CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 109, No. 2 February 2001

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 109, NO. 2, FEBRUARY 2001

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 109, NO. 2, FEBRUARY 2001

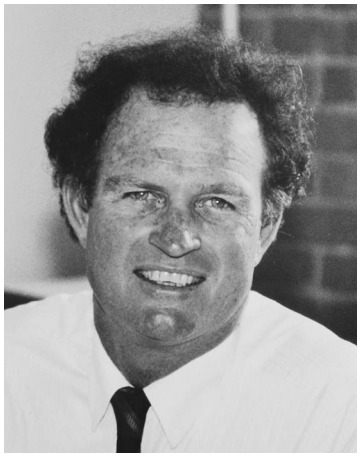
CONTENTS—*Continued from preceding page*

ACOUSTICAL NEWS—USA

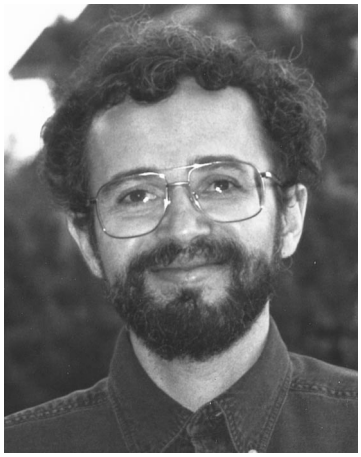
Elaine Moran
Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502

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.*

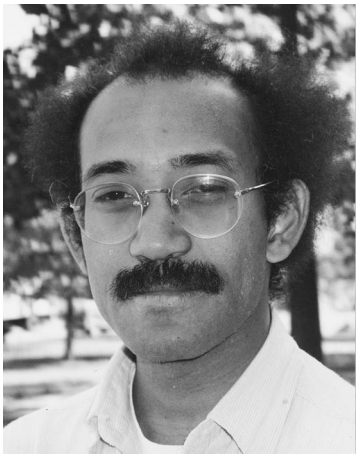
New Fellows of the Acoustical Society of America



Brian G. Fergusson—For contributions to atmospheric and underwater acoustic systems.



Oleg Godin—For contributions to wave propagation theory.



Charles Thompson—For contributions to theoretical and computational acoustics.

A way to ensure the future by honoring the past: The Acoustical Society Foundation

The Acoustical Society Foundation's sole purpose is to gather contributions and invest them in the Acoustical Society endowment so that it reaches a level large enough to fund the activities of the Society despite any changing Societal income streams in the future. A bittersweet role of the General Secretary is to assist Society members and friends in making contributions in memory of those who have gone before us. Such contributions can honor our friends and colleagues who have died, and at the same time expand the Acoustical Society endowment for our younger and future members. If donations in honor of a member who has died reach certain levels, named scholarships, prizes, or other Societal programs or initiatives can be implemented to remember that person. For instance, several years ago a scholarship in honor of Raymond Stetson was established by major contributions from Mac Pickett and Arthur Benton through the newly established Acoustical Society Foundation. This scholarship is successfully in place, with over a dozen noteworthy applicants for the scholarship each year. In addition, significant funds have been donated to the Foundation in honor of Dan Martin, who served the Society for many years, including the important position of our Journal Editor-in-Chief. More recently, contributions in honor of Dennis Klatt, Franklin Cooper, and Jack Purcell have been made.

We at the Foundation encourage you to remember and honor deceased family, friends, and colleagues in this way. For more information on ways to share in the Foundation's goals and mechanisms for making donations that allow you to give while receiving, please contact Dr. Bob Frisina at 716-275-8130 or asf@q.ent.rochester.edu. Thank you.

ROBERT FRISINA
Foundation General Secretary

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.

2001

15–17 March	Annual Meeting, American Auditory Society, Scottsdale, AZ [Wayne J. Staab, Ph.D., American Auditory Society, 512 E. Canberbury Ln., Phoenix, AZ 85022, Tel.: 602-789-0755; Fax: 602-942-1486; E-mail: amaudsoc@aol.com; WWW: www.amauditorysoc.org].
22–25 March	''New Frontiers in the Amelioration of Hearing Loss,'' St. Louis, MO [Sarah Uffman, CID Department of Research, 4560 Clayton Ave., St. Louis, MO 63110, Tel.: 314-977-0278; Fax: 314-977-0030; E-mail: suffman@cid.wustl.edu].
30 April–3 May	2001 SAE Noise & Vibration Conference & Exposition, Traverse City, MI [Patti Kreh, SAE Int'l., 755 W. Big Beaver Rd., Suite 1600, Troy, MI 48084, Tel.: 248-273-2474; Fax: 248-273-2494; E-mail: pkreh@sae.org].
4–8 June	141st Meeting of the Acoustical Society of America, Chicago, IL, Palmer House Hilton Hotel [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: 2 February 2001
7–9 June	International Hearing Aid Conference VI. Novel Processing and Fitting Strategies, Iowa City, IA [Rich Tyler, Tel.: 319-356-2471, E-mail: rich-tyler

	@uiowa.edu, WWW: www.medicine.uiowa.edu/otolaryngology/news/news].
9–13 July	2001 SIAM Annual Meeting, San Diego, CA [Society for Industrial and Applied Mathematics (SIAM), Tel.: 215-382-9800; Fax: 215-386-7999; E-mail: meetings@siam.org; WWW: www.siam.org/meetings/an01/].
15–19 August	ClarinetFest 2001, New Orleans, LA [Dr. Keith Koons, ICA Research Presentation Committee Chair, Music Dept., Univ. of Central Florida, P.O. Box 161354, Orlando, FL 32816-1354, Tel.: 407-823-5116; E-mail: kkoons@pegasus.cc.ucf.edu].
19–24 August	Asilomar Conference on Implantable Auditory Prostheses, Pacific Gove, CA [Michael Dorman, Dept. of Speech and Hearing Science, Arizona State Univ., Tempe, AZ 85287-0102: Tel.: 480-965-3345; Fax: 480-965-0965; E-mail: mdorman@asu.edu].
4–6 October	Ninth Annual Conference on the Management of the Tinnitus Patient, Iowa City, IA [Rich Tyler, Tel.: 319-356-2471; E-mail: rich-tyler@uiowa.edu; WWW: www.medicine.uiowa.edu/otolaryngology/news/news].
7–10 October	2001 IEEE International Ultrasonics Symposium Joint with World Congress on Ultrasonics, Atlanta, GA [W. O'Brien, Electrical and Computer Engineering, Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801; Fax: 217-244-0105; WWW: www.ieeeuffc.org/2001].
3–7 December	142nd Meeting of the Acoustical Society of America, Ft. Lauderdale, FL [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].

2002

3–7 June	143rd Meeting of the Acoustical Society of America, Pittsburgh, PA [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].
2–6 December	Joint Meeting: 144th Meeting of the Acoustical Society of America, 3rd Iberoamerican Congress of Acoustics and 9th Mexican Congress on Acoustics, Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org/cancun.html].

ACOUSTICAL NEWS—INTERNATIONAL

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

25th anniversary of the Instituto de Acústica, Spain

1975 marked the founding of the Instituto de Acústica, emerging from the Acoustics Department of the Center for Physical Research ''L. Tores Quevedo'' which had its origins in 1945. To commemorate the event, the Instituto de Acústica has been organizing a series of lectures covering most of the main topics of acoustics. These lectures will be given by internationally recognized specialists and will be published in a book.

The program spans about 12 months, and the lectures will be presented by scientists from the following countries: France, The Netherlands, Italy, U.K., Portugal, USA, Denmark, Sweden, and Spain. The lecture series has started in May 2000 and will be concluded in April 2001.

Proceedings available

The European Acoustics Association has published a series of proceedings of meetings held in Europe since 1996. These publications are part of the collection *Documenta Acustica* maintained by the EAA. The first of these documents is the 347-page proceedings of the International Symposium on Hydroacoustics and Ultrasonics (in English) held in 1996 at Jurata, Poland. Complete listings and information about these publications can be obtained from *Fenestra Acustica*, the name of the Web site of EAA, which is ⟨eaa.essex.ac.uk/eaa/⟩. Additional information can be obtained via e-mail from the French Acoustical Society, sfa@cal.enst.fr or by Fax to +33 1 48 88 90 60.

International Meetings Calendar

Below are announcements of meetings to be held abroad. Entries preceded by an * are new or updated listings with full contact addresses given in parentheses. *Month/year* listings following other entries refer to meeting announcements, with full contact addresses, which were published in previous issues of the *Journal*.

March 2001

26–29 **German Acoustical Society Meeting (DAGA 2001)**, Hamburg-Harburg, Germany. (e-mail: dega@aku.physik.uni-oldenburg.de) *10/00*

April 2001

9–11 **Acoustical Oceanography**, Southampton, UK. (Fax: +44 1727 850553; Web: www.ioa.org.uk) *8/00*

23–25 **1st International Workshop on Thermoacoustics**, s'Hertogenbosch. (Web: www.phys.tue.nl/index.html) *12/00*

May 2001

21–25 **5th International Conference on Theoretical and Computational Acoustics (ICTCA2001)**, Beijing. (Fax: +1 303 497 3577; Web: www.etl.noaa.gov/ictca01) *12/00*

22–25 ***8th International School on Acousto-Optics and Application**, Gdańsk-Jurata, Poland. (A. Sikorska, Institute of Experimental Physics, University of Gdańsk, ul. Wita Stwosza 57, 80-952 Gdańsk, Poland; Fax: +48 58 341 3175; e-mail: fizao@univ.gda.pl)

28–31 **3rd EAA International Symposium on Hydroacoustics**, Jurata. (Fax: +48 58 625 4846; Web: www.amw.gdynia.pl/pta/sha2001.html) *12/00*

July 2001

2–5 **Ultrasonics International Conference (UI01)**, Delft. (Fax: +1 607 255 9179; Web: www.ccmr.cornell.edu/~ui01/) *12/00*

2–6 **8th International Congress on Sound and Vibration**, Kowloon, Hong Kong. (Fax: +852 2365 4703; Web: www.iiav.org) *8/00*

August 2001

28–30 **INTER-NOISE 2001**, The Hague. (Web: internoise 2001.tudelft.nl) *6/99*

September 2001

2–7 **17th International Congress on Acoustics (ICA)**, Rome. (Fax: +39 6 4976 6932; Web: www.ica2001.it) *10/98*

10–13 **International Symposium on Musical Acoustics (ISMA 2001)**, Perugia. (Fax: +39 75 577 2255; e-mail: perusia@classico.it) *10/99*

October 2001

17–19 **32nd Meeting of the Spanish Acoustical Society**, La Rioja. (Fax: +34 91 411 76 51; Web: www.ia.csic.es/sea/index.html) *10/99*

25–26 ***Fall Meeting of the Swiss Acoustical Society**, Walllis/Valais, Switzerland. (Suva Akustik, P.O. Box 4358, 6002 Luzern, Switzerland; Web: www.sga-ssa.ch)

November 2001

21–23 ***Australian Acoustical Society Annual Meeting**, Canberra, Australia. (Acoustics 2001, Australian Defense Force Academy, Canberra, ACT 2600, Australia; e-mail: nit@adfa.edu.au)

March 2002

4–8 **German Acoustical Society Meeting (DAGA 2002)**, Bochum, Germany. (Web: www.ika.ruhr-uni-bochum.de) *10/00*

June 2002

4–6 ***6th International Symposium on Transport Noise and Vibration**, St. Petersburg, Russia. (East-European Acoustical Association, Moskovskoe Shosse 44, St. Petersburg 196158, Russia; Fax: +7 812 127 9323; e-mail: noise@mail.rcom.ru)

10–14 **Acoustics in Fisheries and Aquatic Ecology**, Montpellier. (Web: www.ices.dk/symposia/) *12/00*

August 2002

19–23 **16th International Symposium on Nonlinear Acoustics (ISNA16)**, Moscow. (Fax: +7 095 126 8411; Web: acs366b.phys.msu.su/isna) *12/00*

September 2002

16–21 **Forum Acusticum 2002 (Joint EAA-SEA-ASJ Meeting)**, Sevilla. (Fax: +34 91 411 7651; Web: www.cica.es/aliens/forum2002) *12/00*

December 2002

2–6 **Joint Meeting: 9th Mexican Congress on Acoustics, 144th Meeting of the Acoustical Society of America, and 3rd Iberoamerican Congress on Acoustics**. (e-mail: sberista@maya.esimez.ipn.mx or Web: asa.aip.org) *10/00*

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).]*

Vibrations of Shells and Rods

Khanh Chau Le

Springer, Berlin, Heidelberg, New York, 1999.
423 pp. Price: $112.00 hc ISBN: 3540645160.

Dynamics of shells, plates, and rods is an engineering problem, which involves complicated mathematics due to the fourth-order partial differential equations that describe flexural vibrations. As a result there are two types of books related to this problem: one is written by engineers or physicists and the other by mathematicians. The reviewed book is clearly written by a mathematician, and, in my view, is written for mathematicians who specialize in mechanics and structural dynamics.

This book is based on a hybrid variational-asymptotic approach that reduces the three-dimensional equations of elastic theory of plates, shells, and rods to two- and one-dimensional approximate equations of motion using small parameters. This variational-asymptotic (VA) method was developed by Professor V. L. Berdichevsky about 20 years ago studying small amplitude long-wave vibrations. The method is a synthesis of two methods that are widely used in mechanics: asymptotic and variational. The VA method starts by approximating the system Lagrangian of the functional, instead of a system of differential equations. Neglecting a small term in the Lagrangian is equivalent to neglecting several terms in the differential equations, which are not always easy to recognize as small ones. Next, approximate differential equations are obtained by varying the approximated functional. This approach simplifies the analysis and provides greater flexibility for modifying and solving the equations.

The author of the book, Professor Le, has extrapolated this method into the high-frequency short-wave range. Specifically, he applies the VA method to analyze vibrations of piezoelectric shells, plates, and rods, and devotes roughly half of the book to this topic.

The book starts with two introductory chapters that discuss the historical and mathematical background of tensor analysis, the geometry of curves and surfaces, the dynamic theory of elasticity and piezoelectricity, and the variational-asymptotic method. After this introduction the book is divided into two equal parts: the first part describes low-frequency vibrations and the second high-frequency vibrations. Both parts contain four chapters with the same titles: Elastic Shells, Elastic Rods, Piezoelectric Shells, and Piezoelectric Rods. Each section ends with problems and exercises. The bibliography consists of 62 references mostly related to the methods described in the book.

Overall, the book describes an interesting and powerful method of deriving and solving complicated equations of shell and rod vibrations. However, as a physicist and practitioner, I hestitate to recommend it to ''engineers who deal with vibrations of shells and rods in their everyday practice.'' I would rather agree that it is ''for mathematicians who seek applications of the variational and asymptotic methods in elasticity and piezoelectricity,'' as announced on the back cover of the book.

DIMITRI M. DONSKOY
CEOE Department
Stevens Institute of Technology
Hoboken, New Jersey 07030

Adaptive Structures: Dynamics and Control

Robert L. Clark, William R. Saunders, and Gary P. Gibbs

John Wiley & Sons, 1997.
$125.00 (hard cover), ISBN: 0-471-12262-9.

Adaptive Structures is a relatively new text that covers the subject of implementing active control in structures. The authors have worked rather extensively in this area, and the text represents an overview of much of the work that they have been involved in over the past decade. It is generally fairly well written, and would be a useful text to those working in the area. There are nine chapters in the text, with five of those being largely review types of chapters, and four of those covering much of the work that the authors have been involved in recently.

Chapter one provides an overview of adaptive structures, and is in many ways important to understand the authors' perspective in approaching the subject. They define an adaptive structure somewhat differently than perhaps what I would, but that is their prerogative. For purposes of this text, ''an adaptive structure is defined to be a structure configured with distributed actuators and sensors and directed by a controller capable of modifying the dynamic response of the structure in the presence of time-varying environmental and operational conditions.'' With this definition, one can have an adaptive structure that makes use of a fixed-gain controller, which the authors refer to as ''constrained adaptive'' or an adaptive structure that makes use of an adaptive control system, which they refer to as ''purely adaptive.'' While both types of adaptive structures are considered in this text, the emphasis is certainly on constrained adaptive structures, as most of the systems considered make use of classical or optimal control techniques, which are fixed-gain control systems.

Chapter two gives a brief review of structural dynamics, covering the process of obtaining the governing equations of motion for a system based on Lagrange's equations. Solutions based on modal expansions and on propagating wave representations are covered, as are the effects of damping. This chapter is written much in the spirit of Meirovitch's texts (for example, his text *Dynamics and Control of Structures*). However, there is not as much detail in this text as there is in Meirovitch's text, and as a result, it may be a little difficult reading for one who has not been exposed to these concepts previously.

The third chapter discusses linear systems and signals, and is again intended as a review chapter. Linear time invariant systems are discussed, which are the focus of the material covered in this text, as are various norms that can be considered in the control problem. Finally, two well-known optimal controllers are discussed: the linear quadratic regulator (LQR) and the linear quadratic Gaussian (LQG) controllers. Chapter 4 provides a brief review of signal processing and digital filters. Topics include correlation and spectral density functions, windowed functions, and finite impulse response (FIR) and infinite impulse response (IIR) filters. In addition, there is a brief coverage of adaptive filter theory, including the Wiener filter, Newton's algorithm, and the LMS algorithm. Chapters two through four are largely review chapters, but written with the assumption that the reader is rather familiar with the concepts. As a result, the reader who is not familiar with the concepts will likely struggle in following everything.

Chapter five covers transduction devices and how to incorporate those devices into the model of the physical system. The emphasis is on piezo-

electric devices, and particularly on the concept of ''sensoriactuators,'' which is an area in which at least one of the authors has spent considerable effort. Sensoriactuators are capable of operating both as actuators and sensors simultaneously and allow one to have truly colocated sensors and actuators, which can be a significant advantage when using feedback based controllers.

Chapter six deals with integrating both spatial and temporal signal processing. This deals with developing sensors (either distributed sensors or arrays of discrete sensors) to obtain desired information. The three objectives covered here include extracting modal amplitudes from the structural vibration, obtaining the amplitudes of propagating waves (in both nondispersive and dispersive media), and obtaining a measure of the acoustic radiation that results from the structural vibration.

Chapter seven covers classical control methods, although again it assumes that the reader has previous exposure to the subject. Techniques such as root-locus and Bode plots are used extensively in the development, but are not developed here. The eighth chapter continues on to discuss techniques for integrating the control system into the physical system. Such topics as global versus local control and narrow-band versus broadband control are covered as they relate to choosing the control system design. Feedback control systems are covered, as are both fixed feedforward and adaptive feedforward systems. Finally, the possibility of developing hybrid systems using both feedback and adaptive feedforward control is covered. The bias of the authors comes through a little in these chapters. For example, they indicate that feedback control is required for global control of a system, whereas I would suggest that it has been shown that global control can be achieved with feedforward systems as well. The authors also indicate that a weakness of the filtered-x algorithm is that the control-to-error transfer function model is required, which is normally done off-line. While this transfer function is required, it is rather common today to obtain that transfer function on-line using one of a couple of methods available for identifying that transfer function. In spite of these biases, the authors do cover some interesting work that they have been involved in for controlling structures.

The final chapter is essentially a culmination chapter, in which the authors address about six different control problems and propose a systematic way of developing an appropriate control system for a given problem. These steps include problem overview; description of the plant; transducer selection, design, and placement; development of the augmented plant model; control strategy; and simulation of experimental results. The problems covered include active control of vibration, active structural acoustic control (ASAC) with a frequency-shaped cost function, ASAC using a distributed piezoelectric transducer, similar problems with the model and control system based on experimentally obtained data, and active control of power flow in beams. These represent problems that the authors have been involved with over the past number of years, and provide a good overview of some of their work.

The text also comes with a MATLAB disk, intended to help the reader to reproduce results shown in the text and to further investigate the concepts by varying parameters within the simulations. It should be noted, however, that to run the files, the reader will also need to have a number of MATLAB Toolboxes, including the robust control toolbox, the mu analysis and synthesis toolbox, the control system toolbox, and the signal processing toolbox. I personally would have liked to see a little less reliance on the toolboxes, as many readers getting started in this area, who may really benefit from being able to play with the programs, will likely not have all the toolboxes available to them in order to run the software.

The authors have been involved extensively in the subject of this text, and there is considerable information contained in the text. It does have a rather different slant than a number of other texts that have come out over the past few years, and as such, provides a useful alternative approach to the subject of adaptive structures. It is, however, better suited for those who have some previous background in the area, as many of the background details are rather sketchy in the presentation.

SCOTT D. SOMMERFELDT
Department of Physics
Brigham Young University
Provo, Utah 84602-4673

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.

