeneous wood structures of irregular shape as found, e.g., in violins, guitars, etc. (Fletcher, 1998). It is also one of the few instruments that has many adjustable setup parameters, the adjustment of which make significant differences in tone. For example, the tension in the head (see Fig. 1) is easily changed, and heads of different thickness and bridges of different mass are easily replaced. This is undoubtedly part of the reason banjo players and technicians are frequently changing settings and parts to achieve better tone, and a large body of ''black magic'' has grown in the art of banjo setup.

There appears to be no previously published analytical model of the structural dynamics and sound radiation of the banjo.

It is hoped that the present results will provide some insight and perhaps guidance relating to the role of various materials and setup adjustments for the banjo. As a minimum, the present study confirms much of the folklore and black magic associated with banjo setup. It should be noted that the black magic is highly evolved and detailed, and there are a great many subtle nuances in tone and associated notions among banjo technicians that are beyond the scope of the present analysis.

The model is a structure composed of a set of connected one-dimensional dynamic, wave-bearing systems (Maidanik and Maga, 1986). A one-string model, with one system representing the string, and one representing the head, is developed first. Then, a five-string model is developed with six interacting dynamic systems, one for each string and one for the head. In both models, one string is driven harmonically by a unit force, and a linear structural impulse response matrix is developed for the composite and used to assess the velocity response. The matrix is of dimension two for the one-string model, there being two dynamic systems involved (the string and the head), and of dimension six for the five-string model. Propagating wave solutions are assumed in the strings and head, and linear governing equations of motion for these systems are tacitly assumed. The parameters for and general description of the model are presented in Sec. II, and the mathematical aspects of the model are presented in the Appendix.

For both the one- and five-string models, the spatial/frequency head velocity response and the radiated power as functions of frequency and time are presented.

It should also be noted that good tone is a subjective notion; however, there is general agreement, at least among bluegrass banjo players, that there is a correlation between the intrinsic loudness of the instrument and quality of tone. *Intrinsic* here is used to refer to the loudness with moderate excitation and not loudness derived by overdriving. The harmonic content or the spectral envelope (Rossing, 1994) is an important tonal quality that can be varied during performance and also varied by the setup of the instrument. Some players prefer warm tones, while some prefer the crisp tones relatively rich in higher harmonics. The banjo is typically played in a staccato fashion, and for this style, there is also general agreement that a relatively fast decay of the tone is desirable. Accordingly, figures of merit, FOMs, are defined for: (1) total radiated power; (2) slope (in amplitude/frequency) of the average power spectral density curve; and (3) the temporal decay rate for a pulse excitation. These FOMs are respectively, *loudness*, *brightness*, and *decay*, and are computed and discussed for the five-string model. The behavior of the FOMs, calculated from the model, and displayed as results, are generally in agreement with conventional banjo setup practices.

a)Electronic mail: dickey@jhu.edu

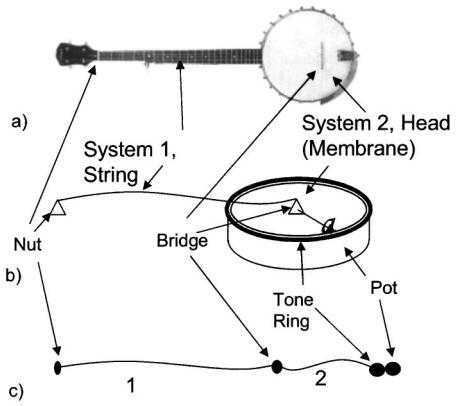


FIG. 1. The banjo simplified. (a) an actual banjo; (b) the one-string caricature defining the dynamic systems; and (c) the model.

II. THE MODEL

The dynamic systems used to model the banjo are defined in Fig. 1. In the model, system 1 is driven at a point harmonically by a unit amplitude external force. The distance of the drive from the bridge is a variable in the study. The *in situ* input admittance is calculated and used to convert the force drive to a velocity; and velocity is the dynamic quantity propagated in the model. The string, of length L_1 and linear density ρ_1 supports a single nondispersive transverse wave of speed c_1 and propagation loss factor η_1. It is terminated on one end by the *nut* having a mass-like impedance. This determines the reflection coefficient for velocity. It is terminated at the other end by being connected, through a mass-like *bridge*, to the center of a circular membrane under tension. The bridge mass is a variable in the study. The perimeter of the head is drawn over the pot and held in place with a tensioning hoop. The hoop is in turn drawn down by a series of hooks, typically 24, which provides the tension. This membrane, referred to as the *head*, constitutes system 2 and is characterized by a surface mass density, ρ_2 tension, T_{H}, and propagation loss factor, η_2. The head tension and propagation loss factor are also variables in the study. The frame supporting the membrane is known as the *pot* and is generally laminated wood and may, or may not, have a metal *tone ring*, typically bronze, between it and the membrane. The pot and tone ring, if present, are modeled as mass-like impedances without loss.

For specificity, the one string on this banjo is taken to be the ''first,'' or bottom, string on the traditional five-string banjo (Note that in Fig. 1a, the banjo is ''upside down''.) As such, it is taken to be of standard gauge and the tension is adjusted to give the fundamental resonance to be the traditional $D=293$ Hz. The set of ''standard'' parameters used for the one-string banjo are given in Table I. The parameters and standard values for the variables listed in Table I are, in most cases, measured from an actual banjo (the author's) and considered to be typical for bluegrass-style banjos. The propagation loss factors in the head and strings were assumed because reasonable string losses are inconsequential and head-loss data could not be found. The strings used as standard were D'Addario light gauge and the densities were calculated based on the published diameters. Since frequency is adjusted in the process of tuning, the string tension is not needed; however, the fundamental frequencies, f_n, determine the wave speeds, c_n, on the string using $\sin(2\pi f_n L_n/c_n)=0$, and c_n, along with ρ_n, are then used to calculate the string impedance.

TABLE I. The standard values for variables and parameters.

	Variable	Standard value
m_{b}	Bridge mass	2.5×10^{-3} kg
f_{H}	Head fundamental frequency	380 Hz
η_{H}	Propagation loss factor in head	0.2
x_1'	Drive position on string	0.1 m
	Parameter	**Value**
m_{N}	Nut mass	0.01 kg
String parameters		
$\mathrm{L}_{1,2,3,4}$	Length	0.686 m
$\eta_{1,2,3,4}$	Propagation loss factor	0.01
f_1	Fundamental frequency	293 Hz
ρ_1	Density	3.01×10^{-4} kg/m
f_2	Fundamental frequency	247 Hz
ρ_2	Density	3.7×10^{-4} 1kg/m
f_3	Fundamental frequency	196 Hz
ρ_3	Density	6.30×10^{-4} kg/m
f_4	Fundamental frequency	146.8 Hz
ρ_4	Density	1.49×10^{-3} kg/m
L_5	Length	0.513 m
η_5	Propagation loss factor	0.01
f_5	Fundamental frequency	392 Hz
ρ_5	Density	3.01×10^{-4} kg/m
Head parameters		
a	Radius	0.127 m
ρ_2	Surface density	0.24 kg/m^2
$\mathrm{T_H}$	Tension (calculated)	4470 N/m
c_2	Transverse wave velocity (calculated)	126 m/s
Pot parameters		
m_{TR}	Tone ring mass density	1.54 kg/m
m_{P}	Pot mass density	0.635 kg/m