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
DAVID PREVES, *Songbird Hearing, Inc., 5 Cedar Brook Drive, Cranbury, New Jersey 08512*
ERIC E. UNGER, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*

6,079,214

43.10.Pr STANDING WAVE PUMP

Richard Patten Bishop, assignor to Face International Corporation
27 June 2000 (Class 62/6); filed 6 August 1998

Two transducers, located at opposite ends of a fluid-filled cylindrical cavity, are made to oscillate in opposite phase, so as to generate a standing wave in the cavity. The pressure difference between the pressures at nodes and antinodes is used to pump the fluid through inlets and outlets located at the nodes and antinodes. Check-valves at the output ports prevent the fluid medium from flowing from the higher-pressure plenum back into the cavity.—EEU

6,079,262

43.10.Pr COIN IDENTIFICATION PROCEDURE

Francisco Ibanez Palomeque and Jose Luis Pina Insausti, assignors to Azkoyen Industrial, S.A.
27 June 2000 (Class 73/163); filed in Spain 28 September 1995

This patent pertains to a micro-processor-based approach for identifying valid coins inserted into automatic vending devices. The coin is made to fall edge-on onto a hard surface and the resulting sound is sensed by a microphone. The low-frequency components of the sound signal, which depend on how the coin impacts, are removed by a high-pass filter. The remaining signal, which is representative of the dynamics of the coin itself, is filtered into a number of bands. The levels observed in these bands are noted, and their various ratios are computed and compared to acceptable values stored in a memory.—EEU

6,087,652

43.35.Zc CONTACTLESS ACOUSTIC SENSING SYSTEM WITH DETECTOR ARRAY SCANNING AND SELF-CALIBRATION

Thomas R. O'Meara and David M. Pepper, assignors to Hughes Electronics Corporation
11 July 2000 (Class 250/214.1); filed 1 May 1997

This patent pertains to laser-based contactless ultrasonic inspection systems. An ultrasonic beam propagating through the workpiece to be inspected causes the piece's readout surface to vibrate in a configuration that is distorted if an internal crack or the like is present. The surface configuration is detected by an optical probe beam that, combined with a reference beam, is directed onto the vibrating surface and made to scan in a pattern corresponding to a detector array. Electro-optic or acousto-optic techniques are used to define a reference phase shift for each pixel of the array. Calibration is obtained by imposing a known modulation on one or more of the probe and reference beams.—EEU

6,079,274

43.38.Fx VIBRATION WAVE DETECTING METHOD AND VIBRATION WAVE DETECTOR

Shigeru Ando and Muneo Harada, assignors to Sumitomo Metal Industries Limited
27 June 2000 (Class 73/649); filed in Japan 22 May 1998

An array of reeds, each with a different length and thus with a different resonance frequency, is used to sense a vibration or sound. A piezoresistor is formed in each reed, and these resistors are connected in parallel so that a signal corresponding to the sum of the vibrations of the reeds can be measured by means of a suitable circuit. By appropriate shaping of each of the piezoresistors one can obtain desired weightings of the contributions from the individual resonators.—EEU

6,064,744

43.38.Ja OMNI-DIRECTIONAL LOUDSPEAKER

Heinz-Juergen Augustin, Storkau, Germany
16 May 2000 (Class 381/160); filed 20 April 1998

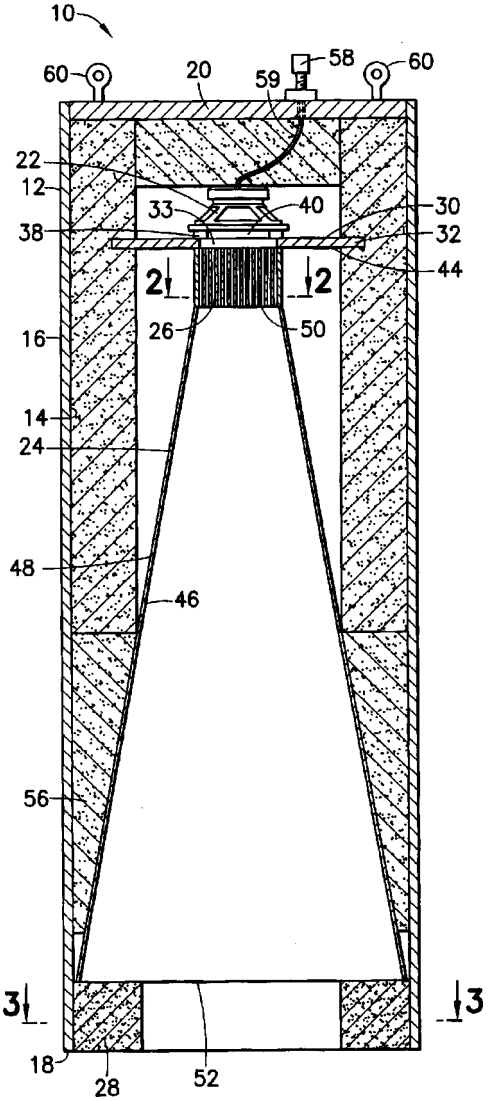
This is yet another nondirectional speaker system in which an upward-firing woofer and a downward-firing tweeter are separated by a large diffusing element. In this case, everything is held together by a central, hollow tube which can house wiring and crossover network components.—GLA

6,055,320

43.38.Ja DIRECTIONAL HORN SPEAKER SYSTEM

David Wiener and Steven Burgess, assignors to Soundtube Entertainment
25 April 2000 (Class 381/343); filed 26 February 1998

At first glance this appears to be little more than a loudspeaker driving a conical horn. However, loudspeaker **22** is spaced away from baffle **30**, providing some combination of front and rear sound pressure to drive throat **33**. Moreover, the sound must pass through in-line phase plug **26**, which



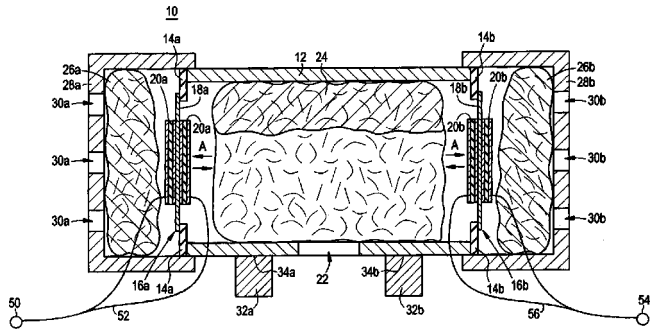
Claim 1.d describes as ''first wavefront manipulation means for altering the wavefront.'' Finally, a ring of absorptive material **28** is located at the horn mouth. All of which is supposed to produce a highly directional coverage pattern over a broad frequency range.—GLA

6,064,746

43.38.Ja PIEZOELECTRIC SPEAKER

Takeshi Nakamura and Yoshiaki Heinouchi, assignors to Murata Manufacturing Company, Limited
16 May 2000 (Class 381/351); filed 2 June 1997

This appears to be a variant of United States Patent 4,549,631, assigned to Bose Corporation (not reviewed). In this case ''sounding members'' **16a** and **16b** are piezoelectric transducers coupled to a central Helm-



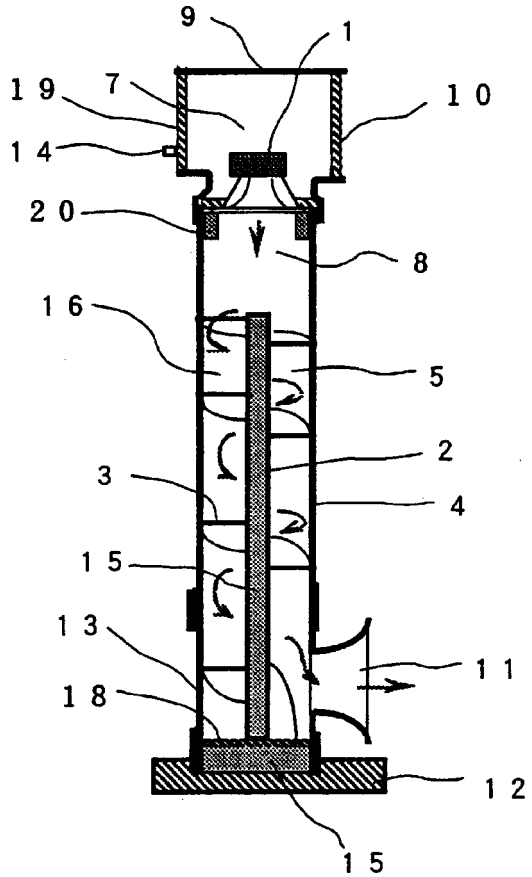
holtz resonator **12** and two outer resonators **28a**, **28b**. The two transducers can be connected to left and right stereo signals.—GLA

6,078,676

43.38.Ja SPEAKER SYSTEM WITH A THREE-DIMENSIONAL SPIRAL SOUND PASSAGE

Masaaki Takenaka, Chiba-ken, Japan
20 June 2000 (Class 381/338); filed in Japan 13 February 1998

A loudspeaker is mounted to one end of a coiled tube having the form of an Archimedes screw. Back chamber **7** and coupling chamber **8** may include damping material. The cross-sectional area of the spiral tube gradually expands from inlet to outlet, and sound emerges through a short flared



mouth **11**. The patent text concentrates on performance rather than theory, but it is possible that the path length difference between central and outer areas tames pipe resonances to some degree.—GLA

6,028,946

43.38.Kb MICROPHONE WITH ASSOCIATED AMPLIFIER

Helmut Jahne, assignor to Stage Tec Entwicklungsgesellschaft für professionelle Audiotechnik mbH
22 February 2000 (Class 381/122); filed in Germany 6 February 1996

In this best of all digital worlds it would seem logical to perform analog-to-digital conversion as early in the chain as possible, making it part of the preamplifier circuitry inside a microphone. The patent text explains that this usually results in a reduction of dynamic range and a host of other problems ranging from excessive digital jitter to special, expensive cables. The invention attempts to avoid these ills by transmitting outputs from two ADCs through standard shielded cable and then converting the two channels to a single-channel digital signal.—GLA

6,061,455

43.38.Lc AUDIO SYSTEM

Darby Edward Hadley and David John Stuart, assignors to Ford Motor Company
9 May 2000 (Class 381/57); filed in the United Kingdom 21 June 1996

An AGC circuit attenuates audio signal level when power amplifier distortion exceeds a predetermined threshold. The threshold is controlled by the setting of the manual volume control. Thus, greater amplifier distortion is permitted at higher volume settings.—GLA

6,028,980

43.38.Md IMAGE AND SOUND RECORDING APPARATUS USING A COMMON RECORDING HEAD AND AUDIO SIGNAL MEMORY

Shigeo Yamagata, assignor to Canon Kabushiki Kaisha
22 February 2000 (Class 386/96); filed in Japan 13 April 1987

Slow-rate video (say, two frames per second) and synchronized audio are recorded efficiently on a conventional hard disk. Those familiar with computer terminology and flow charts will find the patent interesting and easy to follow.—GLA

6,069,959

43.38.Si ACTIVE HEADSET

Owen Jones, assignor to Noise Cancellation Technologies, Incorporated
30 May 2000 (Class 381/71.6); filed 30 April 1997

This noise-cancelling headset allows the user to adjust the amount of noise reduction without changing the effective bandwidth of the system. The invention is said to cancel noise in both high and low frequency ranges yet minimize the subjective pressure within the ears felt by the user.—GLA

6,072,878

43.38.Vk MULTI-CHANNEL SURROUND SOUND MASTERING AND REPRODUCTION TECHNIQUES THAT PRESERVE SPATIAL HARMONICS

James A. Moorer, assignor to Sonic Solutions
6 June 2000 (Class 381/18); filed 24 September 1997

The patent presents the notion of spatial harmonics and explains how they are preserved by the inventor's encoding/decoding scheme. One goal is to generate precise phantom images from almost any arrangement of loudspeakers around the listening area. The concept is interesting, but a skeptical listener might want to hear it before believing it.—GLA

6,075,868

43.38.Vk APPARATUS FOR THE CREATION OF A DESIRABLE ACOUSTICAL VIRTUAL REALITY

Barry S. Goldfarb *et al.*, assignors to BSG Laboratories, Incorporated
13 June 2000 (Class 381/301); filed 25 April 1995

What we have here is ''A portable collapsible seat with an advanced five-driver integral audio system...'' intended to produce a psychologically gripping effect. And well it may.—GLA

6,090,147

43.40.At COMPUTER PROGRAM MEDIA, METHOD AND SYSTEM FOR VIBRATION AND ACOUSTIC ANALYSIS OF COMPLEX STRUCTURAL-ACOUSTIC SYSTEMS

Paul M. Bremner and Robin S. Langley, assignors to Vibro-Acoustics Sciences, Incorporated
18 July 2000 (Class 703/1); filed 5 December 1997

This computer program seeks to overcome the limitations of finite-element analysis (FEA) and statistical energy analysis (SEA) by combining the two. Its approach consists of partitioning a problem into subproblems of large-scale, long wavelength motions and of small-scale, short wavelength motions. The larger-scale, lower-frequency global motions are analyzed by means of FEA or similar methods, whereas the smaller-scale, higher-frequency local motions are evaluated by means of SEA.—EEU

6,087,936

43.40.Yq VIBRATION SENSOR

Randall Woods, Prescott, Arizona
11 July 2000 (Class 340/566); filed 29 December 1998

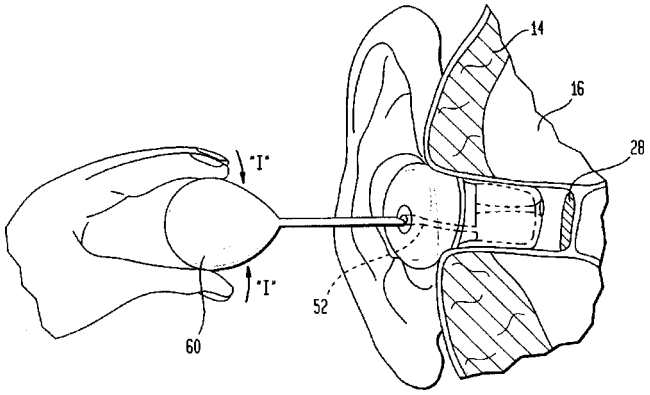
A typical embodiment of the sensor described in this patent consists of a cylindrical housing whose floor is formed by a conical cup. A number of electrical conductors are positioned above this floor, extending essentially radially. A conductive rod extends along the axis of the cylinder, but does not make electrical contact with the cup or the other conductors. In the absence of vibrations, a conductive ball rests on the floor of the housing and connects one or more of the radial conductors to the axial electrode. In the presence of a disturbance, the ball moves within the housing, establishing contact between the central electrode and different radial conductors. The changes in the characteristics of the electrical signal that results from contact interruptions and changes are analyzed to determine the magnitude, duration, and other characteristics of the vibration.—EEU

6,094,494

43.66.Ts HEARING AID DEVICE AND METHOD FOR PROVIDING AN IMPROVED FIT AND REDUCED FEEDBACK

Olaf Haroldson, Princeton, New Jersey
25 July 2000 (Class 381/328); filed 13 August 1998

A hearing aid housing is attached to a balloon that is inflated after insertion into an ear canal. The balloon contains a sound transmission duct that mates to the receiver output tube emanating from the attached hearing



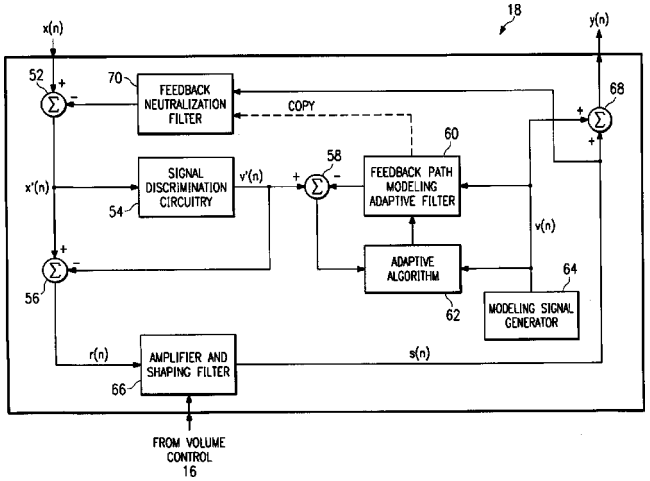
aid housing. As the balloon inflates, it conforms to the shape of the ear canal so as to provide a comfortable fit with reduced acoustical feedback without requiring a custom earmold or an ear impression.—DAP

6,097,823

43.66.Ts DIGITAL HEARING AID AND METHOD FOR FEEDBACK PATH MODELING

Sen M. Kuo, assignor to Texas Instruments, Incorporated
1 August 2000 (Class 381/312); filed 17 December 1997

Without interrupting amplification, a digital hearing aid is said to cancel acoustical feedback that results primarily from venting by continuously modeling the feedback path using an adaptive filter. Feedback neutralization is performed on the digitized input sound signal before signal processing and amplification occurs. The input signal consists of the desired sound input, an output sound feedback component, and a modeling output sound



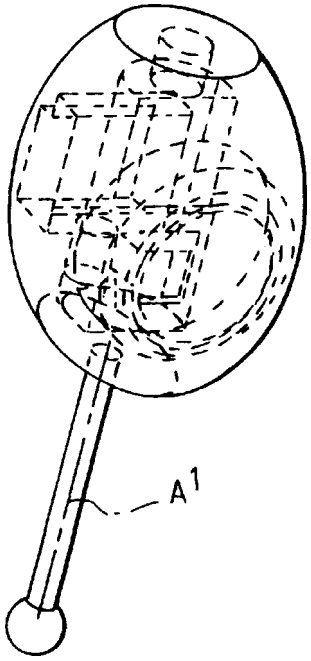
feedback component deliberately produced by a continuously running modeling signal generator. Signal discrimination circuitry generates a modified modeling signal by subtracting a predicted noise signal, containing no component of the feedback signal, from the feedback-neutralized input signal.—DAP

6,097,825

43.66.Ts HEARING AIDS WITH STANDARDIZED SPHEROIDAL HOUSINGS

Robert Yoest *et al.*, assignors to Beltone Electronics Corporation
1 August 2000 (Class 381/322); filed 19 September 1996

One-size-fits-all standard spheroid-shaped hearing aid housings made in mass production are said to fit deeply and comfortably into the ear canals of a large number of hearing aid wearers. Housings are formed either with a deformable or spongelike material covering the hearing aid components or with deformable material acting as a layer covering a molded plastic spher-



oid housing that houses the hearing aid components. Different sized ear canals can be accommodated by a limited number of different sized standard housings. When inserted into an ear canal, the deformable spherical housings conform in either a concave or convex shape to the wearer's outer ear canal.—DAP

6,097,826

43.66.Ts HEARING AID TO BE CARRIED COMPLETELY IN THE AUDITORY CANAL AND INDIVIDUALIZED BY A CAST BODY

Jurg Clavadetscher and Marcel Aeschilmann, assignors to Bernafon AG
1 August 2000 (Class 381/324); filed in Switzerland 24 July 1996

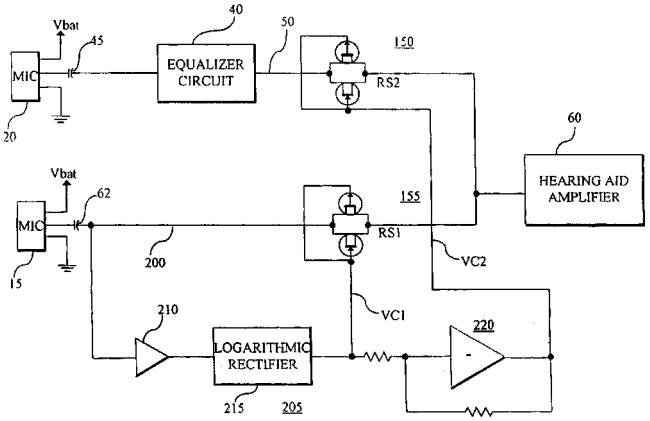
A completely-in-the-canal (CIC) hearing aid housing is customized by casting an outside body in an auditory canal or in a model of an auditory canal. During the casting process, a tubelike positioning tip centers in the ear canal the inner part of the hearing aid that contains the electro-acoustical components. After casting, the positioning tip is removed and the inner part of the hearing aid is rotated so as to not touch the auditory canal wall.—DAP

6,101,258

43.66.Ts HEARING AID HAVING PLURAL MICROPHONES AND A MICROPHONE SWITCHING SYSTEM

Mead Killion *et al.*, assignors to Etymotic Research, Incorporated
8 August 2000 (Class 381/321); filed 13 April 1993

This patent describes a small, switchable omnidirectional/directional microphone system for use in hearing aids. Omnidirectional mode is used for listening in quiet or to music. Noise reduction is achieved in directional mode for which the system operates as either a first-order gradient directional microphone or a second-order directional microphone that is constructed with two first-order directional microphones. Switching gradually from omnidirectional mode to directional mode can be made automatic by sensing when the ambient noise level rises above a certained predetermined value.—DAP



6,104,821

43.66.Ts ELECTRICAL HEARING AID DEVICE WITH HIGH FREQUENCY ELECTROMAGNETIC RADIATION PROTECTION

Kunibert Husung, assignor to Siemens Audiologische Technik GmbH
15 August 2000 (Class 381/312); filed in European Patent Office 2 October 1996

Hearing aid operation can be degraded during mobile radiotelephone use if radiated RF currents are allowed to be transduced and amplified. Improved protection against high-frequency electromagnetic signals is provided by placing a filter directly in the electrical terminals of the input and/or output of the integrated circuit hearing aid amplifier. Connection of RF-quality filter capacitors to reference potential, such as a metallic hearing aid case or to an extra ground layer in a printed circuit board, ideally forms a star-point-like structure.—DAP

6,104,822

43.66.Ts DIGITAL SIGNAL PROCESSING HEARING AID

John L. Melanson and Eric Lindemann, assignors to AudioLogic, Incorporated
15 August 2000 (Class 381/320); filed 10 October 1995

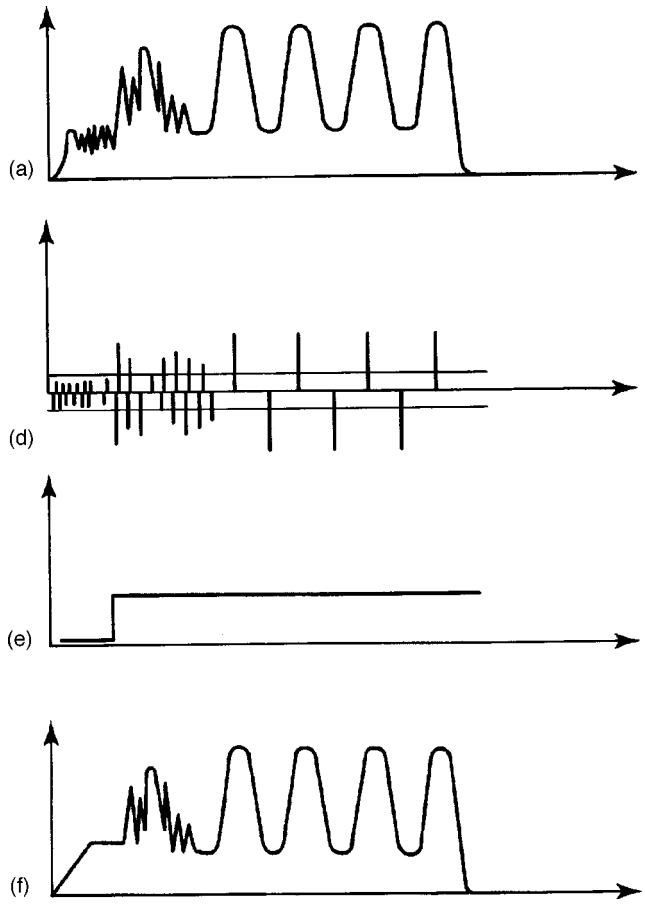
A multi-memory hearing aid allows the wearer to select the most appropriate digital signal processing strategy for a given listening situation. Changing strategies involves switching between programs executed by the digital processor, which, in effect, changes the number of bandpass filters and filter bandwidths into which the input signal is divided. Noise suppression algorithms require fine frequency resolution in the filter banks, which requires more battery current. For reduced power consumption, more simple dynamic range compression algorithms requiring less bands can be selected. The patent teaches how fine frequency resolution can be achieved in the filters at low frequencies without incurring a large delay through the system.—DAP

6,070,135

43.72.Ar METHOD AND APPARATUS FOR DISCRIMINATING NON-SOUNDS AND VOICELESS SOUNDS OF SPEECH SIGNALS FROM EACH OTHER

Chui Hong Kim and Jum Han Bae, assignors to Samsung Electronics Company, Limited
30 May 2000 (Class 704/215); filed in Republic of Korea 30 September 1995

This patent is concerned with playing back intelligible fragments of speech from an audio track being played at high speed, such as during a tape fast wind. Fast speech may be produced by shortening the vowels, but consonants must be kept at roughly the original length. Nonspeech noises should be eliminated. However, voiced/unvoiced detectors rarely make these distinctions. A method is described by which noise may be separated from voiceless speech on the basis of amplitude. This is shown by the acceptance thresholds as seen in the second part of the figure.—DLR



6,073,094

43.72.Gy VOICE COMPRESSION BY PHONEME RECOGNITION AND COMMUNICATION OF PHONEME INDEXES AND VOICE FEATURES

Lu Chang *et al.*, assignors to Motorola
6 June 2000 (Class 704/223); filed 2 June 1998

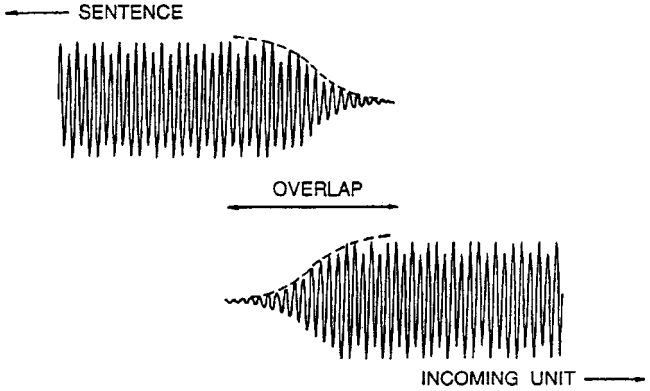
This phoneme vocoder is designed for use in a broadcast system, such as a digital radio, in which the speaker characteristics need not be preserved. A simplified form of recognition is used to extract a phonetic unit sequence from the input signal. This consists of a decorrelation of the log magnitude of spectral bands, which is then filtered and transformed back into the spectral domain. An autocorrelation then provides a voicing measure and a set of spectral features which are used to look up the corresponding phonetic unit in a reference library. The voicing information may be optionally transmitted to improve the reproduced speech quality.—DLR

6,067,519

43.72.Ja WAVEFORM SPEECH SYNTHESIS

Andrew Lowry, assignor to British Telecommunications public limited company
23 May 2000 (Class 704/264); filed in European Patent Office 12 April 1995

This patent describes a method for joining waveform segments from different recordings, providing a smooth transition from the left segment to the right segment. Each waveform segment may represent a different speech sound, such as a different phoneme, where all phonemes were pronounced by the same speaker. Each segment to be joined is categorized as voiced or



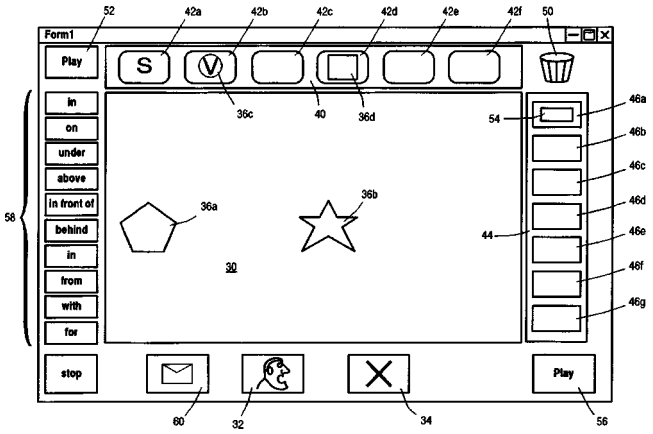
unvoiced and, if voiced, analyzed for glottal closure instants. Various category-specific rules govern the joining process so that, for example, voiced segments are aligned so as to match the glottal closure instants.—DLR

6,068,485

43.72.Ja SYSTEM FOR SYNTHESIZING SPOKEN MESSAGES

Marcia C. Linebarger and John F. Romania, assignors to Unisys Corporation
30 May 2000 (Class 434/116); filed 1 May 1998

This patent describes a sentence synthesizing system designed for aphasic patients who can speak individual words and short phrases, but who have difficulty assembling such phrases into complete sentences. Such patients typically can manipulate symbols which represent the short component items. These items can be separately recorded and each is associated with a geometric figure. The user can then manipulate the figures using a mouse, keyboard, or touch screen and cause the entire sentence to be synthesized.—DLR



6,068,487

43.72.Ja SPELLER FOR READING SYSTEM

Mark S. Dionne, assignor to Lernout & Hauspie Speech Products N.V.
30 May 2000 (Class 434/178); filed 20 October 1998

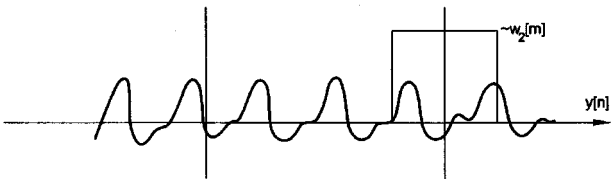
Patients with certain types of dyslexia or reading difficulties can be aided by a system for spelling out selected words or their synonyms. The device described here, intended for use with or as a part of a machine reading system, allows the user to select a word from a document, perhaps during the course of having that document read aloud by the machine, and have the selected word spelled aloud. During the spelling, letters of the word may be highlighted on the screen.—DLR

6,073,100

43.72.Ja METHOD AND APPARATUS FOR SYNTHESIZING SIGNALS USING TRANSFORM-DOMAIN MATCH-OUTPUT EXTENSION

Alan G. Goodridge, Jr. of Mountain View, California
6 June 2000 (Class 704/258); filed 31 March 1997

This high-quality audio synthesis system is said to be applicable for text-to-speech synthesis, audio editing, musical effects, and low-delay voice transformations, among others. Several input sections are defined using a variety of transform-domain signal representations. A candidate signal is



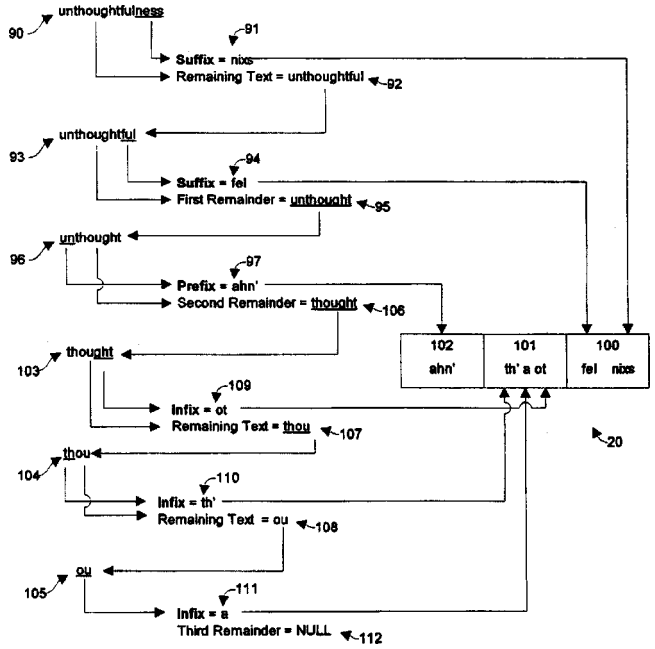
then transformed and matched against the input sections. Matching processes include frequency shifts, time shifts, and resampling factors. Selections are based on maximum cross-correlation, maximum sum of dot products, or minimum sum of squared differences.—DLR

6,076,060

43.72.Ja COMPUTER METHOD AND APPARATUS FOR TRANSLATING TEXT TO SOUND

Ginger Chun-Che Lin and Thomas Kopec, assignors to Compaq Computer Corporation
13 June 2000 (Class 704/260); filed 1 May 1998

This text-to-speech synthesizer is based on multiple sets of letter-to-phoneme rules. Separate rule sets are stored for portions of words; a set for prefixes, a set for suffixes, and a third set for any remaining central portion



of the word. It is argued that such rule sets can do a better job of extracting affixes from the head morpheme than traditional word-based rule sets. A few examples are given, but no complete rule sets are presented.—DLR

6,067,515

43.72.Ne SPLIT MATRIX QUANTIZATION WITH SPLIT VECTOR QUANTIZATION ERROR COMPENSATION AND SELECTIVE ENHANCED PROCESSING FOR ROBUST SPEECH RECOGNITION

Lin Cong and Safdar M. Asghar, assignors to Advanced Micro Devices, Incorporated
23 May 2000 (Class 704/243); filed 27 October 1997

A typical application of vector quantization in speech recognition is to construct a codebook from frames, or sets of acoustic analysis coefficients. If additional analysis vectors from preceeding or following frames are included, the process is known as matrix quantization. In the recognizer described here, the vectors or matrices are partitioned along the frequency axis into two or more bands and quantization is performed separately in each band. The band analysis results are then recombined using fuzzy logic methods to determine a final result.—DLR

6,067,517

43.72.Ne TRANSCRIPTION OF SPEECH DATA WITH SEGMENTS FROM ACOUSTICALLY DISSIMILAR ENVIRONMENTS

Lalit Rai Bahl *et al.*, assignors to International Business Machines Corporation
23 May 2000 (Class 704/256); filed 2 February 1996

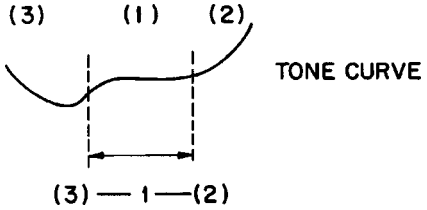
For use in an information kiosk where information may arrive from multiple news broadcasts, this speech recognizer includes multiple acoustic analysis modules, each trained with speech having a different environmental characteristic. During training, the clean speech training items are modified by the addition of noise or a certain type of distortion and then passed to the appropriate special case module. Several strategies are disclosed for selecting the most appropriate module during recognition.—DLR

6,067,520

43.72.Ne SYSTEM AND METHOD OF RECOGNIZING CONTINUOUS MANDARIN SPEECH UTILIZING CHINESE HIDDEN MARKOU MODELS

Lin-Shan Lee, assignor to Lee and Li
23 May 2000 (Class 704/270); filed 29 December 1995

This patent states that spoken Chinese is difficult to recognize because the relatively small syllable inventory leads to greater confusion among syllables. The problem is solved by a combination of methods, including speaker-dependent input, isolated syllable pronunciation, and a streamlined error correction process. The basic system uses standard hidden Markov model classification of consonant and vowel phones. A special pitch analy-



sis process includes Mandarin tone analysis. As a syllable is recognized, a corresponding character is displayed, which may be corrected by the user. Word-length sequences with tones are looked up in a syllable-oriented dictionary to find the most likely word. A further two-gram word sequence test helps disambiguate homophonic words.—DLR

6,070,136

43.72.Ne MATRIX QUANTIZATION WITH VECTOR QUANTIZATION ERROR COMPENSATION FOR ROBUST SPEECH RECOGNITION

Lin Cong and Safdar M. Asghar, assignors to Advanced Micro Devices, Incorporated
30 May 2000 (Class 704/222); filed 27 October 1997

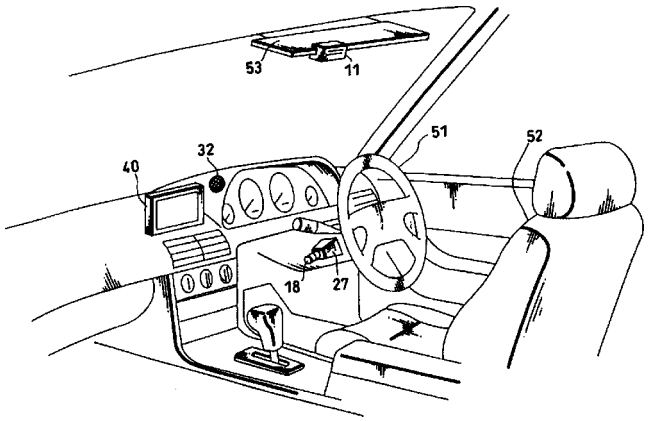
This patent is closely related to United States Patent 6,067,515, reviewed in this issue, in that it uses the same vector and matrix quantization methods. In the device presented here, a recognition controller uses any of several strategies to trade off between the matrix and vector quantization results to achieve the best overall recognition accuracy.—DLR

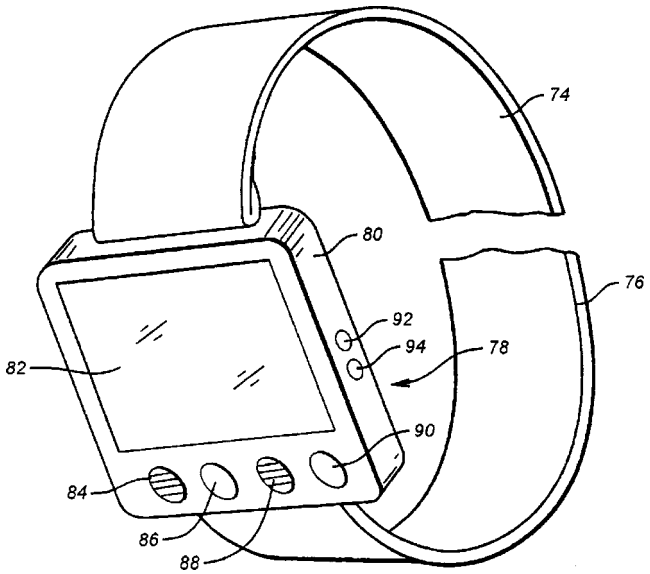
6,067,521

43.72.Ne INTERRUPT CORRECTION OF SPEECH RECOGNITION FOR A NAVIGATION DEVICE

Kazuo Ishii *et al.*, assignors to Sony Corporation
23 May 2000 (Class 704/275); filed in Japan 16 October 1995

This car navigation system uses speech recognition, along with GPS tracking and car speed sensing, to display a map of the local area. The primary task of the speech input seems to be in selection of the size of the region to be displayed. For the Japanese model, place names including prefectures, cities, wards, towns, and villages would be recognized. The microphone control uses push-to-talk and driver's seat localization to minimize false entry.—DLR





6,070,139

43.72.Ne BIFURCATED SPEAKER SPECIFIC AND NON-SPEAKER SPECIFIC SPEECH RECOGNITION METHOD AND APPARATUS

Yasunaga Miyazawa *et al.*, assignors to Seiko Epson Corporation
30 May 2000 (Class 704/275); filed in Japan 21 August 1995

This speaker-independent speech recognition system uses a two-processor method to improve on the recognition accuracy, hopefully approaching that of a speaker-dependent system. When a new speaker uses the system, recognition is performed both by a speaker-dependent processor, seen here as a remote terminal unit, and by a speaker-independent support processor, described as a base unit. As the new voice is learned by the terminal, it makes progressively less demands on the central unit by building up a local memory of the new speaker's characteristics.—DLR

6,070,140

43.72.Ne SPEECH RECOGNIZER

Bao Q. Tran of Houston, Texas
30 May 2000 (Class 704/275); filed 5 June 1995

Voice-controlled computer wakeup is desirable for small, portable devices. However, such a capability is typically difficult to design because some audio analysis is required to separate extraneous noises from valid wakeup keys and, yet, the audio input processing system usually has a large power drain, which should itself be advantageously powered down. This system applies two features toward solving the above problem. A directional multiple-microphone arrangement minimizes background noise and a hardware low-pass filter with 2-KHz cutoff minimizes nonspeech noises. A sufficient level of signal passing these constraints activates the full recognition system.—DLR

6,073,091

43.72.Ne APPARATUS AND METHOD FOR FORMING A FILTERED INFLECTED LANGUAGE MODEL FOR AUTOMATIC SPEECH RECOGNITION

Dimitri Kanevsky *et al.*, assignors to International Business Machines Corporation
6 June 2000 (Class 704/9); filed 6 August 1997

N-gram word probability estimates are more helpful in speech recognition as the value of *n* increases. However, the collection and management problems multiply exponentially with *n*. This is particularly a problem for inflected languages, such as Russian, where 400 000 word forms are needed to represent 99% of the spoken language. This patent presents a solution in the form of a layered probability model. Word forms are first ranked by frequency of occurrence and assigned to frequency classes. The entire corpus is then scanned to assign *n*-gram probabilities hierarchically, via the frequency class assignments. The result is a substantial reduction in memory and lookup time.—DLR

6,073,095

43.72.Ne FAST VOCABULARY INDEPENDENT METHOD AND APPARATUS FOR SPOTTING WORDS IN SPEECH

Satyanarayana Dharanipragada *et al.*, assignors to International Business Machines Corporation
6 June 2000 (Class 704/242); filed 15 October 1997

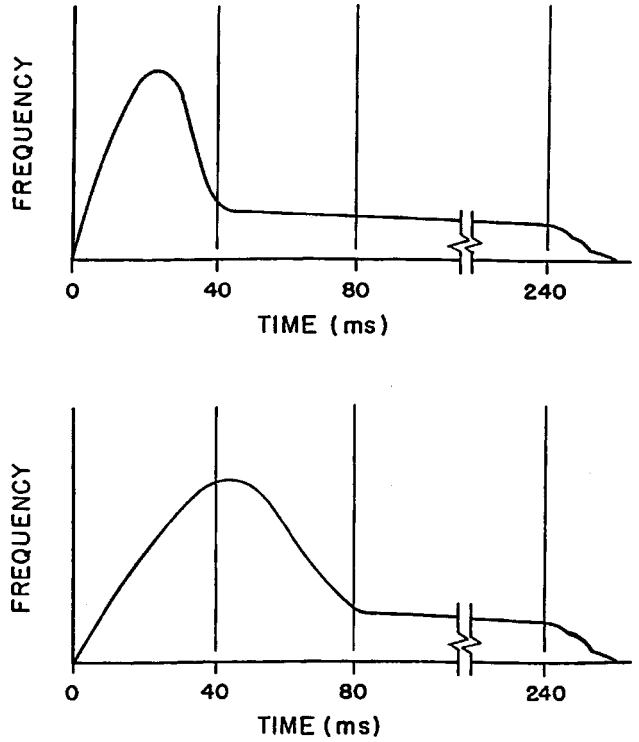
A vocabulary-independent word spotting system is described in which likely hits are recorded based on a precomputed table of phoneme *n*-gram statistics. The *n*-gram table also provides an efficient way to locate examples of the target word in the training speech data. When a possible hit is detected in the incoming speech, a detailed comparison of the speech waveform is done to further qualify whether the hit is a valid occurrence of the target word.—DLR

6,071,123

43.72.Ne METHOD AND DEVICE FOR ENHANCING THE RECOGNITION OF SPEECH AMONG SPEECH-IMPAIRED INDIVIDUALS

Paula Anne Tallal *et al.*, assignors to The Regents of the University of California; Rutgers, the State University of New Jersey
6 June 2000 (Class 434/116); filed 8 December 1994

A battery of psychological tests is described showing that speech perception in certain disadvantaged individuals can be enhanced by slowing down the rate of change of frequency transitions in the speech signal, as shown in the figures. The patent describes a device for producing these rate



changes in real time from a normal speech input signal. The same effect is said to apply to adult foreign language learning. The latter assertion should be met with a healthy skepticism.—DLR

6,073,096

43.72.Ne SPEAKER ADAPTATION SYSTEM AND METHOD BASED ON CLASS-SPECIFIC PRE-CLUSTERING TRAINING SPEAKERS

Yuqing Gao *et al.*, assignors to International Business Machines Corporation
6 June 2000 (Class 704/245); filed 4 February 1998

This speaker-adaptive speech recognition system initially builds clusters of speaker characteristes based on speaker-dependent training for each of the training speakers. The trained models, in the form of HMM or other speech models, are clustered using bottom-up methods, such as Gaussian likelihood, or top-down methods, such as Euclidean distance, or by searching for common acoustic characteristics. When a test speaker uses the system, the nearest training cluster is adapted using a transformation technique, such as maximum likelihood linear regression, to bring the cluster closer to the test speaker's acoustic space.—DLR

6,073,097

43.72.Ne SPEECH RECOGNITION SYSTEM WHICH SELECTS ONE OF A PLURALITY OF VOCABULARY MODELS

Joel M. Gould *et al.*, assignors to Dragon Systems, Incorporated
6 June 2000 (Class 704/251); filed 13 November 1992

The patent describes a number of functions and capabilities of a speech recognition system designed as a user interface for operating a personal computer. These functions include a prompting system to prompt the user for specific words or phrases and then decide whether to adapt the current model for a particular phrase or to prompt for additional inputs. The system has the ability to obtain information from other running programs and from the screen display in order to select speech models for particular sets of words or phrases and can create new vocabulary states associated with each such display or running program.—DLR

6,073,101

43.72.Ne TEXT INDEPENDENT SPEAKER RECOGNITION FOR TRANSPARENT COMMAND AMBIGUITY RESOLUTION AND CONTINUOUS ACCESS CONTROL

Stephane Herman Maes, assignor to International Business Machines Corporation
6 June 2000 (Class 704/275); filed 2 February 1996

This system provides both speaker-independent speech recognition and text-independent speaker recognition. It is intended for control applications, such as vehicle control, computer control, voice dialing, or customer data access. Acoustic feature vectors are extracted from overlapping speech frames and compared with vector parameters and variances stored in a codebook for enrolled speakers. A count is maintained of frames which are sufficiently similar to one of the enrolled speakers. This count is then used to identify an enrolled speaker or to detect a new unenrolled speaker.—DLR

6,073,102

43.72.Ne SPEECH RECOGNITION METHOD

Hans-Ulrich Block, assignor to Siemens Aktiengesellschaft
6 June 2000 (Class 704/275); filed in Germany 19 April 1996