Proper head tension is not quite as straightforward as string tension; indeed, head tension is one of the great mysteries in the art of banjo maintenance and is therefore one of the variables used in this study. Many banjo technicians will assert that proper head tension is such that the head resonates at ''about G, or a little less'' (Smith, 2002). This G is at 392 Hz and the head tension, T_{H}, which gives the fundamental frequency of 380 Hz with a conventional mylar head will be used as a reference value. The value for T_{H} is calculated by assuming the lowest eigenmode of the clamped membrane to be at $f_0=380$ Hz. That is

$$J_0(k_0a)=0,$$

$$\Rightarrow k_0a=2.4048=2\pi f_0a/c,$$

$$T_H=c^2\rho_2,$$

where J_0 is the zeroth-order Bessel function, k_0 is the radial wave number on the head, and $c\equiv c_2$ of Table I. The tone ring mass density was taken from catalog descriptions of typical banjo tone rings and is assumed to be lossless here.

The velocity response waves generated on the string by the drive travel away from the source in both directions,

reflect and transmit into system 2, and generally cause the entire complex to reverberate. The transverse velocity distribution on the head is calculated and combined with a calculated radiation efficiency to obtain the radiated power, $\Pi_{\mathrm{rad}}(\omega)$. The $\Pi_{\mathrm{rad}}(\omega)$ is then integrated over frequency to obtain a total relative radiated power level, $\tilde{\Pi}_{\mathrm{rad}}$. No attempt is made to reference this relative power to known benchmarks for perceived loudness; however, it is assumed that variations in both the $\tilde{\Pi}_{\mathrm{rad}}$ calculated here and in the perceived loudness resulting from changing model parameters are valid and useful. Accordingly, we define the FOM associated with loudness as $\tilde{\Pi}_{\mathrm{rad}}$. The $\Pi_{\mathrm{rad}}(\omega)$ is also used to derive an FOM associated with the brightness of the tone. For this, we use the slope of a straight line fitted to the $\Pi_{\mathrm{rad}}(\omega)$, in a least-squares sense, relative to the value of this quantity for the standard conditions of Table I.

The transient response of the structure is derived from the Fourier transform of the frequency response. The resulting temporal distribution is then fitted, in a least-squares sense, with an exponential curve, and the decay constant of this curve is taken as a measure of the decay of the tone and assigned as the decay FOM.

The major assumptions are

(1) *Bridge placement*. The head is driven at the center. In most modern banjos, the bridge, and therefore the drive point for the head, is placed about two-tenths of the head radius away from the center of the head in the direction away from the nut. Centering the bridge, as done in the model, will clearly influence the modal pattern on the head, but placing it at its traditional position would complicate the analysis and is not incorporated in the model. An additional simplifying assumption in the model is that the bridge does not rotate and transmit a moment excitation to the membrane.
(2) There is no *tailpiece*. The strings on a banjo extend over the bridge and attach to a *tailpiece* which is attached to the pot. The strings thus bend downward as they pass over the bridge and exert some constant downward force keeping the bridge firm on the head. On many banjos, the *tailpiece* provides some additional downward pressure on the strings between the bridge and rim. This is not part of the model.
(3) *The head is a membrane*. The model assumes a nondispersive transverse wave propagating in the head.
(4) *Fluid loading*. The fluid loading is not considered in deriving the response of the head. The transverse wave speed on the head is subsonic with respect to the air, and so the radiation impedance is small and is considered to have a small effect on the membrane dynamics and less effect on the FOMs.
(5) *Drive polarization*. The string is driven and responds in one polarization only, namely in the plane normal to the head.

III. RESULTS

The model described in general terms in the previous section, and in more detail in the Appendix, was used to calculate the response of both the one- and five-string banjos.

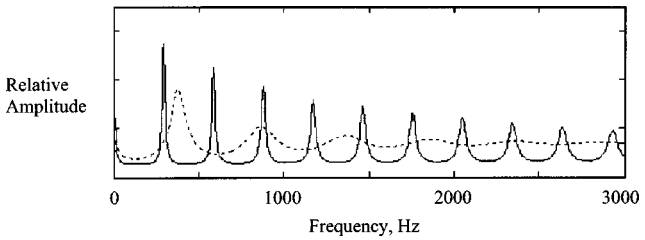


FIG. 2. Resonance response characteristics of the two isolated systems in the one-string banjo; the string, tuned to 293 Hz, solid line, and the head, tuned to 380 Hz, dashed line.

Results that are independent of the number of strings or procedures that pertain to both are discussed first; then, results specific to the one- or the five-string banjo are presented in subsections A and B, respectively.

The resonance characteristics of the string and the head in isolation are presented first in Fig. 2 to aid in the interpretation of results for both the one- and five-string banjos. The family of string resonances starts with the fundamental at 293 Hz and forms an harmonic progression. The family of head resonances starts with the fundamental at 380 Hz and progresses as the series of frequencies satisfying $J_0(k_n a) = 0$ (Morse, 1968). One would expect relatively large response from the banjo at frequencies where string and head resonances coincide.

The velocity distribution of transverse waves on the head is calculated for both the one- and five-string cases. These data are presented in subsections A and B below in three-dimensional plots of velocity amplitude magnitude versus frequency and radial position. Radiated power is given by

$$\Pi_{\mathrm{rad}}(\omega) = (\rho c)_{\mathrm{air}} \sigma(\omega) \int dS |v(x_2, \varphi, \omega)|^2, \tag{1}$$

where $\sigma(\omega)$ is the radiation efficiency of the surface, and the integration is over the area of the head. Radiation efficiency may be defined easily in terms of the wave number variable, κ, the Fourier conjugate of x_2, as the ratio of the radiating (supersonic) portion of the spectrum to the entire spectrum. Specifically (Maidanik *et al.*, 1991)

$$\sigma(\omega) = \frac{\int_0^{\omega/c} \kappa d\kappa |V(\kappa, \omega)|^2 [1 - (\kappa c/\omega)^2]^{-1/2}}{\int_0^{\infty} \kappa d\kappa |V(\kappa, \omega)|^2}. \tag{2}$$

The model gives the spatial velocity distribution, $v(x_2, \omega)$, directly, and it is computationally efficient to numerically transform this into $V(\kappa, \omega)$ and calculate the radiation efficiency using Eq. (2). The result is shown in Fig. 3. It should

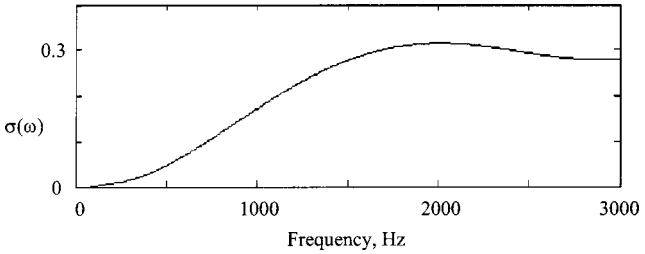


FIG. 3. Radiation efficiency of the banjo head.

be noted that the results shown are approximate in two respects. First, the numerical fast Fourier transform, FFT, implemented in the calculation left discontinuous data and Fig. 3, shows these data after smoothing. Second, the spatial integration needed to be extended for several head radii in order to extend the wave number spectrum to the range of interest, and the extension adopted was to consider the head to be in a rigid baffle.[1] The radiation efficiency is small since the transverse wave on the head is subsonic.

The smoothed radiation efficiency was then used to calculate radiated power in accordance with Eq. (1) and these results are shown in subsections A and B.

The temporal characteristics were investigated by taking the Fourier transform of the radiated power distribution. This gives $\pi_{\mathrm{rad}}(t)$, and, here again, the discretely calculated nature of the spectral data and the FFT resulted in temporal data with a lot of artificial structure. Rather than smooth the data, as was done in the wave number transform, an exponential curve was fitted, in a least-squares sense, to the temporal data. The raw transformed data and the exponential fitted curve are plotted together and shown for the one- and five-string banjos in subsections A and B.

Fianally, FOMs are defined and calculated only for the five-string banjo. These results are presented in subsection B.

The results for a one-string banjo are easier to interpret than those for the five-string, and are presented first.

A. The one-string banjo

This expectation of high response where string and head resonances coincide is borne out in the data of Fig. 4(a), where the magnitude of the head surface velocity is presented as a function of the head radial coordinate, x_2, and frequency. The first and third string resonances are reinforced by head resonances and are clearly more pronounced than the second string resonance that falls close to an head antiresonance. One can clearly see the head mode shapes in this data; the first two string resonances excite the first head mode and the third string resonance excites the second head mode. The response tends to be small at the rim because the impedance of the tone ring and pot are substantially higher than the membrane.

The head velocity distribution is then used to compute the radiated power in accordance with Eq. (1). These results are shown in Fig. 4(b). The string resonances dominate the response, but the modulation superposed on the series of string resonances by the head resonances is evident. The logarithmic scale used here shows harmonic structure at higher frequencies than can be discerned in the linear plot of Fig. 4(a). The fundamental and first six overtones of the first string are shown in the figure and numbered for reference. Note that $\Pi_{\mathrm{rad}}(\omega)$ is richer in high frequencies than is the velocity amplitude because of the efficiency with which the higher frequencies radiate. Note also the relatively low response near the sixth overtone (labeled 7 in the figure) resulting from a mismatch in the resonance characteristics of the string and the head. There is discussion in the classic literature on the contributions by the various overtones to the timbre of musical instrument sounds (Jeans, 1938), and it seems generally accepted that contributions from the sixth

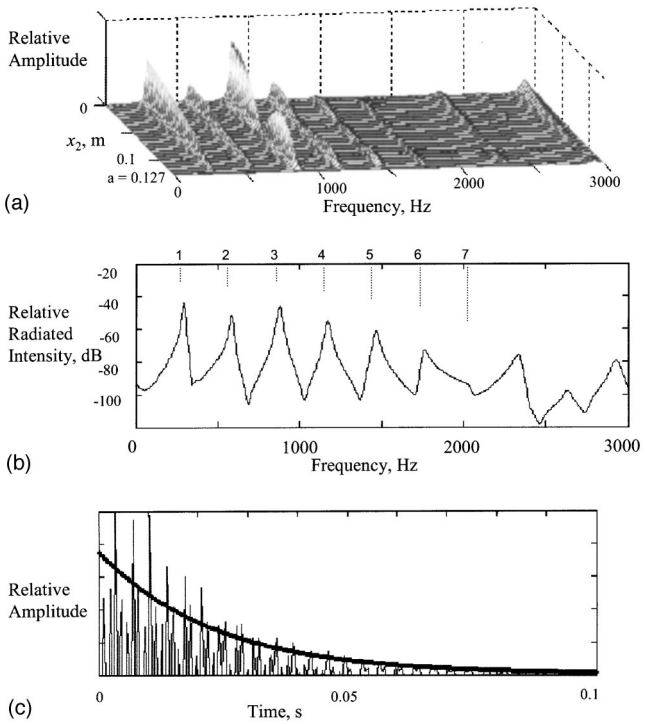


FIG. 4. (a) The velocity amplitude on the one-string banjo head as a function of the radial position, x_2, and frequency. (b) Radiated power spectrum of the sound radiated by the one-string banjo. The series of numbers associated with the dashed markers indicate the frequencies of the first seven string resonances. (c) The relative radiated power as a function of time, $\Pi_{\mathrm{rad}}(t)$, as derived from $\Pi_{\mathrm{rad}}(\omega)$. The bold line is the least-squares exponential approximation.

overtone and beyond add dissonance and are undesirable. The data are not presented, although it should be noted that the diminution of the sixth overtone does not seem to be very sensitive to setup parameters.

The temporal characteristics for the one-string banjo are presented in Fig. 4(c). The fitted curve is shown as the heavy line superposed on $\pi_{\mathrm{rad}}(t)$.

B. The five-string banjo

Results analogous to the one-string banjo are presented in Fig. 5. The spatial–temporal velocity magnitude distribution on the head, Fig. 5(a), shows considerably more complexity, and, importantly, generally higher amplitude [The scales are the same in Figs. 4(a) and 5(a)]. The higher amplitude results from the other strings being driven sympathetically; these strings, in turn, add considerable drive to the head.

The increased drive to the head and resulting response of the five-string vs the one-string banjo is demonstrated by the radiated power spectrum shown in Fig. 5(b). This curve is plotted to the same scale as Fig. 4(b) and the increase in radiated power is evident. There is also evidence of other string resonances in the detail of this spectrum. Two of the more prominent ones, the fundamental and first overtone of the second string (tuned to 247 Hz), are identified in the figure. As in the one-string banjo, the sixth overtone is conspicuously low.