This speech recognition system is intended for use in a control application. As far as this reviewer can tell from the confusing patent text, a system implementer can edit a list of possible key words, along with corresponding actions to be taken by the system upon recognition. There is some manner of provision for dealing with multiple action candidates, perhaps that multiple key words can be defined to execute a particular action.—DLR

6,076,053

43.72.Ne METHODS AND APPARATUS FOR DISCRIMINATIVE TRAINING AND ADAPTATION OF PRONUNCIATION NETWORKS

Biing-Hwang Juang and Filipp E. Korkmazskiy, assignors to Lucent Technologies, Incorporated
13 June 2000 (Class 704/236); filed 21 May 1998

This patent pertains to a method of storing alternative pronunciations of a word in the form of a network so as to be usable by a speech recognition system. The recognition system uses HMM subword or phonetic units, such that the *n*-best responses to a spoken word will be likely to encompass alternate pronunciations of that word. In other words, ''how the word was most likely spoken'' is reinterpreted as ''how the word might possibly be spoken.'' Detailed scoring procedures are presented for effectively using the pronunciation networks so constructed.—DLR

6,076,054

43.72.Ne METHODS AND APPARATUS FOR GENERATING AND USING OUT OF VOCABULARY WORD MODELS FOR SPEAKER DEPENDENT SPEECH RECOGNITION

George J. Vysotsky and Vijay R. Raman, assignors to Nynex Science & Technology, Incorporated
13 June 2000 (Class 704/240); filed 29 February 1996

The patent describes a recognizer system in which multiple speaker-dependent (SD) and speaker-independent (SI) recognizers are operated simultaneously using the same speech input signal. Each of the recognizers may include various out-of-vocabulary, or garbage, word models. In this case, words which are to be correctly recognized by a SI system are covered by a garbage model in a SD system. This reduces the risk of obtaining conflicting recognition results from the two systems.—DLR

6,076,056

43.72.Ne SPEECH RECOGNITION SYSTEM FOR RECOGNIZING CONTINUOUS AND ISOLATED SPEECH

Xuedong D. Huang *et al.*, assignors to Microsoft Corporation
13 June 2000 (Class 704/254); filed 19 September 1997

Isolated-word speech recognizers typically work well as long as the words are in fact spoken in an isolated manner. However, users would often prefer to speak naturally until the system begins to make errors. At that point, they will revert to an isolated-word speaking style. This recognition system is able to combine previously trained speech reference data from both isolated word and continuous recognition systems. New models are trained for each word, resulting in a system which can benefit from the advantages of each style of speaking for the recognition task.—DLR

6,076,057

43.72.Ne UNSUPERVISED HMM ADAPTATION BASED ON SPEECH-SILENCE DISCRIMINATION

Shrikanth Sambasivan Narayanan *et al.*, assignors to AT&T Corporation
13 June 2000 (Class 704/256); filed 21 May 1997

The background section of this patent argues that a speech/silence discriminator (SSD) can be used to help with the more general problem of speech recognition. The patent discloses a procedure for training a SSD without requiring previously marked training data. During normal speech recognizer operation, a SSD HMM is adapted using an unidentified discriminative training algorithm (the abstract suggests that a ''GPD'' might be used). The adapted SSD may then be used stand-alone or together with a speech recognizer.—DLR

6,076,058

43.72.Ne LINEAR TRAJECTORY MODELS INCORPORATING PREPROCESSING PARAMETERS FOR SPEECH RECOGNITION

Rathinavelu Chengalvarayan, assignor to Lucent Technologies, Incorporated
13 June 2000 (Class 704/256); filed 2 March 1998

The patent asserts that Mel-warped Fourier transform features, if transformed appropriately, offer speech recognition performance superior to that of the standard Mel cepstral coefficients. Fourier (DFT) coefficients are processed into Mel-warped, log filter bank coefficients using known methods. A linear combination of succeeding frames is then computed as a new, augmented feature vector. The weights are found by gradient descent optimization based on a minimum classification error criterion.—DLR

Excitation and propagation of non-axisymmetric guided waves in a hollow cylinder

Jian Li[a)]
Matec Instrument Companies, Inc., 56 Hudson Street, Northboro, Massachusetts 01532

Joseph L. Rose[b)]
212 Earth & Engineering Sciences Building, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 25 February 2000; accepted for publication 7 August 2000)

Excitation and propagation of non-axisymmetric guided waves in a hollow cylinder is studied by using the normal mode expansion method (NME). Different sources such as angle beam, tube end excitation with normal beam, and comb transducer possibilities are discussed based on the derivations of the NME method. Numerical calculations are focused on the case of angle beam partial loading. Based on the NME method, the amplitude coefficients for all of the harmonic modes are obtained. Due to the difference of phase velocities for different modes, the superimposed total field varies with propagating distances and hence makes particle displacement distribution patterns (angular profile) change with distance. This varying non-axisymmetric angular profile of guided waves represents a nonuniform energy distribution around the hollow cylinder and thus has an impact on the inspection ability of guided waves. The angular profiles of an angle beam source are predicted by theory and then verified by experiments. The predicted angular profiles also provide information for determining the transducer location to find defects in a certain position on the hollow cylinder. [DOI: 10.1121/1.1315290]

PACS numbers: 43.20.Mv [DEC]

I. INTRODUCTION

Pipe and tube testing has been studied for decades as a result of thousands of miles of tubing and piping in all sorts of power generation, petrochemical, and manufacturing plants. Recent studies have been made on the utilization of ultrasonic guided waves for structural inspection over long distances and also for structures under insulation and coatings.[1–10]

The first paper on a complete theoretical investigation of guided wave propagation in pipes was written by D. C. Gazis.[1] He obtained the general solution of harmonic waves propagating in an infinitely long hollow cylinder. Numerical calculations show that there are a doubly infinite number of modes for a circular cylinder, which includes an infinite number of torsional modes, an infinite number of longitudinal modes and a doubly infinite number of flexural modes.[2] Flexural modes are non-axisymmetric.[2,6] The generation of different modes depends on source loading conditions. Axisymmetric modes like longitudinal modes have been studied carefully in earlier years because of their uniform circumferential energy distribution. Different kinds of generation methods for axisymmetric modes have been widely used. One example is the comb transducer for long distance pipe inspection;[8] Cawley used a dry coupled normal beam transducer at tube ends for heat exchanger inspection source loading.[9] A transducer array made of a series of normal beam transducers was also used for axisymmetric guided wave generation.[10] Besides all of the methods mentioned, non-axisymmetric source loading is desired in certain circumstances such as when only a portion of the cylinder might be accessible. When non-axisymmetric guided waves are generated, the acoustic field is more complicated and the energy distribution of the wave propagation needs to be known in order to evaluate the guided wave inspection ability and to perform frequency and angle tuning.

The hollow cylinder source loading problem was been studied by John Ditri by using the NME method.[4] The amplitude factors of the different harmonic modes were obtained. However, the generated non-axisymmetric field distribution was not discussed, which is very important in terms of the influence on the inspection ability of guided waves. In this article, the normal mode expansion (NME) technique is used to study the acoustic field distribution generated by partial loading. Some special features such as resulting angular profiles will be discussed. Numerical results show that the angular energy profile of the guided wave varies with distance. The angular profile is carefully studied in both theory and experiment. Its impact on guided wave inspection of hollow cylinders is also discussed.

II. THE NORMAL MODE EXPANSION METHOD

The hollow cylinder studied is illustrated in Fig. 1, where guided waves propagate along the $+z$ and $-z$ directions. In order to detect defects in a hollow cylinder, it would be useful to understand the acoustic field generated by the source. One traditional method for solving this problem employs a classical integral transform method. Instead of using the integral transform method, we use here the normal mode expansion method. In this technique, analogous to the expan-

a)Electronic mail: jxl352@psu.edu
b)Electronic mail: jlresm@engr.psu.edu

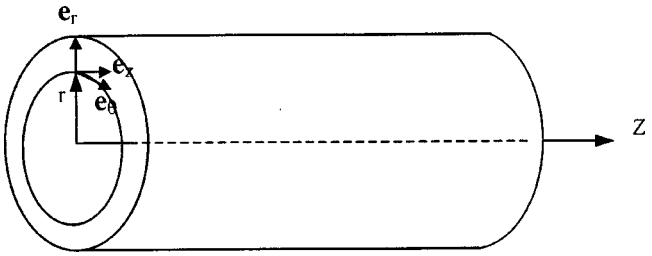


FIG. 1. Cylindrical coordinate system for the hollow cylinder.

sion of a function in terms of any orthogonal set of functions, the fields generated in the cylinder are expanded in the form of an infinite series of the normal modes of the cylinder.[4] The generated particle velocity can therefore be expanded as follows:

$$\mathbf{v}e^{i\omega t}=\sum_{M,n} A_n^M \mathbf{v}_n^M e^{i(\omega t)}, \tag{1}$$

where A_n^M, is the normal mode expansion amplitude of the nth mode of Mth circumferential order, contains $e^{i(-k_n^M z)}$, assuming waves are propagating in the $+z$ direction.

A. Dispersion curves

According to Gazis,[1,2] there are a doubly infinite number of propagating modes for the hollow cylinder. We denote the fields of the normal modes by two indices, n and M, so that the velocity field due to the ''nth'' mode of the ''Mth'' family can be written[1,2]

$$\mathbf{v}_n^M e^{i(\omega t-k_n^M z)}=\{R_{nr}^M(r)\Theta_r^M(M\theta)\mathbf{e}_r+R_{n\theta}^M(r)\Theta_\theta^M(M\theta)\mathbf{e}_\theta + R_{nz}^M(r)\Theta_z^M(M\theta)\mathbf{e}_z\}e^{i(\omega t-k_n^M z)}, \tag{2}$$

where ω and k are the angular frequency and wave number, respectively. $\mathbf{v}_n^M$ is independent of the cylinder axial direction, z, and the time, t. The particle displacement can be decomposed in three directions: $\mathbf{e}_r$, $\mathbf{e}_\theta$, $\mathbf{e}_z$, which are the unit vectors in a cylindrical coordinate system shown in Fig. 1. Functions $R(r)$ and $\Theta(M\theta)$ denote the radial and angular distributions of the component of the stress produced by the ''nth'' mode of the ''Mth'' family respectively. Θ_r^M, Θ_θ^M, and Θ_z^M are sinusoidal functions, $\cos(M\theta)$ and/or $\sin(M\theta)$. When $M=0$, the modes are axisymmetric and for $M>0$, we have a doubly infinite number of flexural modes. The dispersion curves for the axisymmetric longitudinal and non-axisymmetric flexural modes are shown in Fig. 2, which, as an example, is for a carbon steel pipe with 88-mm outer diameter (o.d.) and 5-mm wall thickness.

From Fig. 2 we can see that the dispersion curves for the longitudinal modes are very similar to those of Lamb waves. Beside the dispersion curves for each longitudinal mode $L(0,n)$, there are an infinite number of flexural modes $F(M,n)$ with phase velocities very close to that of $L(0,n)$. The phase velocity difference between the longitudinal mode and flexural modes will be shown. For convenience, we represent both longitudinal and flexural modes as $L(M,n)$. If $M=0$, it is a longitudinal mode; otherwise it is a flexural mode. For an ordinary source such as angle beam wedge, the

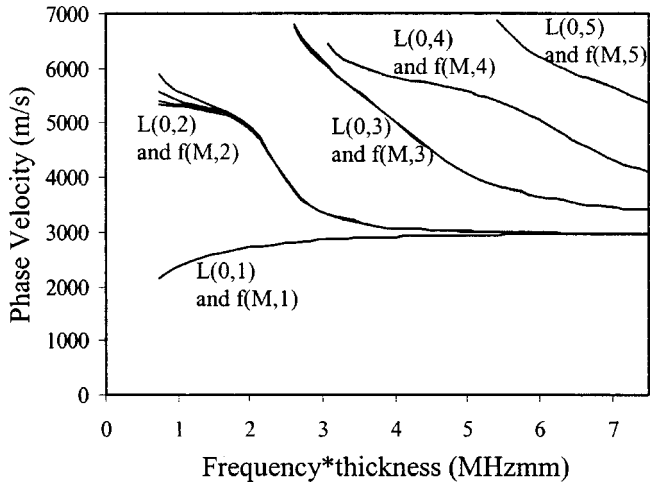


FIG. 2. Dispersion curves for the axisymmetric longitudinal modes and non-axisymmetric flexural modes in a carbon steel hollow cylinder with 88-mm diameter and 5-mm wall thickness.

generated modes are controlled by Snell's law and the total field is the superposition of the longitudinal mode and a group of flexural modes. The slight difference between phase velocities makes the superposition pattern change with the propagating distance, and, therefore, leads to the varying angular profile.

It should be pointed out that the dispersion curves for the guided waves in a hollow cylinder are determined by two geometrical factors that play different roles: cylinder diameter and wall thickness. The numerical calculations show that the phase velocities of the longitudinal modes are mostly determined by wall thickness, which is similar to the plate thickness for Lamb waves. The difference between the flexural modes and corresponding longitudinal modes is determined by the cylinder diameter. Numerical calculations also show that the larger the diameter, the closer the flexural mode phase velocity will be to the longitudinal mode phase velocity. This can also explain, in Fig. 2, why the phase velocities of the flexural modes are so close to those of the longitudinal modes because of a large cylinder diameter. For a smaller diameter tube,[7] the phase velocities of the flexural and longitudinal modes are well separated, and the modes are more easily defined.

B. Orthogonality of normal modes and the normal mode expansion method

The amplitude factor A_n^M can be calculated by using the normal mode expansion method. In order to use this method, first, the orthogonality relation between the waveguide modes should be established. Such an orthogonality relation has already been developed for the layered waveguide case. The hollow cylinder case can start with the complex reciprocity relation, which is described as follows:

$$\nabla\cdot(V_2^*\cdot T_1+V_1\cdot T_2^*)=0, \tag{3}$$

where V_1, T_1 and V_2, T_2 represents the particle velocity and stress fields of the two different solutions to the linear elastic wave propagation equation. Based on this relation, the or-

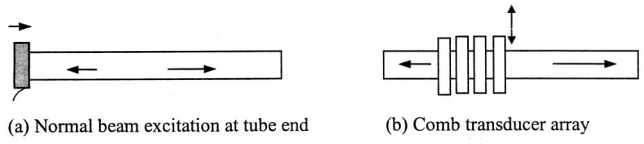


FIG. 3. Axisymmetric source loading considerations.

thogonal relation between wave modes can be derived in terms of an area integral, P,[4]

$$P_{nm}^{MN} = -\frac{1}{4}\int\int_D (V_n^{M*}\cdot T_m^N + V_m^N \cdot T_n^{M*})\cdot \mathbf{e}_z d\sigma, \tag{4}$$

where D is the cross section of the waveguide. For two modes V_n^M, V_m^N,

$$P_{nm}^{MN} = 0 \quad \text{unless} \quad n=m \quad \text{and} \quad M=N. \tag{5}$$

With the orthogonality relationship in hand, we can proceed with the normal mode expansion technique. For the complex reciprocity relation in (3), we choose the total field expressed as formula (1) and the field of mode n of circumferential order M as expressed in formula (2). The complex reciprocity in formula (3) can be reformulated as

$$\nabla_{r\theta}\cdot(V_n^{M*}\cdot T + V\cdot T_n^{M*})e^{i\beta_n^{M*}z} + \frac{\partial}{\partial z}\left(\sum_{N=0}^{\infty}\sum_{\mu} A_\mu^N(z)(V_n^{M*}\cdot T_\mu^N + V_\mu^N\cdot T_n^{M*})\cdot e_z\cdot e^{i\beta_n^{M*}z}\right) = 0. \tag{6}$$

The integral of formula (6) forms the basis for solving various source loading problems. In the following section, normal beam, comb transducer and angle beam partial loading are discussed based on formula (6) and the corresponding orthogonality relations (5).

C. Axisymmetric source loading

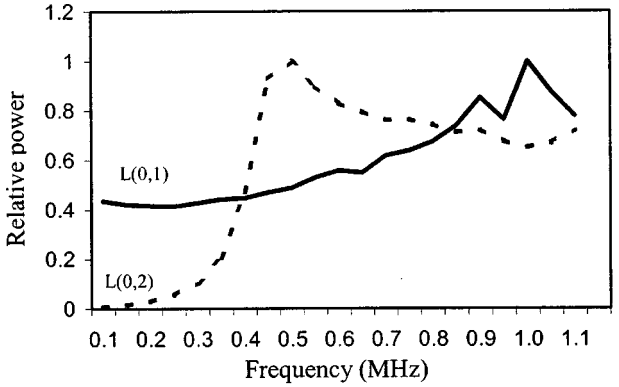
Both normal beam and comb transducers can be treated as axisymmetric source loadings. See Fig. 3. Only axisymmetric longitudinal modes are excited. If the source is placed on the tube end, as in a normal beam method, the inner and outer tube surfaces are traction-free, and the integral of the first term in formula (6) will vanish. Hence the amplitude factor $A_n(z)$ can be derived based on formula (6):

$$A_n(z) = S\cdot\frac{e^{-i\beta_n z}}{4P_{nn}}\cdot(\overline{V_{nz}^*}\cdot p + V\cdot\overline{T_{nzz}^*}), \tag{7}$$

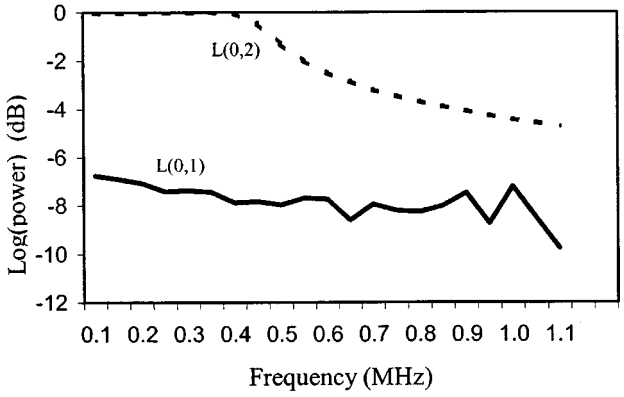
where S is the cross-sectional area of the tube end at $z=0$; $\overline{V_{nz}}$ and $\overline{T_{nzz}}$ are the average normal particle velocity and normal stress at the tube end for the longitudinal mode n. V and p are the particle velocity and pressure of the source.

For another case, if the surface of the cylinder is not traction-free, which is the case of partial or comb transducer loading (source is placed on the outer boundary), formula (6) can be reformulated as

$$A_n^M(z) = \frac{e^{-i\beta_n^M z}}{4P_{nn}^{MM}}\int_{-L}^{L} e^{i\beta_n^M \eta}\left(\oint_{\partial D} V_n^{M*}\cdot(T\cdot\mathbf{e}_r)\,ds\right)d\eta, \tag{8}$$



(a) One comb element source loading.



(b) Normal beam source loading

FIG. 4. Relative power of $L(0,1)$ and $L(0,2)$ modes sent by the single comb element and normal beam source loadings on a carbon steel pipe with 88-mm o.d. and 5-mm wall thickness.

where $-L$ and L are the starting and ending positions of the source along the cylinder, and ∂D is the outer boundary. For a comb transducer, the outer boundary loading is axisymmetric and hence, only axisymmetric modes can be generated. In this case, formula (8) can be put in a simpler form:

$$A_n(z) = S^*\frac{V_{nr}^*\cdot p}{4P_{nn}}e^{-i\beta_n z}, \tag{9}$$

where S is the transducer area touching the cylinder and p is the normal pressure of the transducer. V_{nr} and P_{nn} are the partial velocity and power density of that mode, respectively. Note that formula (9) employs an assumption that the comb element widths are much shorter than the wavelength. For a transducer array with N elements and element distance D, the amplitude factor will be

$$A_n(z) = S^*\frac{V_{nr}^*\cdot p}{4P_{nn}}e^{-i\beta_n z}* \sin\frac{(2\pi^* N^* D/\lambda)}{(2\pi^* D/\lambda)}, \tag{10}$$

where D is the element distance. $A_n(z)$ will attain a maximum value when D is equal to the wavelength. Therefore, mode control can be implemented by changing D.

For the normal beam source or for just one comb element, all possible longitudinal modes will be generated, although with very different sensitivities. Figure 4 shows the power carried by $L(0,1)$ and $L(0,2)$, which are excited by

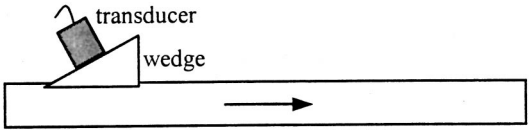


FIG. 5. Angle beam source loading for a hollow cylinder.

the normal beam and a comb element, respectively. Suppose that the transducers for both source loadings had a broadband response and that the comb element widths were much shorter than the wavelength. The frequency range for the calculation for Fig. 4 is 50 KHz to 1.1 MHz. Figure 4 shows that at low frequencies, efficiency to generate the $L(0,1)$ mode is greater for the comb transducer, but the $L(0,2)$ mode produced by the normal beam method is better. This can be explained on the basis of the wave structures of both modes. For example, for the $L(0,1)$ mode at low frequency, the particle displacement in the radius direction dominates and carries most of the power transferred by the wave. Therefore, it is much easier to couple with the $L(0,1)$ mode for the comb transducer at low frequencies.

D. Non-axisymmetric partial loading

A partial loading source is not axisymmetric and covers only the pipe surface over some circumferential angle. Partial loading generates both longitudinal and flexural modes in the hollow cylinder. Due to the slight difference of the phase velocities of the flexural modes, the particle displacement distribution changes with propagating distance.

For an angle beam wedge loading, shown in Fig. 5, the traction on the outer boundary can be formulated as in Fig. 4.

$$T\cdot n_1=\begin{cases} -p_1(\theta)p_2(z)e_r, & |z|\leqslant L, \quad r=b, \quad \theta<|\theta_0|, \\ 0, & |z|>L, \quad r=b, \quad \theta>|\theta_0|. \end{cases} \tag{11}$$

Putting formula (11) into (8), we obtain a simpler form for the amplitude factor:

$$A^M_{+n}(z)=-\frac{R^{M*}_{nr}(b)e^{-i\beta^M_n z}}{4P^{MM}_{nn}}\langle\Theta^M_r,p_1(\theta)\rangle\cdot\langle e^{i\beta^M_n z},p_2(z)\rangle. \tag{12}$$

From formula (12), we see several ways of increasing the amplitude factor of a particular mode, i.e., hence implementing mode control. In order to make the amplitude factor a maximum, we have to set the function $p_2(z)$ equal to $e^{-i\beta^M_n z}$, which leads to an expression of Snell's law. We also have to make the inner product between $p_1(\theta)$ and Θ^M_r a maximum in order to maximize the amplitude of the mode $L(M,n)$. For most angle beam wedges, it is simpler to control the first term by Snell's law. For the second term, we cannot change it and must therefore approximate $p_1(\theta)$ as a constant value in certain areas.