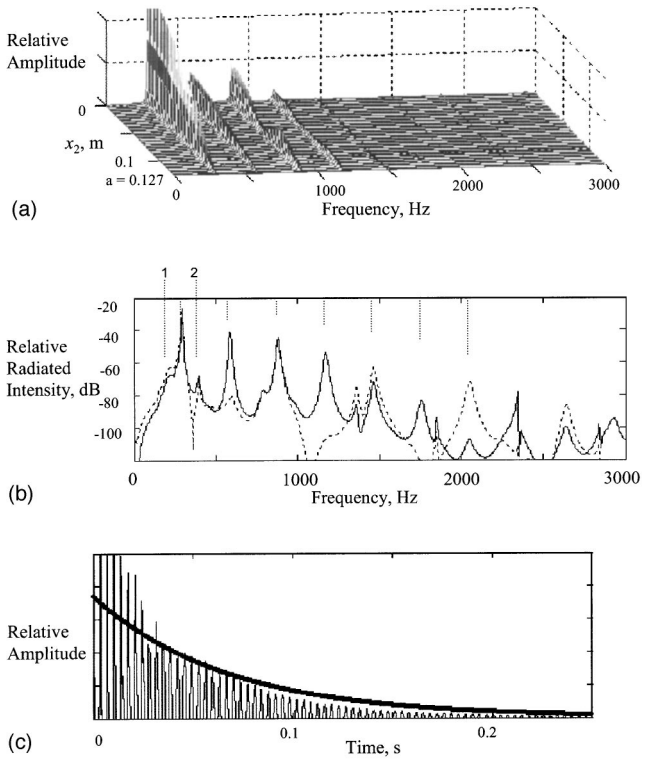


FIG. 5. (a) The velocity amplitude on the five-string banjo head as a function of the radial position, x_2, and frequency. (b) Radiated power spectrum of the sound radiated by the five-string banjo. The first seven *in situ* string resonances for the first, i.e., the driven string, are indicated as before and the first two resonances of the second string are indicated and numbered. In the dashed curve, the string is driven at its midpoint. (c) The relative radiated power as a function of time, $\Pi_{\mathrm{rad}}(t)$, as derived from $\Pi_{\mathrm{rad}}(\omega)$. The bold line is the best-fit exponential approximation, and the decay constant for this is the decay FOM.

An example of some interest is when the string is driven at its midpoint. The power spectrum for this case is shown as the dashed line in Fig. 5(b), and one notes the absence of the odd harmonics, i.e., the modes that would have a node at the string's midpoint.

The time-domain response for the five-string banjo is shown in Fig. 5(c). Here, it can be seen that the decay of the five-string banjo is significantly less than the one-string banjo; i.e., it rings longer. One proposed explanation for this is that there are more strings and thereby more string modes in the five-string case, and since the strings are less damped than the head, there is a higher proportion of lightly damped modes excited in the five-string case. In both the one- and five-string cases, there appears to be more than one decay constant: an initial fast decay, followed by a slower one. This has been observed in guitars and pianos and attributed to energy shifting between string modes of different polarizations or shifting to other string modes. (Fletcher, 1998, pp. 255 and 384; and Weinreich, 1977). Only one polarization is allowed here and a compound decay constant may be the result of energy being distributed among other strings and the head.

C. The figures of merit

The *loudness FOM* is calculated as the integral of the radiated power distribution, Fig. 5(b), over the frequency range (0,3000) Hz normalized by its value at the standard conditions listed in Table I. The values of all the FOMs are adjusted to be unity under standard conditions. A positive *loudness FOM* indicates that the banjo is louder.

The *brightness FOM* is evaluated by calculating a least-squares fitting of a straight line to the radiated power level on the interval (0,3000) Hz. The slope of this line, normalized by its slope under standard conditions, defines the FOM. Note that both the slope and its normalization are typically negative, making the ratio a positive quantity that increases with the proportion of higher frequencies in the spectrum.

The *decay FOM* is defined as the negative of the decay constant of the exponential curve superimposed on the temporal response in Fig. 5(c). The FOM is normalized by its value under standard conditions, and an increase in the FOM indicates a faster decay of the tone, i.e., less sustain.

The dependence of the three FOMs (*loudness*, *brightness*, and *decay*) on four parameters is investigated and presented in Fig. 6. The FOMs are intended to give a qualitative indication of the effect of the parameters on the relevant qualities of banjo tone. They were investigated for the five-string banjo only. Note that the assumed standard values of the parameters under investigation are indicated by vertical dashed lines in the figures, and that the FOMs at these values are, by definition, unity.

1. The propagation loss in the head, Fig. 6(a)

Most modern banjo heads are made of Mylar and some traditional heads are still made of animal skins. Mylar and skin heads are probably at the low and high end, respectively, of the loss factor range used here, and propagation loss factor was chosen as a parameter because there are several modifications to Mylar heads designed to simulate skin. In fact, most Mylar heads are frosted and some banjo technicians maintain that unfrosted heads are brighter. Assuming that the frosting adds slightly to the loss factor, this contention is borne out in Fig. 6(a). Here, it can be seen that there is relativly little change in *loudness* accompanied by significantly faster *decay* and loss of *brightness* with increasing loss factor. The implication is consistent with the commonly held belief that skin heads produce warmer (less bright) sounds than Mylar (especially unfrosted).

2. The fundamental frequency of the head, Fig. 6(b)

The dependence of *loudness* and *brightness* on head fundamental frequency, or tension, is in agreement with the commonly held notion that tighter heads produce louder and brighter sounds. No explanation is proffered for the complicated behavior of *decay* with variations in head tension.

3. The mass of the bridge, Fig. 6(c)

Note the scale on the abscissas of Fig. 6(c) and that the FOMs are less sensitive to bridge mass than to the other parameters. The data show a moderate increase in *decay* and *loudness*, and a slight decrease in *brightness* with increasing bridge mass. The increase in loudness is not in agreement with experience, and the effect of bridge mass is generally lower than might be expected.

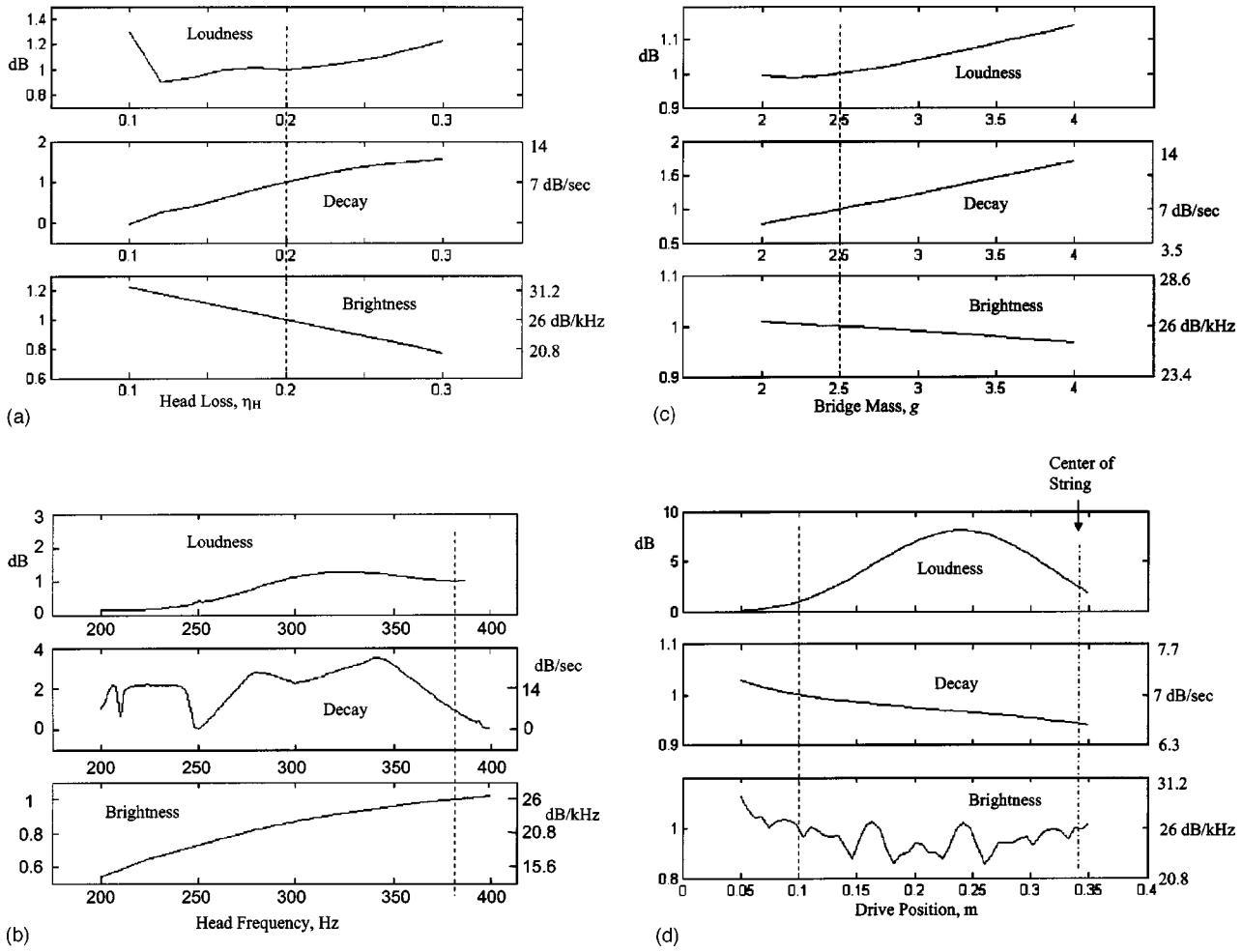


FIG. 6. The figure-of-merits, FOMs. The vertical dashed lines represent the standard parameter values. (a) The FOMs as a function of propagation loss factor in the head. (b) The FOMs, as a function of head tension as indicated by the frequency of the fundamental mode of the head. (c) The FOMs, as a function of the mass of the bridge. (d) The FOMs, as a function of the drive position measured from the bridge.

4. The position along the string where it is driven, Fig. 6(d)

Changing the plucking or strumming position along the string is a common technique among banjo players for altering the tone of the instrument. The tone gets warmer as the player moves away from the bridge. This is demonstrated in Fig. 6(d). It is interesting to note that *brightness* is a sensitive and complicated function of the distance of the excitation from the bridge. *Brightness* and *loudness* variation with drive position are the most highly varying FOMs studied. Plucking or strumming the banjo very close to the bridge, within 5 cm, produces a crisp tone with many high harmonics and a rapid decay, whereas strumming it at 15 cm, near where the neck joins the pot, produces a warm and louder tone with relatively few harmonics. Also note that *loudness* increases significantly at about 0.2 m, where the string's first overtone has an antinode. Note also the general symmetry of the *loudness* FOM about this antinode position and that the drive (recalling that it is unit force) produces very little response when applied close to the bridge or at the center of the string. The midpoint of the string is indicated in the figure.

IV. CONCLUSIONS

The structural dynamic model of the five-string banjo reinforces many of the commonly held notions regarding tone among banjo players and luthiers. Figures-of-merit are defined for several tonal qualities and the variation of these FOMs with several setup adjustments is investigated. Specifically

(1) *Loudness*—for a loud tone, use high head tension and play near where the neck joins the pot.
(2) *Brightness*—for a brighter tone, use a head material with low propagation loss (e.g., unfrosted Mylar), use high head tension, and pluck or strum close to the bridge.
(3) *Decay*—for fast decay, use a head material with high propagation loss (e.g., frosted Mylar or even animal skin), and use a heavy bridge.

Other indications from the model are the complicated relationship and sensitivity of *decay* to slight variations in head tension and of *brightness* to slight variations in the location where the string is plucked.

The major assumptions in the model are (1) the mem-

brane is driven at a point at its center; (2) there is no tailpiece; (3) the head is a membrane responding in flexure and the radiation loading does not influence the response; and (4) the string is driven and responds only in the polarization normal to the head. All of these assumptions clearly have an influence on banjo tone. For this reason, the goals and results of this study were to qualify changes in tone resulting from various material and setup changes, not to describe the tone in any absolute sense.

The indications derived from the model are generally consistent with commonly observed setup practices; however, it should be stated that there are a great many material and setup choices outside the scope of the model. Some of these may some day be shown to be important influences on banjo tone.

ACKNOWLEDGMENTS

The author is indebted to Dr. Ray Wakeland (a colleague and fellow banjo player) for helpful discussions; to Charles Buice (senior editor with the Paris Review) for much needed help with the language; and to the editor and reviewers for many constructive comments.

APPENDIX: THE MATHEMATICS

The model is a connected set of one-dimensional wave-bearing systems. Each string and the head constitute separate systems. Thus, the one-string banjo is a two-system structure and the five-string banjo is a six-system structure. The response of the structure is assumed time stationary and linear, and is formulated in terms of an impulse response matrix, $\underline{\mathbf{g}}$, which is a matrix of dimension equal to the number of systems.[2] We have

$$\mathbf{v}(\mathbf{x},\omega)=\int d\underline{\mathbf{x}}'\,\underline{\mathbf{g}}(\mathbf{x}|\mathbf{x}',\omega)\mathbf{v}(\mathbf{x}',\omega), \tag{A1}$$

where $\mathbf{v}$ and $\mathbf{x}$ are vectors of dimension two or six (for the one-string or five-string banjo, respectively), ω is the frequency, and $d\underline{\mathbf{x}}'=(dx_i'\,\delta_{ij})$, the integrating differential formed into a diagonal matrix compatible with the integrand. The elements of $\mathbf{x}$ are specific values of the coordinates at which the response is assessed, and similarly, $\mathbf{x}'$ is the vector containing information about which system(s) and where within the system(s) the structure is driven. The elements of the drive vector, $\mathbf{v}(\mathbf{x}',\omega)$, give the spatial and frequency distribution of the drive. In the present example, only system 1 is driven and this drive is harmonic, independent of the response, of unit force, and applied at the single point, x_1', giving a velocity response at the drive point of

$$\begin{aligned}&\mathbf{v}(\mathbf{x}',\omega)=\{\mathrm{v}_1(\omega),0\},\\&\mathrm{v}_1(\omega)=V_1^0(\omega)\,\delta(x_1-x_1')e^{i\omega t},\\&V_1^0=g_0P_1^0=g_0,\end{aligned} \tag{A2}$$

g_0 being the drive point admittance, and $P_1^0=1$. We formulate $d\underline{\mathbf{x}}'$ as a diagonal matrix, and the impulse response matrix, $\underline{\mathbf{g}}$, relates every point in the drive coordinates to every point in the response coordinates. Further, $\underline{\mathbf{g}}$ is dimensionless in the formulation used here and thus the response vector, $\mathbf{v}$, is velocity.