III. NUMERICAL CALCULATIONS

For numerical calculations, consider a Plexiglas wedge of size 30×50 mm^2. The hollow cylinder is an 88-mm-o.d. carbon steel pipe with 5-mm wall thickness. The wedge covers 45 degrees in the circumferential direction of the pipe. If

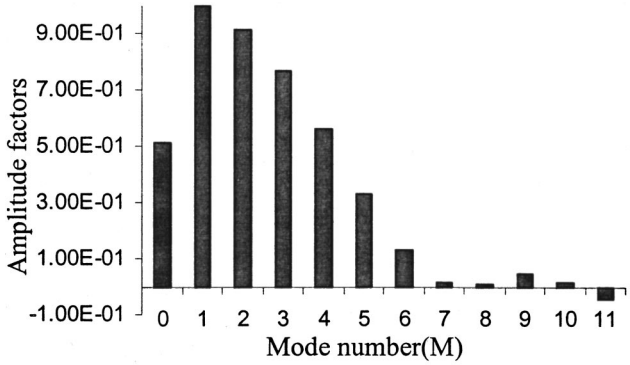


FIG. 6. Amplitude factors for $L(M,2)$ modes generated by a 70-degree angle beam at 290 KHz.

we make the wedge angle 70 degrees and frequency 290 KHz for example, most of the generated modes will be $L(M,2)$. Using the normal mode expansion method just discussed, the amplitude factors of each circumferential mode $L(M,2)$ are computed, and then plotted in Fig. 6.

The total generated field can be calculated based on the amplitude factor, phase velocity and wave structure of each mode. By superimposing all of the modes together, we obtain the particle displacement angular profiles of the total field at different propagation distances, as shown in Fig. 7. In Fig. 7, angular profiles are illustrated for $L(M,2)$ modes at 290 and 375 KHz and $L(M,1)$ modes at 375 KHz, respectively. From Fig. 7, we can see that the variables that change the angular profiles include angle beam size, frequency, mode and propagating distance. At a distance close to the transducer, in this case, at a 0-degree circumferential position, the angular profiles are focused in the transducer location area. When the wave propagates, the angular profiles change and spread to various patterns. Even at the same distance, the angular profiles are different for the same mode at different frequencies, or different modes at the same frequency. Interestingly, Fig. 7 also show that when the propagating distance is 3.4 m, most of the energy is focused on the opposite circumferential position to the source loading for $L(M,2)$ modes at the frequency of 290 KHz.

It should be pointed out that $L(M,1)$ modes are ignored for the angular profile calculations in Fig. 7. Although all existing modes at the excitation frequency are generated as we discussed, the phase velocities of the $L(M,1)$ modes are very different from those of the $L(M,2)$ modes. Therefore, when the incident angles are adjusted to generate the $L(M,2)$ modes, the third terms of formula (12) for the $L(M,1)$ modes are far less than those of the $L(M,1)$ modes, and thus the small amplitudes of the $L(M,1)$ modes can be ignored.

The major conclusion drawn from one illustrative example shown in Fig. 7 is that guided wave modes generated by partial loading cannot find all of the defects in a pipe with only one wedge angle and frequency because of the nonuniform angular profiles. Although this problem can be overcome by rotating the transducer around the pipe, there is an alternative by sweeping the frequency or transferring to another mode to obtain different energy distribution/angular profiles. The later method is valuable when only a part of the pipe is accessible. We can see from Fig. 7 that the three

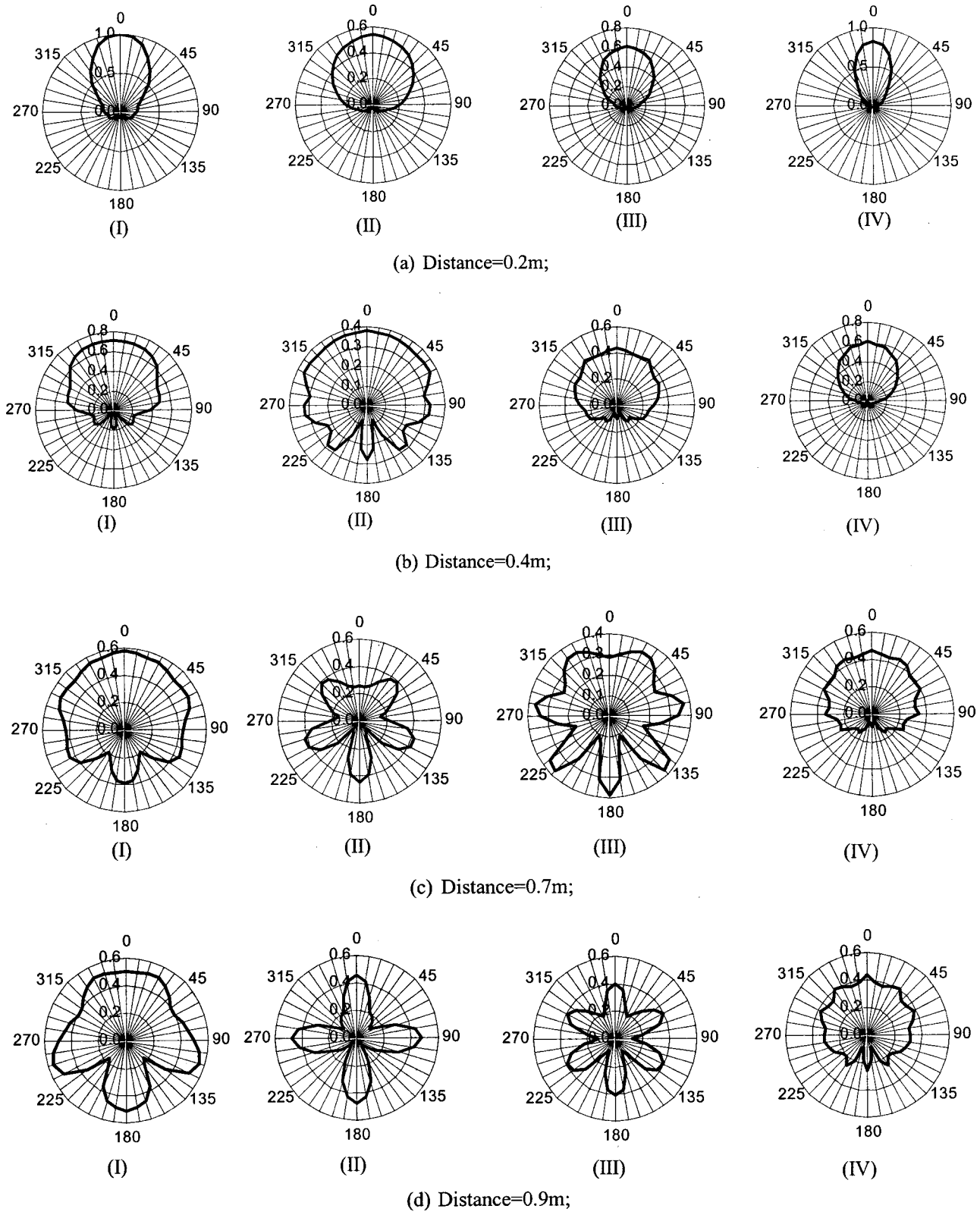
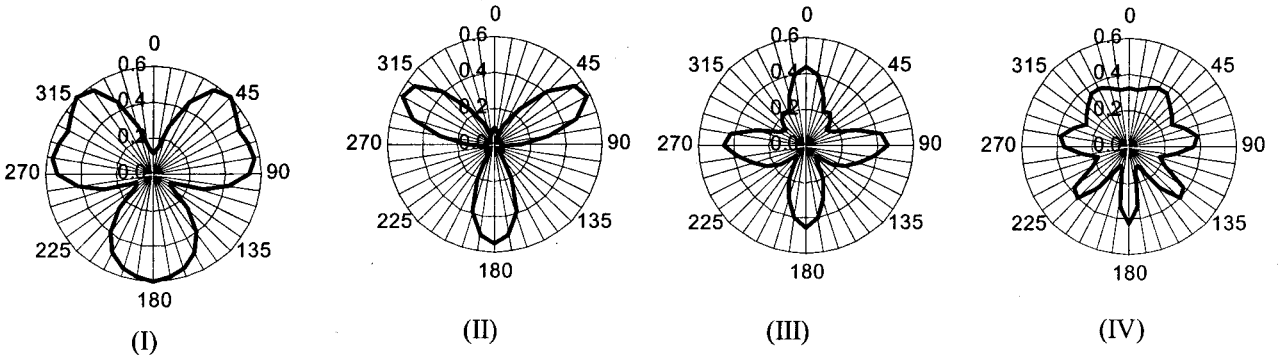


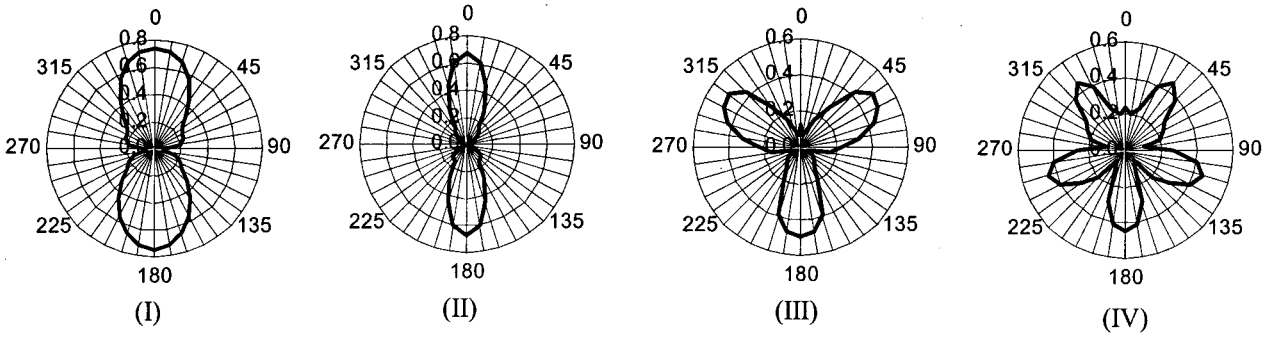
FIG. 7. Angular profiles at different distances on a carbon steel pipe with 88-mm o.d. and 5-mm wall thickness. Angle beam covers a 90-degree circumferential angle for (I) and a 45-degree circumferential angle for (II), (III) and (IV); (I) $L(M,2)$ modes at 270 KHz; (II) $L(M,2)$ modes at 270 KHz; (III) $L(M,2)$ modes at 375 KHz; and (IV) $L(M,1)$ modes at 375 KHz.

angular profiles together can cover almost all pipe areas at any of the propagating distances.

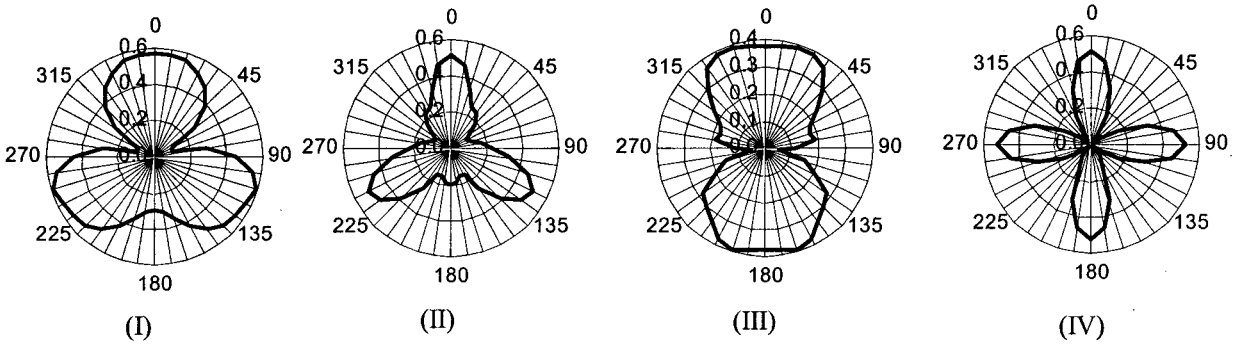
Since the total fields are the superimposition of all of the modes with different phase velocities, the slight difference of phase velocities leads to a variation of the phase match between the different modes, therefore leading to variations of the angular profiles along the propagating distance. This is why phase velocity difference plays an important role in the



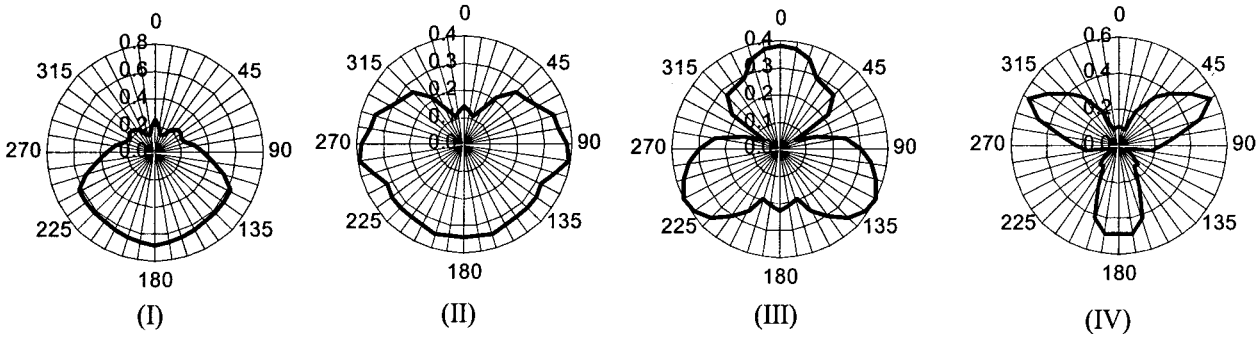
(e) Distance=1.25m;



(f) Distance=1.85m;



(g) Distance=2.45m;



(h) Distance=3.4m.

FIG. 7. (*Continued.*)

development of the angular profiles. Figure 7 indicates that the angular profiles of $L(M,2)$ at 290 KHz change faster than those of the other two groups. This can be explained based on the greater phase velocity difference between $L(M,2)$ modes shown in Fig. 8.

IV. EXPERIMENT

Experiments were conducted to verify the theoretical conclusions on the angular profiles. An eight-element angle beam transducer array is placed around an 88-mm-o.d. 5-mm

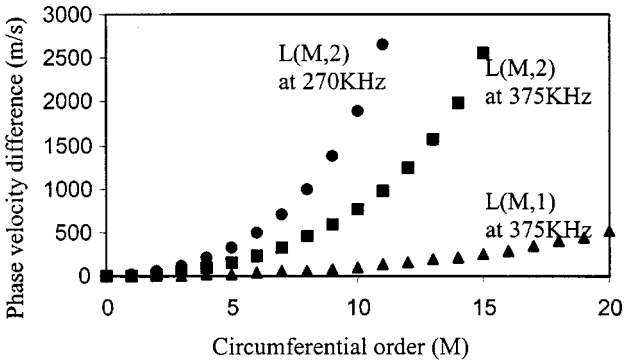


FIG. 8. The phase velocity differences between the flexural modes ($M>0$) and corresponding axisymmetric longitudinal modes ($M=0$) for a carbon steel pipe with 88-mm o.d. and 5-mm wall thickness.

wall thickness carbon steel pipe. The transducer array is shown in Fig. 9. All transducers are 0.5-MHz broadband piezocomposite transducers supplied by Krautkramer Branson. An angle beam sending transducer was placed at a specific distance from the eight-element receiving transducer array. The guided waves excited by the sender were received by each array element separately. As each element was placed at a different circumferential location, the signal amplitudes from all transducers produced an angular profile of the guided waves at that distance. Before the experiment was run, each transducer was calibrated for its sensitivity.

For the experiments, we set the frequency of the tone burst signal of 290 kHz and a transducer angle of 30 degrees. From the dispersion curves, we can see that $L(M,2)$ modes will be generated. If the transducer angle is set to 70 degrees, $L(M,1)$ modes will be generated. The pipe was 2.95 m long. We measured the angular profiles at distances from 0.85 to 2.7 m, as shown in Fig. 10. For comparison purposes, each measured angular profile was plotted together with the theoretical results. Figure 10 shows that the experiment results matched with the theory quite well.

In order to show the impact of the angular profiles on the inspection ability of guided waves, we ran an experiment to find an artificial defect by using the $L(M,2)$ modes at 290 kHz. Each element of the transducer array shown in Fig. 9 serves both a sender and receiver independently. At a distance of 1.8 m from the transducers, a 20 mm wide and 20% deep notch was fabricated on the outer surface of the pipe. The defect was at the same circumferential location as element 1. Both were on the top of the pipe. The angular profiles in Fig. 7 show us that if a transducer is placed on the top of the pipe, most of the energy of the guided waves would focus on only the top and bottom of the pipe at the distance of 1.8 m. Therefore, it was not surprising to find out that only

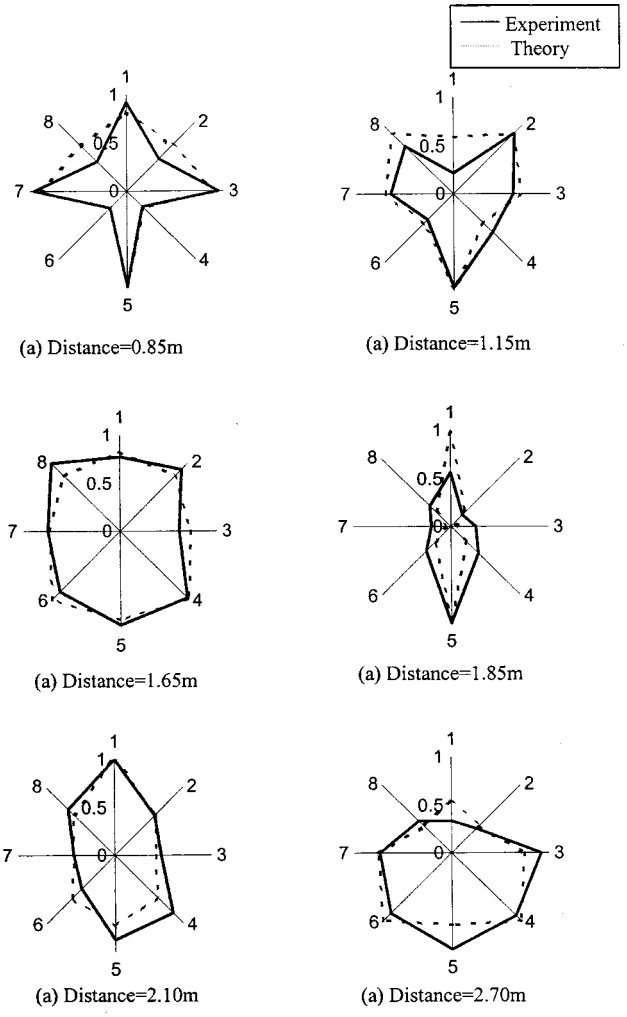


FIG. 10. Experimental and theoretical results of the angular profiles for the $L(M,2)$ modes at the frequency of 290 kHz.

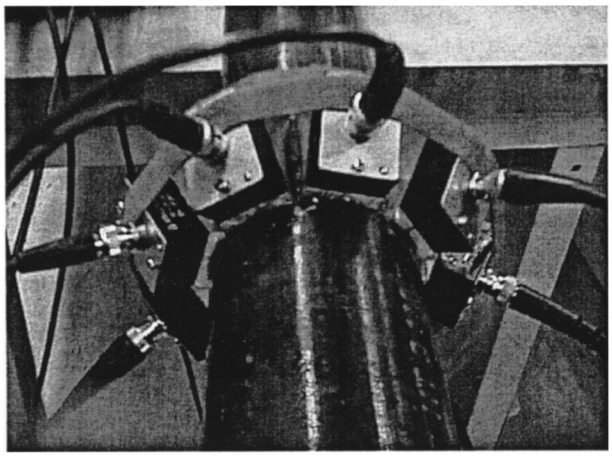


FIG. 9. Eight-element transducer array around a pipe.

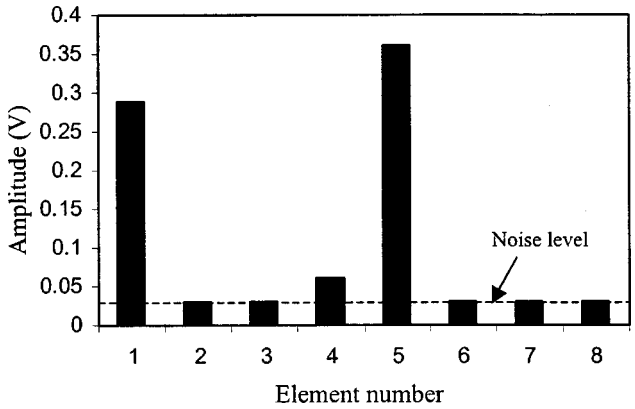
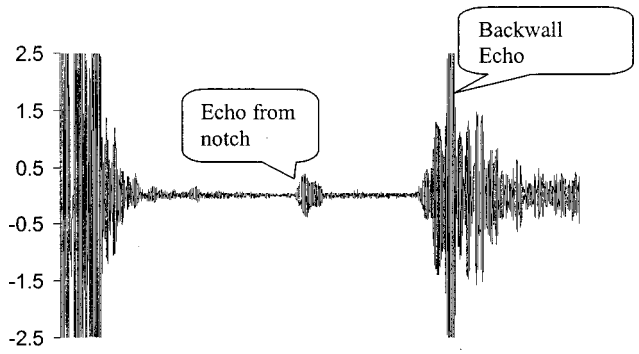
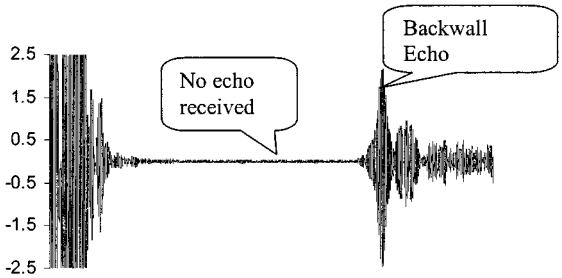


FIG. 11. Amplitude of the notch echo for each element.