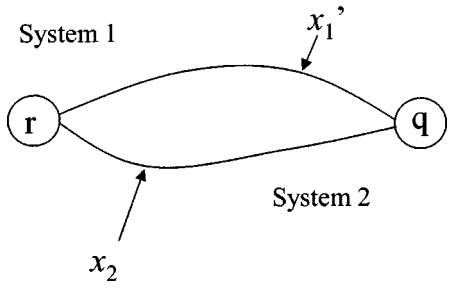


FIG. 7. The most general connection of two systems.

The impulse response matrix is developed from wave considerations and accounts for all the reverberant interactions within and among systems. Figure 7 shows the most general connection for two systems. The six-system diagram would simply have six systems in place of two; the slight loss of generality in using only two systems will be addressed when it occurs. For specificity, we designate system 1 as the string and system 2 as the head (refer to Fig. 1) and the r junction as the connection of these through the bridge. The q junction then represents the peghead end of the string and the outer periphery of the head. Since there is no direct connection between the peghead end of the string and the head, the figure can be redrawn, as in Fig. 1(c).

Since we have assumed linearity, superposition of solutions allows us to write,

$$\underline{\mathbf{g}}=\underline{\mathbf{g}}_{\mathrm{direct}}+{}_q\underline{\mathbf{g}}_r+{}_r\underline{\mathbf{g}}_r+{}_r\underline{\mathbf{g}}_q+{}_q\underline{\mathbf{g}}_q,$$

where the subscript ''direct'' indicates the component which is generated by the source at x' and travels directly, without any interaction at boundaries, to the point x. This direct component is zero in the present model since the drive and response are always in separate systems, i.e., string and head, respectively. For the other four components, ${}_\beta\underline{\mathbf{g}}_\alpha$, represents the component of response which originated from the drive going towards the α junction, and arriving at the assess point coming from the β junction. All components have a time dependence of $e^{i\omega t}$ which has been omitted for brevity. We have

$$\begin{aligned}&{}_q\underline{\mathbf{g}}_r=\underline{\mathbf{t}}(\mathbf{x}|\mathbf{x}_q)\underline{\mathbf{Q}}\underline{\mathbf{t}}(\mathbf{x}_q|\mathbf{x}_r)\underline{\mathbf{R}}\underline{\mathcal{D}}_r\underline{\mathbf{t}}(\mathbf{x}_r|\mathbf{x}'),\\&{}_r\underline{\mathbf{g}}_r=\underline{\mathbf{t}}(\mathbf{x}|\mathbf{x}_r)\underline{\mathbf{R}}\underline{\mathcal{D}}_r\underline{\mathbf{t}}(\mathbf{x}_r|\mathbf{x}'),\\&{}_r\underline{\mathbf{g}}_q=\underline{\mathbf{t}}(\mathbf{x}|\mathbf{x}_r)\underline{\mathbf{R}}\underline{\mathbf{t}}(\mathbf{x}_r|\mathbf{x}_q)\underline{\mathbf{Q}}\underline{\mathcal{D}}_q\underline{\mathbf{t}}(\mathbf{x}_q|\mathbf{x}'),\\&{}_q\underline{\mathbf{g}}_q=\underline{\mathbf{t}}(\mathbf{x}|\mathbf{x}_q)\underline{\mathbf{Q}}\underline{\mathcal{D}}_q\underline{\mathbf{t}}(\mathbf{x}_q|\mathbf{x}'),\\&\underline{\mathcal{D}}_r=(1-\underline{\mathbf{t}}(\mathbf{x}_r|\mathbf{x}_q)\underline{\mathbf{Q}}\underline{\mathbf{t}}(\mathbf{x}_q|\mathbf{x}_r)\underline{\mathbf{R}})^{-1},\\&\underline{\mathcal{D}}_q=(1-\underline{\mathbf{t}}(\mathbf{x}_q|\mathbf{x}_r)\underline{\mathbf{R}}\underline{\mathbf{t}}(\mathbf{x}_r|\mathbf{x}_q)\underline{\mathbf{Q}})^{-1},\end{aligned} \tag{A3}$$

where the $\mathbf{t}$ matrices propagate signals from one point to another. For example, the $\underline{\mathbf{t}}(\mathbf{x}_\alpha|\mathbf{x}_\beta)$ matrices propagate signals the entire length of the systems, i.e., from x_β to x_α. Other $\mathbf{t}$ matrices in the equations propagate signals a lesser distance in accordance with their arguments. All the $\mathbf{t}$ matrices are diagonal and the element number describes the propagation in that system. Noting that system 1 is a string

with plane-wave type propagation, we have for the (1,1) element

$$\mathrm{t}_{1,1}^{+,-}(\mathbf{x}_\alpha|\mathbf{x}_\beta)=e^{(-ik_1|x_\alpha-x_\beta|)},\quad k_1=\frac{\omega}{c_1}(1-i\eta_1), \tag{A4a}$$

where the absolute value sign insures that propagation in either direction is the same, i.e., $t^+=t^-$, and that a positive propagation loss factor, η_1, insures that the response decays as it propagates. The factors $\underline{\mathcal{D}}_\alpha$ contain the resonances in the structure. The form $(1-\xi)^{-1}$ arises from the identity $\Sigma_{n=0}^{\infty}\xi^n=(1-\xi)^{-1}$, with $\xi=\underline{\mathbf{t}}(\mathbf{x}_q|\mathbf{x}_r)\underline{\mathbf{R}}\underline{\mathbf{t}}(\mathbf{x}_r|\mathbf{x}_q)\underline{\mathbf{Q}}$ accounting for one complete reverberation, ξ^2 accounting for two, and so on.

1. The one-string banjo

For the one-string banjo, the head is designated as system 2. The $t_{2,2}^{+,-}$ elements, or propagation in system 2 is not plane wave, and the propagators there for the outgoing and incoming waves, t^+ and t^-, respectively, are

$$\begin{aligned}
&\mathrm{t}_{2,2}^{+}(x_{2,\alpha}|x_{2,\beta})=\frac{H_0^{(1)}(k_2x_{2,\alpha})}{H_0^{(1)}(k_2x_{2,\beta})},\qquad (x_{2,\alpha}>x_{2,\beta}),\\
&k_2=\frac{\omega}{c_2}(1-i\eta_2),\\
&t_{2,2}^{-}(x_{2,\beta}|x_{2,\alpha})=\frac{H_0^{(2)}(k_2x_{2,\beta})}{H_0^{(2)}(k_2x_{2,\alpha})}.
\end{aligned} \tag{A4b}$$

The $\underline{\mathbf{R}}$ and $\underline{\mathbf{Q}}$ matrices in Eq. (A3) are the reflection and transmission coefficients for the structure. The diagonal elements of these are the amplitude reflection factors (for velocity) when a wave strikes a junction and reflects back into the same system, and the off-diagonal elements are the transmission factors when a wave passes, through a junction, from one system to another. The resonances of the structure are contained in the (near) singularities of the factors $\underline{\mathcal{D}}_\alpha$. The reflection factor for a wave in system 1 (the string), when it strikes the nut is

$$\begin{aligned}
&Q_{1,1}=\frac{z_1-z_{\mathrm{nut}}}{z_1+z_{\mathrm{nut}}},\\
&z_1=\rho_1c_1,\quad z_{\mathrm{nut}}=i\omega m_{\mathrm{nut}},
\end{aligned} \tag{A5a}$$

where (ρ_1,c_1) are the linear density and transverse wave speed in the string, and m_{nut} is the mass of the nut. Since there is no direct connection between systems 1 and 2 through the q junction, we have $Q_{1,2}=Q_{2,1}=0$. The remaining element, $Q_{2,2}$, is the reflection of the outgoing wave on the head from the pot. The reflection and transmission factors at this boundary are similar in form to Eqs. (A5) but complicated somewhat by the circular geometry of system 2. It should first be noted that the impedances of an outgoing and an ingoing wave in circular coordinates are

$$z^+=i\sigma c\frac{H_0^{(1)}(k_2r)}{H_0^{(1)\prime}(k_2r)},\quad z^-=-i\sigma c\frac{H_0^{(2)}(k_2r)}{H_0^{(2)\prime}(k_2r)}, \tag{A6}$$

respectively, where σc is the characteristic impedance of the membrane and where the dash indicates the derivative with respect to the argument of the Hankel function. Thus,

$$\begin{aligned}
&Q_{2,2}=\frac{z_2-z_{\mathrm{rim}}}{z_2+z_{\mathrm{rim}}},\\
&z_2=-i\sigma_{\mathrm{head}}c_{\mathrm{head}}\frac{H_0^{(2)}(k_2a)}{H_0^{(2)\prime}(k_2a)},\\
&z_{\mathrm{rim}}=z_{\mathrm{pot}}+z_{\mathrm{tone\ ring}},\\
&z_{\mathrm{pot}}=i\omega m_{\mathrm{pot}}(1-i\eta_{\mathrm{pot}}),\\
&z_{\mathrm{tone\ ring}}=i\omega m_{\mathrm{tone\ ring}},
\end{aligned} \tag{A5b}$$

where $(\sigma_{\mathrm{head}},c_{\mathrm{head}})$ are the surface mass density and flexural wave speed on the head, m_{pot} is the linear mass density of the wooden pot, and $m_{\mathrm{tone\ ring}}$ is the linear mass density of the tone ring, if present.

The elements of $\underline{\mathbf{R}}$ are the reflection/transmission coefficients associated with the string/bridge/head connection. We have

$$\begin{aligned}
&R_{1,1}=\frac{z_1-(z_{\mathrm{bridge}}+z_{\mathrm{head}}^{+}|_{x_2\to0})}{z_1+(z_{\mathrm{bridge}}+z_{\mathrm{head}}^{+}|_{x_2\to0})},\\
&z_{\mathrm{bridge}}=i\omega m_{\mathrm{bridge}},\quad z_{\mathrm{head}}^{+,-}|_{x_2\to0}=-i\sigma_{\mathrm{head}}c_{\mathrm{head}},\\
&R_{2,1}=\frac{2z_1}{z_1+(z_{\mathrm{bridge}}+z_{\mathrm{head}}^{+}|_{x_2\to0})},\\
&R_{1,2}=\frac{2z_{\mathrm{head}}^{-}|_{x_2\to0}}{z_1+(z_{\mathrm{bridge}}+z_{\mathrm{head}}^{-}|_{x_2\to0})},\\
&R_{2,2}=\frac{z_{\mathrm{head}}^{-}|_{x_2\to0}-(z_1+z_{\mathrm{bridge}})}{z_{\mathrm{head}}^{-}|_{x_2\to0}+(z_1+z_{\mathrm{bridge}})},
\end{aligned} \tag{A7}$$

where z_{bridge} is the mass of the bridge and $z_{\mathrm{head}}^{+}|_{x_2\to0}$ indicates that the impedance is evaluated at $x_2=0$. Use has been made of

$$\operatorname*{Lim}_{r\to0}\left(\frac{H_0^{(1,2)}(k_2r)}{H_0^{(1,2)\prime}(k_2r)}\right)=-1.$$

2. Modifications for the five-string banjo

If other strings were present, as in the analysis of the five-string banjo, all matrices in the model are of dimension six. The system propagator matrices are diagonal, with

$$\mathrm{t}_{\gamma,\gamma}^{+,-}=e^{(-ik\gamma L\gamma)},$$

for $\gamma=1$ to 5 and $\mathrm{t}_{6,6}^{+,-}$ are the same as $\mathrm{t}_{2,2}^{+,-}$ for the one-string case.

The reflection factor at the peghead, Eq. (A5a), is modified to

$$Q_{\alpha,\alpha}=\frac{z_\alpha-\tilde{z}_{\text{nut}}}{z_\alpha+\tilde{z}_{\text{nut}}},\quad \tilde{z}_{\text{nut}}=z_{\text{nut}}+\sum_{\beta\neq\alpha} z_\beta, \tag{A8a}$$

in which β runs over all the strings. There now needs to be a transmission factor between strings at the peghead. This is given by

$$Q_{\beta,\alpha}=\frac{2z_\alpha}{z_\alpha+\tilde{z}_{\text{nut}}}, \tag{A8b}$$

and, again, $Q_{6,\alpha}=Q_{\alpha,6}=0$ since there is no direct communication between the strings and the head at the nut.

The modifications to the reflection and transmission factors at the bridge are similar. Specifically, Eqs. (A7) are changed to

$$R_{\alpha,\alpha}=\frac{z_\alpha-\tilde{z}_{\text{bridge}}}{z_\alpha+\tilde{z}_{\text{bridge}}},$$

$$\tilde{z}_{\text{bridge}}=z_{\text{bridge}}+\sum_{\beta\neq\alpha} z_\beta, \tag{A9}$$

$$R_{\beta,\alpha}=\frac{2z_\alpha}{z_\alpha+\tilde{z}_{\text{bridge}}}.$$

[1]The choice between using a rigid or free baffle here is not clear. If the banjo is played by a banjo player, the question becomes whether the player is rigid or soft, and how big he or she is. The choice is considered moot in light of the fact that only relative effects are considered and the choice of baffle makes little difference.

[2]The convention used here is that vectors are indicated by bold quantities, e.g., **x**, and matrices by bold underlined quantities, e.g., **g**. Thus, the expression $\mathbf{g}(\mathbf{x})$ is a matrix whose elements are vectors. As used here, the dimension of all matrices and vectors are the same, i.e., the number of systems involved. For example, the velocity response of the two system structure is $\mathbf{v}(\mathbf{x},\omega)=\{v_1(x_1,\omega),v_2(x_2,\omega)\}$, and $\mathbf{x}=\{x_1,x_2\}$, where x_1 and x_2 are scalar position coordinates.

Fletcher, N. H., and Rossing, T. D. (**1998**). *The Physics of Musical Instruments* (Springer, Berlin).

Jeans, J. (**1938**). *Science and Music* (Macmillan, New York).

Maidanik, G., and Maga, L. (**1986**). ''Response of coupled one-dimensional dynamic systems,'' J. Sound Vib. **111**, 361–375.