(a) Pulse echo signal for element 5.



(b) Pulse echo signals for element 7.

FIG. 12. Sample pulse echo signals for elements 5 and 7. The notch is at the same circumferential location as element 1.

elements 1 and 5 could receive strong echoes from the notch (element 5 is on the bottom of the pipe). This is shown in Fig. 11. As an example, pulse echo waveforms for elements 5 and 7 are shown in Fig. 12. The difference between the signals shows the significant impact of the angular profiles on the testing ability of guided waves. In order to find the defects at any of the circumferential locations, we have to rotate the transducer around the pipe. It becomes possible to find a defect on the bottom of a pipe, for example, with a transducer on the top of the pipe. Alternatively, we can also sweep frequency or change the mode in order to change the angular profiles, indicating that non-axisymmetric guided waves are generally more tedious than axisymmetric guided waves for practical inspection. Software developments will simplify, however, the use of non-axisymmetric guided waves for efficient flaw detection and analysis. Given the angular profiles, we can then obtain further information about defects, such as circumferential locations, and potentially shape and size information as well.

V. CONCLUSIONS

The normal mode expansion method was used for studying a variety of pipe source loading problems. Discussions were conducted for such cases as normal beam, comb transducer and partial loading. Numerical computations were implemented for the partial loading case. The discussions are focused on the variance of angular profiles for different propagating distances. The changing variables for angular profiles include frequency, wave mode, propagating distance and source loading. The impact of the angular profiles on the testing ability of guided waves was also discussed. Although further consideration, like transducer rotation or frequency tuning, is needed for non-axisymmetric guided waves to find most all defects, non-axisymmetric guided waves can be used to locate both the distance and circumferential location based on the use of angular profiles. Further research is needed, however, to develop test algorithms for defect detection, location, classification, and sizing with non-axisymmetric loading.

ACKNOWLEDGMENTS

Thanks are given to the Lawrence Livermore National Laboratory and to Diane Chinn for technical and financial support of this project.

[1] D. C. Gazis, ''Three-dimensional investigation of the propagation of waves in hollow circular cylinders. I. Analytical foundation,'' J. Acoust. Soc. Am. **31**, 568–573 (1959).
[2] D. C. Gazis, ''Three-dimensional investigation of the propagation of waves in hollow circular cylinders. II. Numerical results,'' J. Acoust. Soc. Am. **31**, 573–578 (1959).
[3] B. A. Auld, *Acoustic Fields and Waves in Solids: Vol. II* (Wiley, New York, 1973).
[4] J. J. Ditri and J. L. Rose, ''Excitation of Guided Wave Modes in Hollow Cylinders by Applied Surface Tractions,'' J. Appl. Phys. **72**(7), 2589–2597 (1992).
[5] J. J. Ditri, ''Utilization of guided waves for the characterization of circumferential cracks in hollow cylinders,'' J. Acoust. Soc. Am. **96**, 3769–3775 (1994).
[6] H. J. Shin, ''Non-Axisymmetric Ultrasonic Guided Waves For Tubing Inspection,'' the Pennsylvania State University, Ph.D. thesis, May 1997.
[7] H. J. Shin and J. L. Rose, ''Guided Wave Tuning Principles for Defect Detection in Tubing,'' J. Nondestruct. Eval. **17**(1), 27–36 (1998).
[8] M. J. Quarry and J. L. Rose, ''Multimode Guided Wave Inspection of Piping Using Comb Transducers,'' Mater. Eval. **57**(10), 1089–1090 (1999).
[9] D. N. Alleyne and P. Cawley, ''The excitation of Lamb Waves in Pipes Using Dry-coupled Piezoelectric Transducers,'' J. Nondestruct. Eval. **15**(1), 11–20 (1996).
[10] D. N. Alleyne and P. Cawley, ''Long Range Propagation of Lamb Waves in Chemical Plant Pipework,'' Mater. Eval. **45**(4), 504–508 (1997).

Least-squares Legendre spectral element solutions to sound propagation problems

Wen H. Lin[a)]
Rocketdyne Propulsion & Power, The Boeing Company, P.O. Box 7922, 6633 Canoga Avenue, Canoga Park, California 91309-7922

(Received 15 March 2000; revised 20 October 2000; accepted 26 October 2000)

This paper presents a novel algorithm and numerical results of sound wave propagation. The method is based on a least-squares Legendre spectral element approach for spatial discretization and the Crank–Nicolson [Proc. Cambridge Philos. Soc. **43**, 50–67 (1947)] and Adams–Bashforth [D. Gottlieb and S. A. Orszag, *Numerical Analysis of Spectral Methods: Theory and Applications* (CBMS-NSF Monograph, Siam 1977)] schemes for temporal discretization to solve the linearized acoustic field equations for sound propagation. Two types of NASA Computational Aeroacoustics (CAA) Workshop benchmark problems [ICASE/LaRC Workshop on Benchmark Problems in Computational Aeroacoustics, edited by J. C. Hardin, J. R. Ristorcelli, and C. K. W. Tam, NASA Conference Publication 3300, 1995a] are considered: a narrow Gaussian sound wave propagating in a one-dimensional space without flows, and the reflection of a two-dimensional acoustic pulse off a rigid wall in the presence of a uniform flow of Mach 0.5 in a semi-infinite space. The first problem was used to examine the numerical dispersion and dissipation characteristics of the proposed algorithm. The second problem was to demonstrate the capability of the algorithm in treating sound propagation in a flow. Comparisons were made of the computed results with analytical results and results obtained by other methods. It is shown that all results computed by the present method are in good agreement with the analytical solutions and results of the first problem agree very well with those predicted by other schemes.
[DOI: 10.1121/1.1336137]

PACS numbers: 43.20.Mv [DEC]

I. INTRODUCTION

Acoustic waves propagating in a flow field are not only of theoretical interest but also of practical importance for the study of aeroacoustics and for the applications in industrial designs. By nature, sound propagation in a flow field is a complicated process: it can be the propagation of an external sound source in a specified flow or the propagation of the aerodynamic sound generated by a flow itself. The first type of problem mainly deals with the flow effects on the propagation of sound waves. The second kind of aeroacoustic problems involves aerodynamic sound generation and propagation and is normally more difficult to solve than the first one. This paper presents a least-squares spectral element method based on Legendre polynomials to solve the first type of aeroacoustic problems, namely, to treat the propagation of a given sound source in a specified flow. The effectiveness of the proposed algorithm and the flow effect on the sound propagation are studied.

Acoustic waves are considered as small perturbations to the mean flow, and they are transported by the mean flow in addition to their own propagation. To account for the flow effect on the propagation of sound waves, one needs to consider the flow and acoustic characteristics simultaneously. Normally, the time scale of acoustic propagation is different from that of flow convection unless the flow is at transonic speed. In addition, the acoustic pressure is much smaller than the mean-flow pressure by three or four orders of magnitude under a normal condition. Therefore, to compute the correct pressure level associated with sound propagation in a transient flow is quite challenging because disparate scales of acoustic and flow fields require stringent rules on numerical accuracy for producing a meaningful solution. With the advent of high-speed computing resources, many numerical schemes have been designed to advance the newly emerging field called computational aeroacoustics (CAA). This new field continues to lead academic, NASA, and industrial teams to develop a practical and efficient numerical tool for computing aeroacoustic signatures for engineering applications.

Since 1994 academia, NASA research centers, and NASA institutes for computer applications in science and engineering have sponsored three CAA workshops to address the numerical methods and issues relevant to solving the aerodynamic sound generation and propagation in a fluid flow.[1–3] The American Institute of Aeronautics and Astronautics (AIAA), the American Society of Mechanical Engineers (ASME), and the Confederation of European Aerospace Societies (CEAS) have also sponsored conferences and programs in computational aeroacoustics since the early 1990's. For example, in the 1995 ASME/JSME Fluids Engineering and Laser Anemometry Conference, there was a special volume addressing the algorithm and issues relating to, and practical usage of, computational aeroacoustics.[4] Also, in a recent joint conference of AIAA/CEAS there were six

[a)]Electronic mail: wen-hwang.lin@boeing.com

sessions of computational aeroacoustics dealing with the development of basic methodologies and the engineering applications of the subject matter.[5] Many articles on computational aeroacoustics have been published in the proceedings and journals since 1990. In general, the numerical methods of these publications can be roughly classified into the following schemes: the dispersion-relation-preserving (DRP) approach and its variations,[6–8] finite difference (FD) method,[9,10] and finite element (FE) method.[11,12] These references are only cited here as illustrating examples and not meant to be complete. More computational aeroacoustic literature can be found from these references.

In the present study we propose a high-order spectral scheme to treat the propagation of sound waves in a flow. Spectral methods offer the flexibility and accuracy via using high-order polynomials to accomplish the spatial interpretation of variables to be solved. Therefore, they can be useful when solving sound propagation problems in a flow field because high resolution is required to solve these problems. Since Gottlieb and Orszag[13] introduced the numerical analysis of spectral methods in the late 1970's, researchers have extensively studied and used the methods or the varied methods to solve various dynamic problems in fluid and solid mechanics. For instance, Patera and his co-workers used the isoparametric spectral element method based on Chebyshev polynomials to solve incompressible viscous flow problems.[14,15] Seriani[16] and Seriani and Priolo[17] used the Chebyshev spectral element method to solve the large-scale acoustic wave propagation in complex geological media. Komatitsch *et al.*[18,19] solved the seismic problems of elastic wave propagation using the Legendre spectral element method. Schumack and Schultz[20] used the spectral method with Chebyshev and Legendre polynomials to solve Stokes equations on nonstaggered grids. Malik *et al.*[21] solved the two-dimensional wall-bounded shear flows with a Fourier–Chebyshev spectral collocation method.

In the following sections, a least-squares spectral element method based on Legendre polynomials is first presented in detail and then the numerical results obtained by this method are discussed. The numerical examples were selected from the benchmark problems of NASA ICASE/LaRC Computational Aeroacoustics Workshop.[1] These examples were chosen to verify the applicability and accuracy of the proposed method in solving aeroacoustic problems of sound propagation. All computed results agree well with the analytical solutions and results of the first test problem also agree well with the results obtained by the dispersion-relation-preserving[6] and standard finite-difference schemes.[6]

II. LEAST-SQUARES LEGENDRE SPECTRAL ELEMENT ANALYSIS

In this section, we develop a least-squares Legendre spectral element analysis for spatial discretization to solve a partial differential equation of space and time. Let us consider an inhomogeneous, partial differential equation written in the following form:

$$[\mathbf{L}][\mathbf{u}]=[\mathbf{f}] \text{ in a domain } D, \tag{1}$$

where $[\mathbf{L}]$ contains the first-order partial differential operator in space and time, $\{\mathbf{u}\}$ the column vector of unknown variables, and $\{\mathbf{f}\}$ the column vector of forcing functions. First, we divide the domain D into a finite number of elements and then assume an approximate solution to Eq. (1) in a typical element can be written as

$$\{\mathbf{u}^e(t,x,y,z)\}=\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\{\mathbf{a}_{ijk}\}\phi_i^e(x)\phi_j^e(y)\phi_k^e(z) \tag{2}$$

for a three-dimensional problem, where $\phi_i^e(x), \phi_j^e(y), \phi_k^e(z)$ are linearly independent basis functions in the x, y, and z directions, $\mathbf{a}_{ijk}$ the expansion coefficients for the unknown variables, and N the total number of basis functions (or degrees of freedom) in each coordinate direction of an element. It is noted that the same kinds of basis functions are employed for all unknowns. This feature is a characteristic of the proposed algorithm and it simplifies the mathematical formulation and numerical implementation. Generally, the basis functions are arbitrary and need not satisfy the differential equation or boundary conditions. However, they must be differentiable once in the domain and at the boundaries to guarantee the C^0 continuity of field variables at the element boundaries.

Substituting Eq. (2) into Eq. (1), forming the integral of the square of residual over the domain, and applying the method of least-mean squares (e.g., Refs. 22 and 23) to the integral with respect to the unknown coefficients, one obtains a least-square element equation

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}K_{lmnijk}\mathbf{a}_{ijk}=\mathbf{F}_{lmn}, \tag{3}$$

for the unknown $\{\mathbf{a}_{ijk}\}$, where

$$\mathbf{K}_{lmnijk}=\int\int\int_e[\mathbf{L}\phi_l(x)\phi_m(y)\phi_n(z)]^T \times[\mathbf{L}\phi_i(x)\phi_j(y)\phi_k(z)]dx\,dy\,dz, \tag{4a}$$

$$\mathbf{F}_{lmn}=\int\int\int_e[\mathbf{L}\phi_l(x)\phi_m(y)\phi_n(z)]^T\mathbf{f}(t,x,y,z)dx\,dy\,dz, \tag{4b}$$

T indicates the transpose of a matrix. Equation (3) shows that the original differential equation becomes a discretized element equation with its coefficients, Eq. (4a), written in terms of the integral of products of the derivatives of basis functions over the domain. The forcing functions are also weighted by the derivatives of basis functions, as shown in Eq. (4b). These equations are fundamental of designing a numerical algorithm to solve the original differential equation.

In order to utilize the spectral element method to evaluate the integrals in Eqs. (4a) and (4b) with the aid of Gauss quadrature rules,[24] we transform the physical domain (x, y, z) into a computational space (ξ, η, ζ) with the aid of isoparametric tensor-product mapping[15,20] defined as

$$x(\xi,\eta,\zeta)=\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} x_{ijk}^{e}\phi_i(\xi)\phi_j(\eta)\phi_k(\zeta),$$

$$y(\xi,\eta,\zeta)=\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} y_{ijk}^{e}\phi_i(\xi)\phi_j(\eta)\phi_k(\zeta), \tag{5}$$

$$z(\xi,\eta,\zeta)=\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} z_{ijk}^{e}\phi_i(\xi)\phi_j(\eta)\phi_k(\zeta),$$

where x_{ijk}^{e}, y_{ijk}^{e}, z_{ijk}^{e} are the coordinates of the Nth node in the physical element, and ξ, η, ζ are the coordinates ranging from -1 to 1 in a computational element. Within each element we use the Lagrange interpolation functions, which in turn are in terms of Legendre polynomials, to accomplish nodal interpolation. The grid nodes in the computational domain are chosen as Legendre–Gauss–Lobatto collocation points,[20,25] which are the roots obtained by solving $(1-\xi^2)L_N'(\xi)=0$, $(1-\eta^2)L_N'(\eta)=0$, and $(1-\zeta^2)L_N'(\zeta)=0$ in the (ξ,η,ζ) space for a three-dimensional problem. The prime in these expressions indicates the derivative of L_N with respective to the corresponding argument and L_N is the Nth-order Legendre polynomial. For a two-dimensional problem only the first two expressions need be solved for the collocation points in the (ξ,η) space, and for the one-dimensional problems only the first one need be solved in the ξ space. As shown in Refs. 20 and 25, the Lagrange interpolation functions in the computational domain can be written in terms of Legendre polynomials as

$$\phi_\alpha(\xi)=\frac{-1}{N(N+1)L_N(\xi_\alpha)}\frac{(1-\xi^2)L_N'(\xi_\alpha)}{\xi-\xi_\alpha},$$

$$\phi_\beta(\eta)=\frac{-1}{N(N+1)L_N(\eta_\beta)}\frac{(1-\eta^2)L_N'(\eta_\beta)}{\eta-\eta_\beta}, \tag{6}$$

$$\phi_\gamma(\zeta)=\frac{-1}{N(N+1)L_N(\zeta_\gamma)}\frac{(1-\zeta^2)L_N'(\zeta_\gamma)}{\zeta-\zeta_\gamma},$$

with $\phi_\alpha(\xi_m)\equiv\delta_{\alpha m}$ (the Kronecker delta function), ξ_α, η_β, ζ_γ are Legendre–Gauss–Lobatto points, and α, β, γ are from 1 to N.

Using these coordinate transformation relations, one can evaluate the integrals in Eqs. (4a) and (4b) in a cube of volume of 8 units for a three-dimensional problem, in a square of area of 4 units for a two-dimensional problem, or in a line of length of 2 units for a one-dimensional problem. That is, for a three-dimensional problem Eqs. (4a) and (4b) become

$$\mathbf{K}_{lmnijk}=\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}(\mathbf{L}(\xi,\eta,\zeta)\phi_l(\xi)\phi_m(\eta)\phi_n(\zeta))^T \times(\mathbf{L}(\xi,\eta,\zeta)\phi_i(\xi)\phi_j(\eta)\phi_k(\zeta))J\,d\xi\,d\eta\,d\zeta, \tag{7a}$$

$$\mathbf{F}_{lmn}=\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}(\mathbf{L}(\xi,\eta,\zeta)\phi_l(\xi)\phi_m(\eta)\phi_n(\zeta))^T \times\mathbf{f}(t,\xi,\eta,\zeta)J\,d\xi\,d\eta\,d\zeta, \tag{7b}$$

where J is the Jacobian of the coordinate transformation. These two integrals can be computed exactly with the aid of Gauss quadrature rules[26] as

$$\mathbf{K}_{lmnijk}=\sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}\sum_{\gamma=1}^{N} W(\xi_\alpha)W(\eta_\beta)W(\zeta_\gamma) \times[\mathbf{L}\phi_l(\xi_\alpha)\phi_m(\eta_\beta)\phi_n(\zeta_\gamma)]^T \times[\mathbf{L}\phi_i(\xi_\alpha)\phi_j(\eta_\beta)\phi_k(\zeta_\gamma)]J(\xi_\alpha,\eta_\beta,\zeta_\gamma), \tag{8a}$$

$$\mathbf{F}_{lmn}=\sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}\sum_{\gamma=1}^{N} W(\xi_\alpha)W(\eta_\beta)W(\zeta_\gamma) \times[\mathbf{L}\phi_l(\xi_\alpha)\phi_m(\eta_\beta)\phi_n(\zeta_\gamma)]^T \times\mathbf{f}(t,\xi_\alpha,\eta_\beta,\zeta_\gamma)J(\xi_\alpha,\eta_\beta,\zeta_\gamma), \tag{8b}$$

where N is the number of Legendre–Gauss–Lobatto points in each integration direction, and W the weighting factors for the Gauss–Legendre quadrature, which again can be written in terms of Legendre polynomials (for example, see Refs. 20 and 25) as

$$W_\alpha=\frac{2}{N(N+1)[L_N(\xi_\alpha)]^2},\quad W_\beta=\frac{2}{N(N+1)[L_N(\eta_\beta)]^2},$$

$$W_\gamma=\frac{2}{N(N+1)[L_N(\zeta_\gamma)]^2}.$$

At this point we have cast the least-squares finite-element equation [i.e., Eq. (3)] into a least-squares Legendre *spectral* element equation with the aid of Eqs. (8a) and (8b).

It should be noted that the time discretizaton is still open at this moment and that it can be accomplished by various second- or higher-order schemes such as Crank–Nicholson[27] or Adams–Bashforth[13] method. Details of time differencing are shown in the following numerical examples. In the first example the Crank–Nicholson scheme was used to discretize the time derivative, and in the second example the Adams–Bashforth scheme was used.

The spectral element equation plus an appropriate time discretization can be solved with initial and boundary conditions for the unknown coefficients, $\{\mathbf{a}_{ijk}\}$, and for other field variables at every node at any time. In the following two sections, the method of least-squares Legendre spectral element is applied to solve the propagation of a one-dimensional Gaussian pulse without a flow, and the reflection of a two-dimensional Gaussian pulse from a rigid wall in a uniform flow of Mach 0.5. These problems were chosen from the benchmark problems of NASA Computational Aeroacoustics Workshop[1] to test the property and accuracy of the proposed method.

III. PROPAGATION OF ONE-DIMENSIONAL WAVES

Consider the initial value problem for the propagation of one-dimensional convective wave in a domain of x between -20 and 450. The differential equation for this wave propagation problem can be written as[1,6]

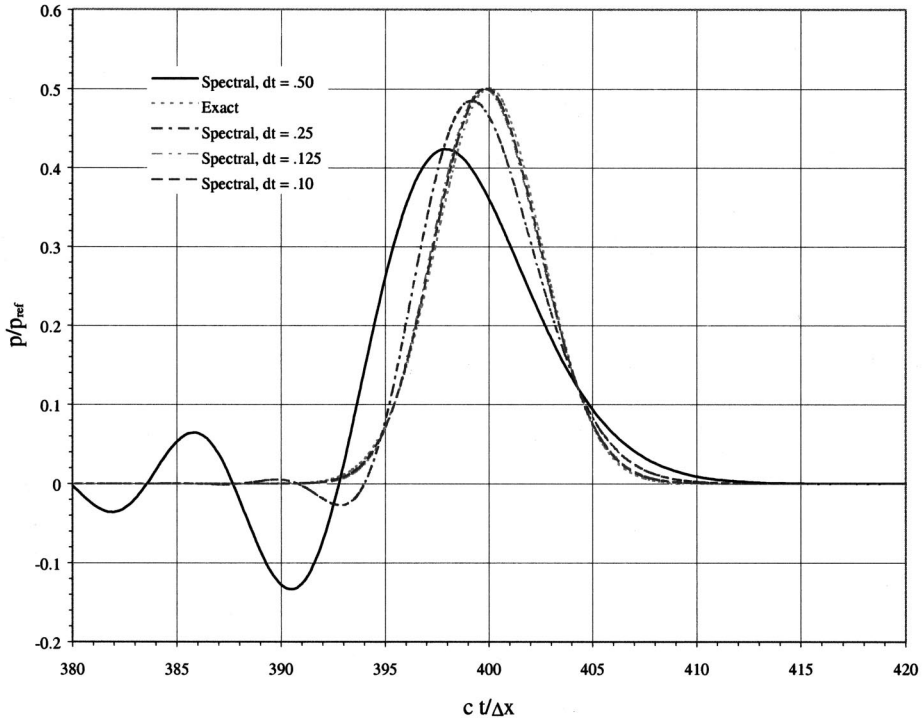


FIG. 1. Variations of pressure amplitudes with time step size for $b=3$ and $t=400$.

$$\frac{\partial p}{\partial t}+\frac{\partial p}{\partial x}=0, \quad -20\leqslant x\leqslant 450,$$
$$p=0.5\exp\left[-\ln 2\left(\frac{x}{b}\right)^2\right] \text{ at } t=0, \tag{9}$$

where x is normalized with grid spacing Δx, t normalized with $\Delta x/c$ ($c=$speed of sound), p the sound pressure normalized with ρc^2 ($\rho=$fluid density), and b the width of half maximum. The exact solution to this problem is

$$p=0.5\exp\left[-\ln 2\left(\frac{x-t}{b}\right)^2\right]. \tag{10}$$

To solve Eq. (9) via the proposed method, we write the differential wave equation in a semidiscrete form as

$$\left(1+\alpha\Delta t\frac{\partial}{\partial x}\right)p=\left(1-[1-\alpha]\Delta t\frac{\partial}{\partial x}\right)p^0, \tag{11}$$

where p^0 is the initial condition, Δt the time step size, and α is a parameter representing various time-marching schemes.

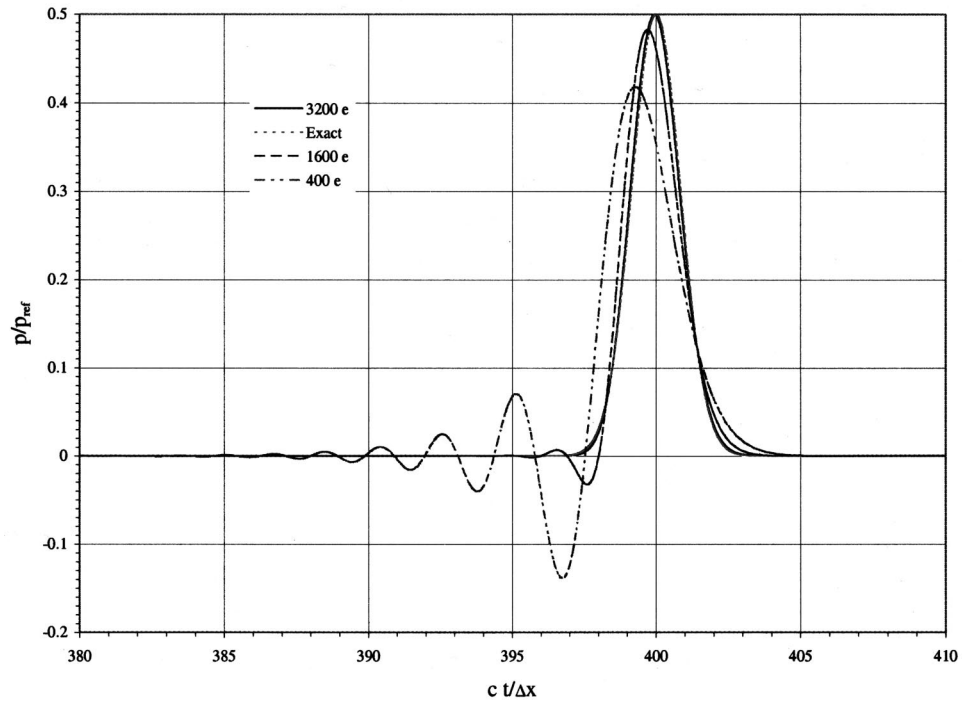


FIG. 2. Variations of pressure amplitudes with element number for $b=1$ and $t=400$.

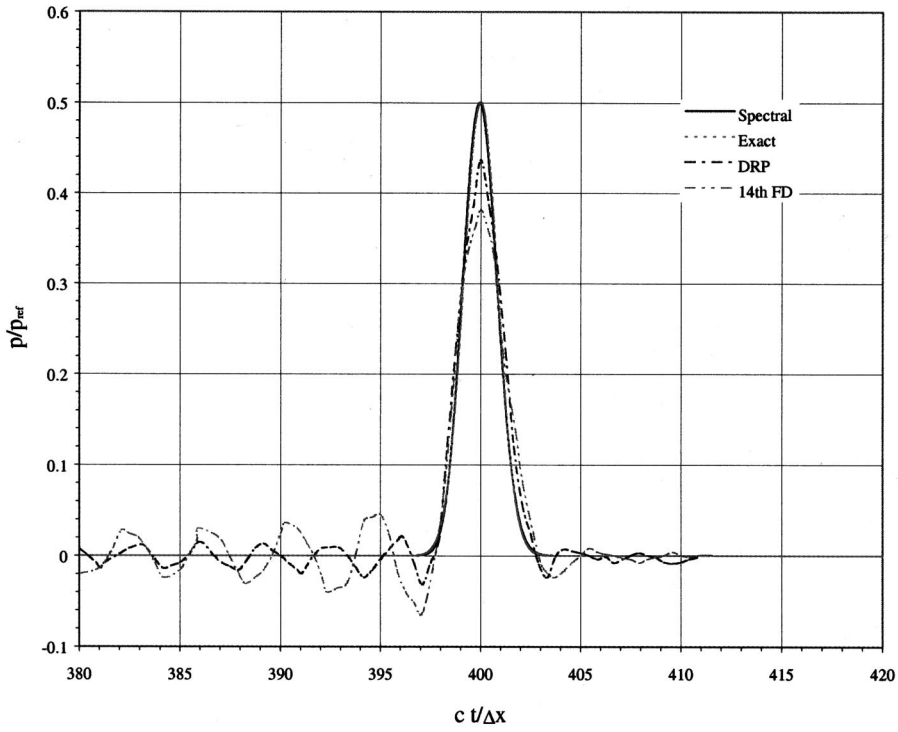


FIG. 3. Comparisons of numerical pressure amplitudes with analytical result for $b=1$ and $t=400$.

For instance, when α is set to 0.5, the algorithm is the Crank–Nicolson scheme, and if α is set to 1, the algorithm is the backward difference scheme. In the present study α is always set to 0.5; therefore, the time discretization for this problem is of a second-order scheme. Equation (11) is in the same form as Eq. (1) with the unknown p, $[1+\alpha\Delta t(\partial/\partial x)]$ the matrix operator, and the right-hand side as the forcing function. Following the formulation described in Sec. II, one has the least-squares spectral element equation for the propagation of one-dimensional waves as

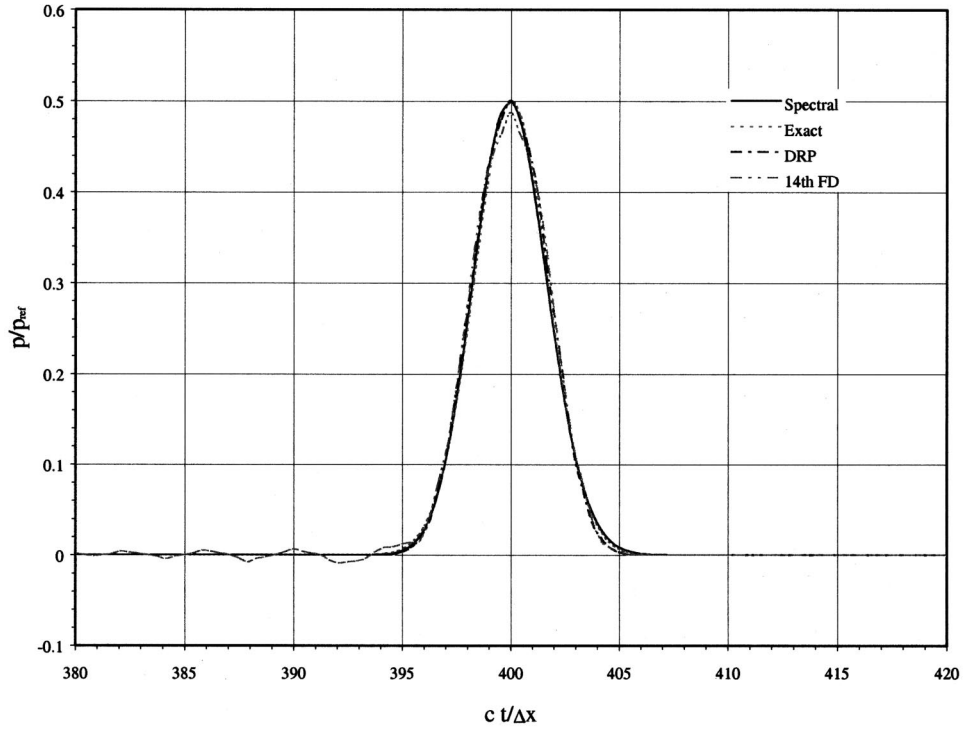


FIG. 4. Comparisons of numerical pressure amplitudes with analytical result for $b=2$ and $t=400$.

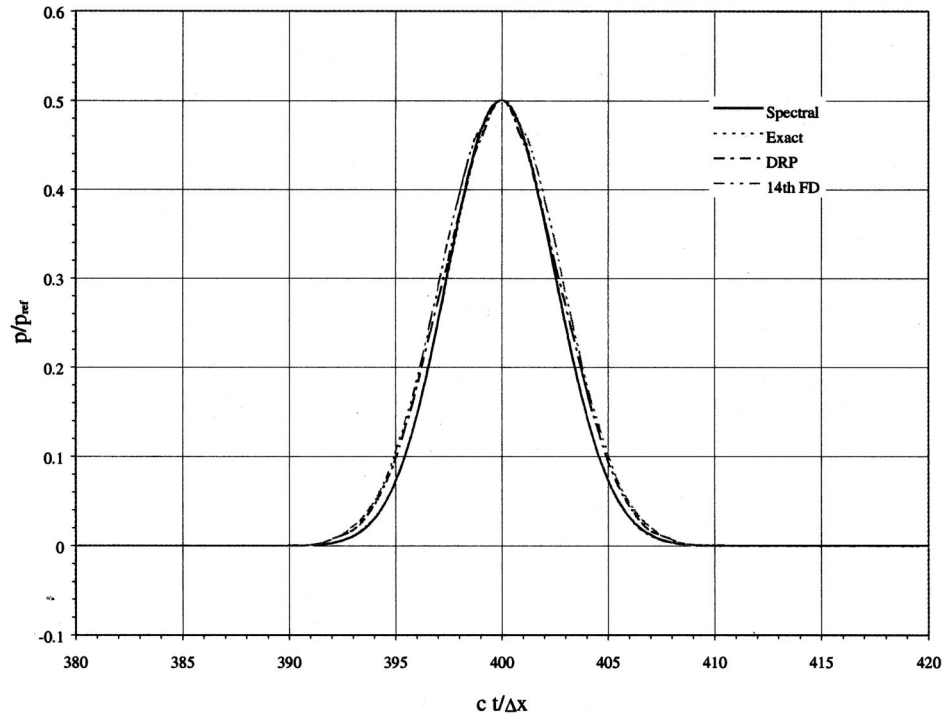


FIG. 5. Comparisons of numerical pressure amplitudes with analytical result for $b=3$ and $t=400$.

$$\sum_{k=1}^{N}\sum_{l=1}^{N}\left[\delta_{kl}+\frac{\alpha\Delta t}{J(\xi_l)}\frac{\partial\phi_k(\xi_l)}{\partial\xi}\right]\left[\delta_{il}+\frac{\alpha\Delta t}{J(\xi_l)}\frac{\partial\phi_i(\xi_l)}{\partial\xi}\right]\times J(\xi_l)W(\xi_l)a_k$$
$$=\sum_{k=1}^{N}\sum_{l=1}^{N}\left[a_k^0\delta_{kl}-(1-\alpha)\frac{\Delta t}{J(\xi_l)}\frac{\partial\phi_k(\xi_l)}{\partial\xi}a_k^0\right]\times\left[\delta_{il}+\frac{\alpha\Delta t}{J(\xi_l)}\frac{\partial\phi_i(\xi_l)}{\partial\xi}\right]J(\xi_l)W(\xi_l), \quad (12)$$

where i is from 1 to N, δ_{ij} the Kronecker delta function, J the Jacobian of coordinate transformation, W the weighting factor for the Gauss–Legendre quadrature, ξ_i the Legendre–Gauss–Lobatto points, a_k^0 is the initial condition relating to p^0, and a_k the unknown to be solved. This equation was solved for sound propagation with the initial pressure given in Eq. (9) for three values of half-maximum width (namely, $b=1,2,3$).

Figure 1 shows the variations of pressure amplitudes with the size of time step for the case of $b=3$. The number of elements and the order of Legendre polynomial used for calculating these results are 200 and 6, respectively. As seen from this figure, the wave amplitude decreases and the wave tail oscillates when the time step size is greater than 0.25. Namely, the wave is dissipative and dispersive when the time step is too large. The wave amplitude drops 4% for $dt=0.25$ and 16% for $dt=0.5$, compared with the exact value. From numerical experiments, it is known that for a fixed size of grid the larger time step is harder than the smaller time step to satisfy the numerical stability condition, namely, $\Delta t/\Delta x\leqslant 1$ in this problem, where Δt and Δx are the time step and grid size, respectively. Therefore, to compute the correct wave amplitude and phase with a given grid, one needs to perform a sensitivity study of time step and grid size (whose effect is discussed below) to prevent numerical dissipation and dispersion.

Figure 2 shows the effect of grid size on pressure amplitudes for $b=1$ with a fixed time step and a fixed order of the interpolation function (i.e., Legendre polynomial). This figure also shows the comparison of the computed results with the analytical solution. As seen from this figure, the wave tail oscillates when the number of elements of a grid is small. In the case of $b=1$ the pulse is very sharp like a delta function, so that it is difficult to model the pulse motion correctly. When the grid is coarse (or the number of elements is small) and not able to resolve the waveform, the computed amplitude is dissipated and its phase is oscillating. For example, the wave oscillates in the trailing edge and does not have the right peak value at the correct time for 400- and 1600-element grids. Also, the wave amplitude drops for the

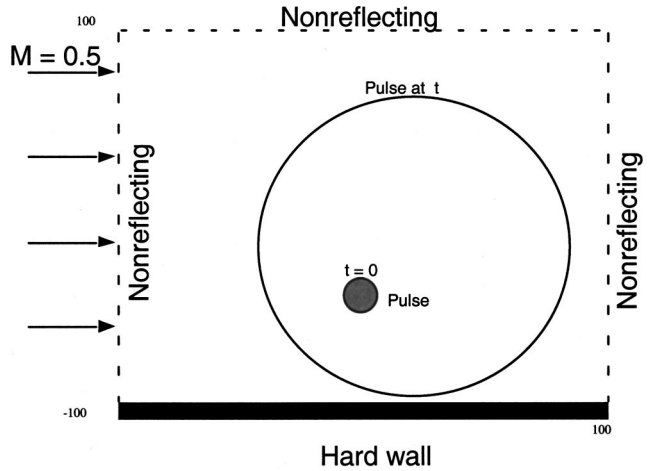


FIG. 6. Geometry for sound waves reflected from a wall in a uniform flow of Mach 0.5.

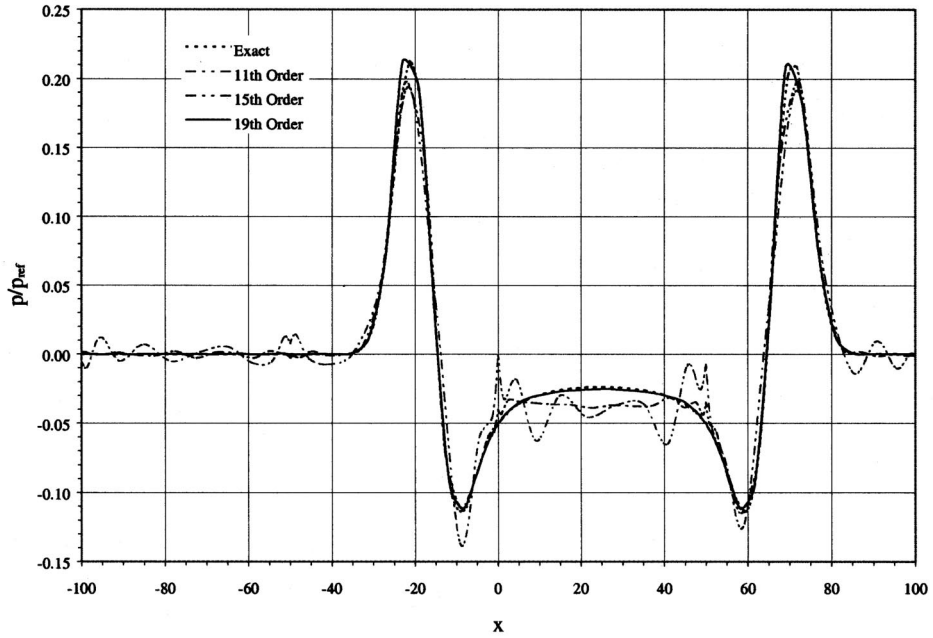


FIG. 7. Effect of grid size on the pressure waveform along the line $y=0$, $t=60$, $p_{\text{ref}}=\rho c^2$, where c is the speed of sound and ρ the fluid density.

400- and 1600-element grids because these grids are not fine enough to resolve the wave magnitudes and shapes. Although the results presented in this figure are for $b=1$, it is expected that similar phenomena will be observed for $b=2$ and 3. However, for a broader pulse fewer grid points are needed to resolve the wave shape, phase, and magnitude.

Figures 3–5 show the comparisons of computed results of various numerical schemes with the analytical results for $b=1,2,3$. It is seen from these figures that when the pulse is narrow (namely, $b=1$), both DRP and standard finite-difference schemes[6] do not predict correct wave magnitude and phase. At $b=2$, both spectral and DRP schemes predict the correct waveform and magnitude, and both are comparable with the analytical solution. However, the 14th-order finite-difference method still does not compute the correct wave phase and magnitude for $b=2$. When the waveform is broader (i.e., $b=3$), all three schemes predict correct results, which are very close to the analytical solution.

Typically, the CPU (central process unit) time for computing the propagation of a broad pulse (whose $b=3$), on a Sun Ultrasparc II single processor of 450 MHz, for $ct/\Delta x=400$ with a time step $\Delta t=0.05$ on a grid of 1000 points is 60 s with the present method, 45 s with the 15-point DRP scheme, and 50 s with the 14th-order finite-difference scheme. The current spectral method uses a grid of 50 elements and the 20th-order Legendre polynomials to obtain the solution, which is of the same accuracy as those obtained by the DRP and FD methods. Normally, if the time-advancing scheme is the same for all methods, then the spectral element method normally takes fewer points to achieve the same accuracy. Therefore, using fewer grid points will compensate the longer CPU time of using the spectral method.

IV. REFLECTION OF AN ACOUSTIC PULSE FROM A RIGID WALL

Consider the reflection of an acoustic pulse off a rigid wall in the presence of a uniform flow of Mach 0.5 in a semi-infinite space bounded by an acoustically rigid wall at $y=0$, as shown in Fig. 6. The acoustic pulse originally located at (0, 25) is released at $t=0$ and transported by the uniform flow. The governing equations for the pulse propagation written in a matrix are

$$\begin{bmatrix} \frac{\partial}{\partial t}+M\frac{\partial}{\partial x} & \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial t}+M\frac{\partial}{\partial x} & 0 \\ \frac{\partial}{\partial y} & 0 & \frac{\partial}{\partial t}+M\frac{\partial}{\partial x} \end{bmatrix} \begin{Bmatrix} p \\ u \\ v \end{Bmatrix} = 0, \tag{13}$$

with M being the Mach number and equal to 0.5 in this problem. Using the Adams–Bashforth scheme with three-time levels as in Ref. 13, one obtain a semidiscrete equation for the above matrix equation as

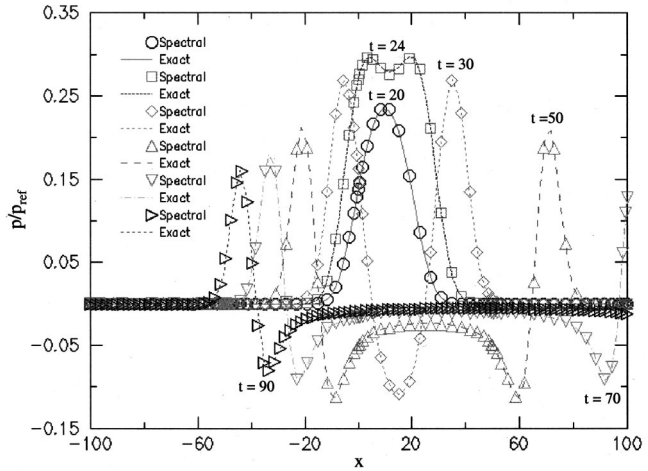


FIG. 8. Comparisons of sound pressures along the line $y=0$ with the exact solutions at various time, $p_{\text{ref}}=\rho c^2$, where c is the speed of sound and ρ the fluid density.