Maidanik, G., Dickey, J., and Ertel, J. (**1991**). ''Partial radiation efficiency of line driven panels,'' J. Sound Vib. **144**(1), 71–86.

Morse, P. M., and Ingard, U. K. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York).

Rossing, T. D., and Fletcher, N. H. (**1994**). *Principles of Vibration and Sound* (Springer, Berlin).

Smith, S. (**2002**). Personal communication.

Weinreich, G. (**1977**). ''Coupled piano strings,'' J. Acoust. Soc. Am. **62**, 1474–1484.

Erratum: "Frequency-dependent attenuation-compensation functions for ultrasonic signals backscattered from random media" [J. Acoust. Soc. Am. 111, 2308–2319 (2002)]

Michael L. Oelze and William D. O'Brien, Jr.
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801

(Received 3 July 2003; accepted for publication 4 August 2003)

 [DOI: 10.1121/1.1612491]

PACS numbers: 43.80.Vj, 43.80.Qf, 43.10.Vx [ADP] Pages: 2967–2968

The initial formulation of the problem in Ref. 1 did not separate out the coherent and incoherent scattering. While the form of the attenuation-compensation terms will not be affected, the error analysis was not complete. In reformulating the theory to separate out the coherent and incoherent scattering we start with revised Eqs. (10) and (11) from Ref. 1

$$r_{dx}(t)=s(t+2d_1/c)+s(t+2d_2/c)+s(t+2d_3/c)+\ldots \tag{10}$$

and

$$|R_{dx}(f)|=|S(f)|\times|e^{-i2\pi f(2d_1/c)}+e^{-i2\pi f(2d_2/c)}+e^{-i2\pi f(2d_3/c)}+\ldots|. \tag{11}$$

The second term on the right hand side of Eq. (11) simplifies as

$$|e^{-i2\pi f(2d_1/c)}+e^{-i2\pi f(2d_2/c)}+e^{-i2\pi f(2d_3/c)}+\ldots|^2 = \sum_{i=j=1}^{\bar{n}} 1+\sum_{i\neq j=1}^{\bar{n}} e^{-i2\pi f(2d_i/c-2d_j/c)}, \tag{11b}$$

where the first term on the right hand side is the incoherent scattering term and the second term is the coherent scattering term, and $\bar{n}$ is the number of scatterers in a unit length, dx.[2] Simplifying Eq. (11b) yields,

$$|e^{-i2\pi f(2d_1/c)}+e^{-i2\pi f(2d_2/c)}+e^{-i2\pi f(2d_3/c)}+\ldots|^2 = \bar{n}+\Phi_{\mathrm{coh}}. \tag{11c}$$

Relating Eq. (11c) to (11) gives

$$|R_{dx}(f)|=|S(f)|\bar{n}^{1/2}\left(1+\frac{\Phi_{\mathrm{coh}}}{\bar{n}}\right)^{1/2}. \tag{11d}$$

The term under the radical can be expanded as

$$\left(1+\frac{\Phi_{\mathrm{coh}}}{\bar{n}}\right)^{1/2}=1+\frac{1}{2}\frac{\Phi_{\mathrm{coh}}}{\bar{n}}-\frac{1}{8}\left(\frac{\Phi_{\mathrm{coh}}}{\bar{n}}\right)^2+\ldots=1+\Phi'_{\mathrm{coh}}, \tag{11e}$$

where $\Phi'_{\mathrm{coh}}<\Phi_{\mathrm{coh}}$ for $\bar{n}\geqslant 1$. As $\bar{n}$ (number density) becomes large, $\Phi'_{\mathrm{coh}}\ll\Phi_{\mathrm{coh}}$. Multiplying by the round trip propagation loss gives a revised Eq. (12) as

$$|R_{dx}(f)|=|S(f)|\bar{n}^{1/2}(1+\Phi'_{\mathrm{coh}})e^{-2\alpha(f)x}. \tag{12}$$

Integrating Eq. (12) over the length of the gate, L, gives

$$|R_L(f)|=|S(f)|\bar{n}^{1/2}\frac{1}{L}\int_0^L(1+\Phi'_{\mathrm{coh}})e^{-2\alpha(f)x}dx. \tag{13}$$

Equations (17)–(19) would then be rewritten as

$$|R_{dx}(f)|^2=|S(f)|^2\bar{n}(1+\Phi_{\mathrm{coh}}), \tag{17}$$

$$|R_{dx}(f)|^2=|S(f)|^2\bar{n}(1+\Phi_{\mathrm{coh}})e^{-4\alpha(f)x}, \tag{18}$$

$$|R_L(f)|^2=|S(f)|^2\bar{n}\frac{1}{L}\int_0^L(1+\Phi_{\mathrm{coh}})e^{-4\alpha(f)x}dx. \tag{19}$$

Omit (20). Equations (24)–(27) would then become

$$W_{OO}(f)=\frac{|P(f)|^2|S(f)|^2}{L^2}\bar{n}\left[\int_0^L(1+\Phi'_{\mathrm{coh}})e^{-2\alpha(f)x}dx\right]^2, \tag{24}$$

$$\epsilon_{OO}(f)=\frac{|P(f)|^2|S(f)|^2}{L}\bar{n}\left(\frac{2}{A_{OO}^{1/2}}\int_0^L\Phi'_{\mathrm{coh}}e^{-2\alpha(f)x}dx+\frac{1}{L}\left[\int_0^L\Phi'_{\mathrm{coh}}e^{-2\alpha(f)x}dx\right]^2\right), \tag{25}$$

and as αL becomes large

$$\epsilon_{OO}(f)=\frac{|P(f)|^2|S(f)|^2}{L}\overline{n}\left\langle 2\int_0^L \Phi'_{\mathrm{coh}}e^{-2\alpha(f)x}dx + \frac{1}{L}\left[\int_0^L \Phi'_{\mathrm{coh}}e^{-2\alpha(f)x}dx\right]^2\right\rangle, \tag{26}$$

$$\epsilon_{OM}(f)=\frac{|P(f)|^2|S(f)|^2}{L}\overline{n}\int_0^L \Phi_{\mathrm{coh}}e^{-4\alpha(f)x}dx. \tag{27}$$

As the number of scatterers, $\overline{n}$, increases the Φ'_{coh} term falls off much more rapidly than the Φ_{coh} term. Consequently, the error term for the Oelze and O'Brien compensation becomes smaller relative to the error term from Eq. (27) for increasing $\overline{n}$.

We are grateful to Anthony Gerig and James Zagzebski at the University of Wisconsin-Madison for bringing thiserror to our attention. This work is supported by NIH F32 CA96419 to MLO.

[1] M. L. Oelze and W. D. O'Brien, Jr., "Frequency-dependent attenuation-compensation functions for ultrasonic signals backscattered from random media," J. Acoust. Soc. Am. **111**, 2308–2319 (2002).
[2] J. F. Chen, E. L. Madsen, and J. A. Zagzebski, "A method for determination of frequency-dependent effective scatterer number density," J. Acoust. Soc. Am. **95**, 77–85 (1994).