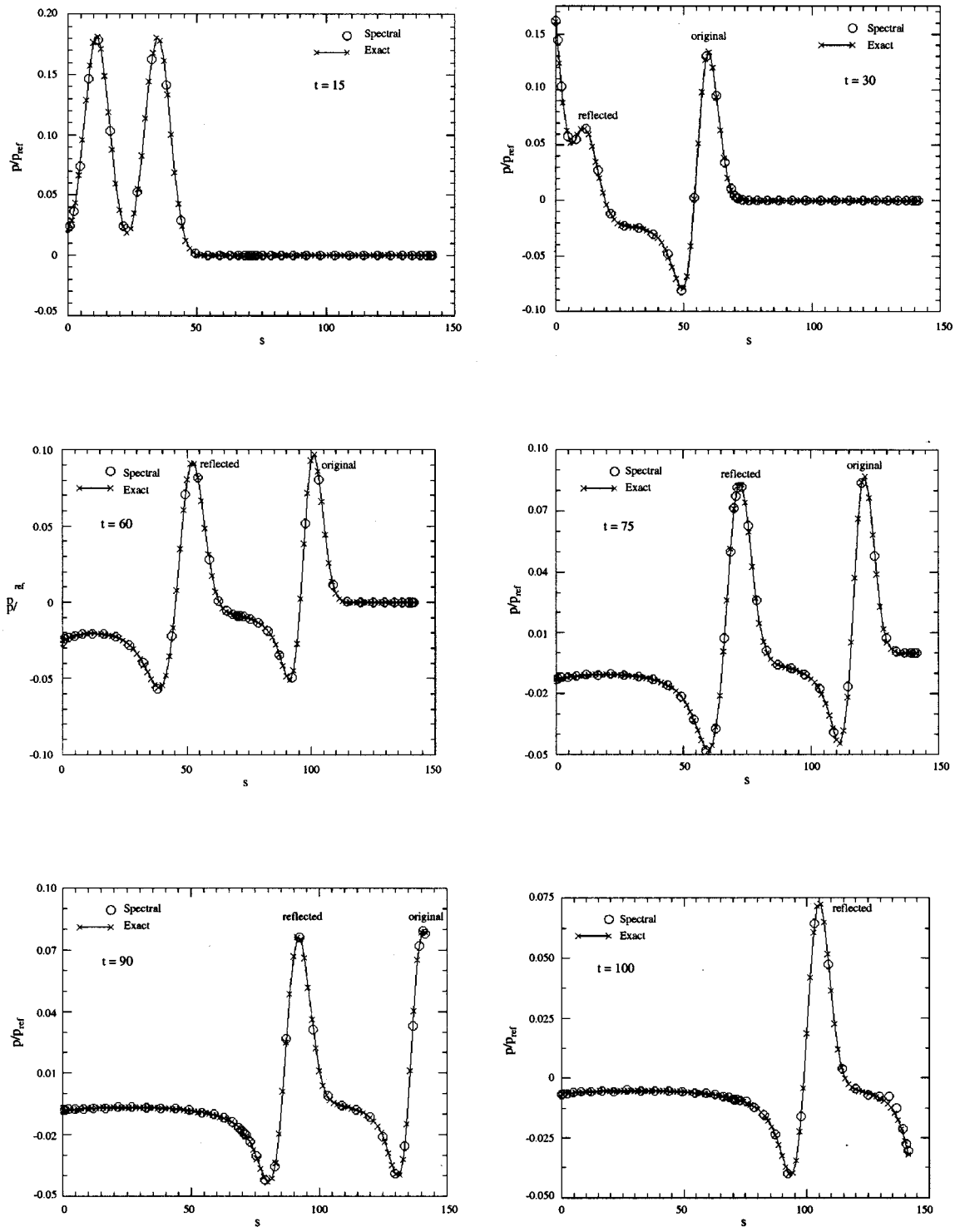


FIG. 9. Pressure waveforms along the line $y=x$, $p_{\mathrm{ref}}=\rho c^2$, where c is the speed of sound and ρ the fluid density, $s=(x^2+y^2)^{0.5}$.

$$\begin{bmatrix} \alpha_1+\Delta t M\dfrac{\partial}{\partial x} & \Delta t\dfrac{\partial}{\partial x} & \Delta t\dfrac{\partial}{\partial y} \\ \Delta t\dfrac{\partial}{\partial x} & \alpha_1+\Delta t M\dfrac{\partial}{\partial x} & 0 \\ \Delta t\dfrac{\partial}{\partial y} & 0 & \alpha_1+\Delta t M\dfrac{\partial}{\partial x} \end{bmatrix} \begin{Bmatrix} p^{n+1} \\ u^{n+1} \\ v^{n+1} \end{Bmatrix} = -\begin{Bmatrix} \alpha_2 p^n+\alpha_3 p^{n-1} \\ \alpha_2 u^n+\alpha_3 u^{n-1} \\ \alpha_2 v^n+\alpha_3 v^{n-1} \end{Bmatrix}, \tag{14}$$

where $\alpha_1=1$, $\alpha_2=1.5$, and $\alpha_3=-0.5$ are the time-differencing parameters associated with the Adams–Bashforth scheme. The square matrix and the column matrix on the left-hand side are the system matrix [**L**] and the unknowns, respectively, and the column matrix on the right-hand side is the initial condition acting as a forcing function. This equation can be used to solve the propagation of sound waves in a uniform flow with a set of initial and boundary conditions. The nondimensional initial conditions[1] for this problem are $u=0$, $v=0$, and

$$p=\exp\left[-\ln 2\left\{\frac{x^2+(y-25)^2}{25}\right\}\right]. \tag{15}$$

Following the formulation outlined in Sec. II, we have developed a least-squares Legendre spectral element code and computed the sound pressures associated with the pulse propagation at various times. The total number of elements used in all calculations for this problem is 16. The computed results are shown in Figs. 7–9 with the comparison of analytical results. It is noted that all exact pressures were computed from the analytical solution of Ref. 1 at the same grid nodes as in the spectral method.

Figure 7 presents the effect of grid size (related to the order of Legendre polynomial) on the computed results. It is seen that both the 11th- and 15th-order polynomials are not adequate to resolve the wave amplitude and phase. The waveform oscillates for the grids with these two polynomials. As the order of spectral polynomial increases, the computed result agrees very well with the exact solution. Again, the coarse grid causes numerical dispersion such that the computed waveforms oscillates and the wave amplitudes are not right. From the numerical experiments on this problem, it is found that normally it takes about eight nodes to represent a complete wave. All grids used for the following calculations employ the 19th-order Legendre polynomial and have eight points or more to resolve a complete wave.

Figure 8 shows the comparison of the pressure waveform along the wall (i.e., $y=0$) and Fig. 9 shows that along the line $y=x$. It is seen from these figures that the computed and analytical results are in good agreement. The maximum error defined as the absolute value of the difference between the exact and computed results varies between 10^{-4} and 10^{-7}. Figure 8 shows that the pulse has a maximum before it reaches the rigid wall and has two maximums after it is reflected from the wall (namely, after $t=24$). The two maximums are due to the reinforcement effect of the reflected wave to the original wave, and they are symmetric with respect to the center of the pulse and their magnitude gradually decreases when time is increasing. Eventually the forward wavefront (referred as the wave in the direction of 0°) leaves the downstream boundary when the nondimensional time, t, greater than 70. This forward wave moves three times faster than the backward wave (referred as the wave in the direction of 180°), as can be seen from the locations of the maximums or the minimums when t is greater than 30 (namely, at $t=50,70,90$).

Figure 9 shows another interesting phenomena about the pulse propagation as time progresses. This figure also shows the comparison of computed pressures, along the line $y=x$, with the exact pressures at various times. The 45° line, $y=x$, cuts through different parts of the original and reflected waves as time increases. For example, when $t=15$, the pulse does not reach the rigid wall so that the two maximums belong to the original wave. At $t=30$, the pulse has been reflected from the rigid wall; therefore, there is one hump associated with the reflected wave and one for the original wave. Because the pulse is just reflected from the wall at $t=30$, the 45° line only cuts a small part of the reflected wave and the corresponding hump is small. As time progresses such that the nondimensional time equal to or greater than 60 (i.e., $t=60,75,90$), the 45° line cuts through the original and reflected waves at points of nearly equal amplitude. At $t=90$, the line barely cuts the original wave because the original pulse front moves further and almost passes that line. At $t=100$, the line only cuts the reflected wave because the original pulse already moves far from the 45° line.

V. CONCLUDING REMARKS

A least-squares spectral element method has been presented with numerical solutions to the problems of one- and two-dimensional sound propagation. The computed results compare very well with the analytical results that were calculated at the same grid nodes from the exact solutions. For the one-dimensional wave, comparison was also made with the results obtained by the DRP and standard finite-difference schemes.[6] In general, the comparisons agree very well for the broad pulses (i.e., for the pulses with half-maximum width $b=2$ and 3). When the pulse width becomes narrow such that it resembles a delta function, it is hard to model the pulse motion accurately with the standard finite-difference or DRP scheme.[6] To compute the correct waveform, phase, and amplitude for a pulse with the half-maximum width $b=1$, we used 3200 elements and the eighth-order Legendre polynomial for a domain of nondimensional length of 470. The result computed with this grid agrees very well with the exact solution. Overall, results obtained from the present study indicate that the least-squares Legendre spectral element method is accurate to treat sound propagation in flows or in a domain without flow. This property makes the method very attractive to developing an accurate tool to treat the second type of aeroacoustic problems, i.e., the aerodynamic sound generation and propagation.

[1] ICASE/LaRC Workshop on Benchmark Problems in Computational Aeroacoustics, edited by J. C. Hardin, J. R. Ristorcelli, and C. K. W. Tam, NASA Conference Publication 3300, 1995.

[2] Second Computational Aeroacoustics (CAA) Workshop on Benchmark Problems, edited by C. K. W. Tam and J. C. Hardin, NASA Conference Publication 3352, 1997.

[3] Third Computational Aeroacoustics (CAA) Workshop on Benchmark Problems, NASA Glenn Research Center, November 8–10, 1999.

[4] Computational Aeroacoustics, edited by A. S. Lyrintzis, R. R. Mankbadi, O. Baysal, and M. I. Hitachi, ASME Fed-Vol 219, 1995.

[5] A Collection of Technical Papers, 5th AIAA/CEAS Aeroacoustics Conference and Exhibit, Volumes 1–2, 1999.

[6] C. K. W. Tam, ''Applied Aero-Acoustics: Prediction Methods,'' Numerical Methods in Computational Aeroacoustics, Lecture Series 1996-04, von Karman Institute for Fluid Dynamics, 1996, pp. 6–10.

[7] F. O. Hu, M. Y. Hussaini, and J. Manthey, ''Low-dissipation and Low-dispersion Runge–Kutta Schemes for Computational Acoustics,'' J. Comput. Phys. **124**, 177–191 (1996).

[8] O. Baysal, D. K. Kaushik, and M. Idres, ''Low Dispersion Scheme for Nonlinear Acoustic Waves in Nonuniform Flow,'' AIAA Paper 97-1582, Proceedings 3rd CEAS/AIAA Aeroacoustics Conference 1997.
[9] J. W. Goodrich and J. Hardin, ''Accurate Finite Difference Algorithms for Computational Aeroacoustics,'' in *Computational Fluid Dynamics Review*, edited by M. M. Hafez and K. Oshima (1998).
[10] Y. Ozyoruk and L. N. Long, ''Computation of Sound Radiating from Engine Inlets,'' AIAA J. **34**, 894–901 (1996).
[11] W. Krebs, G. Walz, and S. Hoffmann, ''Thermoacoustic Analysis of Annular Combustor,'' AIAA Paper 99-1971, 5th AIAA/CEAS Aeroacoustics Conference and Exhibit 1999.
[12] C. Peyret and P. Malbequi, ''A Finite Element Method for Duct Acoustics Analysis,'' 3rd Computational Aeroacoustics (CAA) Workshop on Benchmark Problems, NASA Glenn Research Center 1999.
[13] D. Gottlieb and S. A. Orszag, Numerical Analysis of Spectral Methods: Theory and Applications (CBMS-NSF Monograph, SIAM 1977), p. 110.
[14] A. T. Patera, ''A Spectral Element Method for Fluid Dynamics: Laminar Flow in a Channel Expansion,'' J. Comput. Phys. **54**, 468–488 (1984).
[15] K. Z. Korczak and A. T. Patera, ''An Isoparametric Spectral Element Method for Solution of the Navier–Stokes Equations in Complex Geometry,'' J. Comput. Phys. **62**, 361–382 (1986).
[16] G. Seriani, ''3-D Large-scale Wave Propagation Modeling by Spectral Element Method on Cray T3E Multiprocessor,'' J. Comput. Meth. Appl. Mech. Eng. **164**, 235–247 (1998).
[17] G. Seriani and E. Priolo, ''Spectral Element Method for Acoustic Wave Simulation in Heterogeneous Media,'' Finite Elem. Anal. Design **16**, 337–348 (1994).
[18] D. Komatitsch, J. P. Vilotte, R. Vai, J. M. Castillo-Covarrubias, and F. J. Sanchez-Sesma, ''The Spectral Element Method for Elastic Wave Equations—Application to 2D and 3D Seismic Problems,'' Int. J. Numer. Methods Eng. **45**, 1139–1164 (1999).
[19] D. Komatitsch and J. P. Vilotte, ''The Spectral Element Method: an Efficient Tool to Simulate the Seismic Response of 2D and 3D Geological Structures,'' Bull. Seismol. Soc. Am. **88**, 368–392 (1998).
[20] M. R. Schumack, W. W. Schultz, and J. P. Boyd, ''Spectral Method Solution of the Stokes Equations on Nonstaggered Grids,'' J. Comput. Phys. **94**, 30–58 (1991).
[21] M. R. Malik, T. A. Zang, and M. Y. Hussaini, ''A Spectral Collocation Method for the Navier–Stokes Equations,'' J. Comput. Phys. **61**, 64–88 (1985).
[22] O. C. Zienkiewicz, *The Finite Element Method* (McGraw-Hill, London, 1977), p. 87.
[23] D. S. Burnett, *Finite Element Analysis: From Concepts to Applications* (Addison-Wesley, Reading, MA, 1987), p. 68.
[24] A. H. Stroud and D. Secrest, *Gaussian Quadrature Formulas* (Prentice-Hall, Englewood Cliffs, NJ, 1966), pp. 1–2.
[25] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral Method in Fluid Dynamics* (Springer, Berlin, 1988), pp. 60–65.
[26] P. Davis and P. Rabinowitz, ''Abscissas and Weights for Gaussian Quadratures of High Order,'' J. Res. Natl. Bur. Stand. **56**, RP2645 (1956).
[27] J. Crank and P. Nicolson, ''A Practical Method for Numerical Evaluation of Solutions of Partial Differential Equations of the Heat-Conduction Type,'' Proc. Cambridge Philos. Soc. **43**, 50–67 (1947).

Improved solution for the vortical and acoustical mode coupling inside a two-dimensional cavity with porous walls

Joseph Majdalani
Mechanical and Industrial Engineering Department, Marquette University, Milwaukee, Wisconsin 53233

(Received 2 September 1999; revised 10 November 2000; accepted 17 November 2000)

This work presents an improved solution to a former study that analyzes the oscillatory motion of gases prescribed by vortico-acoustical mode coupling inside a two-dimensional porous cavity. The physical problem arises in the context of an oscillating gas inside a rectangular enclosure with wall transpiration, sublimation, or sweating. Previously, a multiple-scale solution was derived for the temporal field. The asymptotic formulation was based on an unconventional choice of scales. Its accuracy was also commensurate with the size of ε, a parameter that captured the effect of small viscosity. Currently, an exact solution is derived and compared to the previous formulation. A simple WKBJ solution is also constructed for validation purposes. Unlike both asymptotic formulations, the exact solution remains accurate regardless of the range of physical parameters.
 [DOI: 10.1121/1.1340648]

PACS numbers: 43.20.Bi, 43.20.Hq, 43.20.Mv, 43.28.Py [LCS]

I. INTRODUCTION

The purpose of this work is to provide both exact and asymptotic solutions to a boundary value problem that arises in the context of a fluid oscillating inside a rectangular cavity with transpiring walls. The corresponding mean-flow interactions with oscillatory motion involve time-dependent inertial, convective, and diffusive mechanisms. In a previous article,[1] an approximate solution was obtained using multiple-scale perturbation tools. The perturbation technique used was unconventional in the sense that it relied on a nonlinear scaling transformation. This was necessitated by the intricate inner and outer scaling structures caused by complex interactions of convective, inertial, and diffusive mechanisms. The resulting asymptotic formulation was found to be useful in characterizing the time-dependent field. The latter was a by-product of vortical and acoustical mode interactions that were driven, at the solid boundaries, by the acoustic pressure field. The asymptotic solution found by Majdalani[1] was simple, compact, and practical. However, its accuracy was limited since it depended on the size of a perturbation parameter ε that was the reciprocal of the kinetic Reynolds number. Results were accurate as long as ε remained small. The advantages of removing this limitation were therefore obvious.

Motivated by the desire to seek a more general formulation, an exact analytical solution will be derived here that is independent of ε. Not only will the exact solution be applicable to a wider range of physical parameters, but it will also provide the means to verify the former asymptotic solution in a rigorous fashion. This will be followed by applying a standard WKBJ (Wentzel, Kramers, Brillouin, and Jeffreys) approach to arrive at a leading order, uniformly valid approximation. The WKBJ fomulation will be shown to coincide with the leading order term of the former multiple-scale solution. This will justify the unusual, nonlinear scaling transformation used formerly.[1] Due to its accuracy over a range of flow parameters, the compact multiple-scale formulation will be shown to provide a practical equivalent to the exact solution. Moreover, it will clearly exhibit the functional dependence on physical parameters.

II. MATHEMATICAL ANALYSIS

The mathematical model pertains to the acoustic velocity and pressure fields described by Majdalani.[1] For the sake of consistency, the same geometry and notation will be used here.

A. Velocity field

As detailed in the preceding article,[1] we consider a perfect gas performing small oscillations (at a circular frequency ω_0) about a steady two-dimensional field. This field is established inside a long and narrow cavity of length L, height $2H$, and width W. In a Cartesian reference frame anchored at the cavity's head-end center, y and z represent the cross-flow and streamwise coordinates. All spatial coordinates are normalized by H. For pressure oscillations of amplitude A_p and mean pressure p_0 at the cavity's head end, the total velocity profile can be expressed as a sum of steady and temporal components. The temporal component can be expressed, in turn, as a sum of acoustic, pressure-driven, and solenoidal, vorticity-driven modes. As shown in the preceding article,[1] the total dimensional velocity may be represented by

$$\mathbf{u}^*(y,z,t)=\overbrace{V_b\mathbf{U}(y,z)}^{\text{mean}}+\overbrace{a_0[\hat{\mathbf{u}}(y,z,t)+\tilde{\mathbf{u}}(y,z,t)]}^{\text{unsteady}}, \tag{1}$$

where V_b is the gas velocity at the transpiring wall, a_0 is the speed of sound, and t is dimensionless time. In fact, time $t(=a_0t^*/H)$ is made dimensionless by referring the actual time t^* to the average time it takes for an acoustic pressure disturbance to travel from the porous wall to the centerline, (H/a_0). Note that $\mathbf{u}^*$ is composed of a mean and two time-dependent parts. The dimensionless mean-flow velocity $\mathbf{U}$ is given by

$$\mathbf{U}=U_y\mathbf{e}_y+U_z\mathbf{e}_z=-y\mathbf{e}_y+z\mathbf{e}_z, \tag{2}$$

where $y=y^*/H$ is the dimensionless normal distance measured from the porous wall. Furthermore, $z=z^*/H$ is defined to be the (dimensionless) axial distance measured from the cavity's head-end wall. The dimensionless unit vectors $(\mathbf{e}_y,\mathbf{e}_z)$ have their usual significance. As such, the mean velocity component in Eq. (1) is the product of V_b and the nondimensional spatial vector $\mathbf{U}$.

Having fully defined the steady field inside the cavity, the acoustic velocity that appears in the unsteady part of Eq. (1) can be normalized by the speed of sound a_0 and written as[1]

$$\hat{\mathbf{u}}(z,t)=i\,(\varepsilon_w/\gamma)\,\sin(k_m z)\exp(-ik_m t)\mathbf{e}_z+O(M_b). \quad (3)$$

Here $\varepsilon_w=A_p/p_0$ is the pressure wave amplitude, γ is the ratio of specific heats, $k_m=m\pi H/L$ is the dimensionless wave number, and $m=1,2,3,...$ is the acoustic mode number. Note that these parameters are all unitless.

The second unsteady part of Eq. (1) represents the vortical velocity response. This component stems from the linearized, rotational momentum equation known to the order of the cross-flow Mach number $(M_b=V_b/a_0\ll 1)$. This can be expressed by

$$\tilde{\mathbf{u}}(y,z,t)=\mathbf{V}(y,z)\exp(-ik_m t),\quad \mathbf{V}(y,z)=V_y\mathbf{e}_y+V_z\mathbf{e}_z. \quad (4)$$

Noting that $V_y/V_z=O(M_b)$, the corresponding vortical mass and momentum conservation equations are

$$\partial V_y/\partial y+\partial V_z/\partial z=0, \quad (5)$$

$$iV_z=\sigma\left[\frac{\partial}{\partial z}(V_zU_z)+U_y\frac{\partial V_z}{\partial y}\right]-\varepsilon\frac{\partial^2 V_z}{\partial y^2}+O(M_b), \quad (6)$$

where

$$\varepsilon\equiv 1/\mathrm{Re}_k=\nu_0/\omega_0 H^2,\quad \sigma\equiv 1/\mathrm{Sr}=V_b/\omega_0 H. \quad (7)$$

The last two parameters represent the reciprocals of the kinetic Reynolds number, Re_k, and the Strouhal number, Sr. Since, in practice, $\mathrm{Re}_k>10^3$ and $\mathrm{Sr}>10$, both ε and σ can be used as primary and secondary perturbation parameters. Note that the Strouhal number is the product of the circular frequency and characteristic height $(\omega_0 H)$ divided by the velocity at the boundary V_b. Since V_b is two-to-three orders of magnitude smaller than a_0, Sr is much larger than the familiar aeroacoustic Strouhal number based on the speed of sound. The latter extends over the range $[10^{-3},10]$ with a peak in the noise spectrum occurring at $\mathrm{Sr}\approx 0.2$. In the current analysis, Sr extends over the range $[10,10^3]$ with a commonly reported value of $\mathrm{Sr}\approx 50$.

When separation of variables is applied to the momentum equation (6), the vortical velocity component can be solved for. The result is

$$V_z(y,z)=-\frac{\varepsilon_w}{\gamma}i\sum_{n=0}^{\infty}\frac{(-1)^n(k_m z)^{2n+1}}{(2n+1)!}Y_n(y). \quad (8)$$

Here the velocity eigenfunction $Y_n(y)$ must be determined from the boundary-value problem defined by

$$\varepsilon\frac{d^2Y_n}{dy^2}+\sigma y\frac{dY_n}{dy}+[i-(\lambda_n+1)\sigma]Y_n=0,\quad \lambda_n=2n+1, \quad (9)$$

and

$$Y_n(1)=1\ \ \text{(no-slip)}, \quad (10)$$

$$dY_n(0)/dy=0\ \ \text{(centerline symmetry)}. \quad (11)$$

B. Unconventional multiple-scale solution

Using two fictitious scales, $y_0=y$ and $y_1=\varepsilon y^{-2}$, a multiple-scale solution for Y_n was derived before by Majdalani.[1] Using the superscript M for ''multiple scale,'' this solution is reproduced here for convenience and added clarity

$$Y_n^M(y)=y^{2n+2}\exp\{-\xi[1-\sigma^2(2n+1)(2n+2)]\eta - i[\ln y+\xi\sigma^2(4n+3)\eta]/\sigma\}+O(\varepsilon), \quad (12)$$

where $\xi=\mathrm{Sr}^3/\mathrm{Re}_k$ is a viscosity parameter. Note that the asymptotic solution strongly depends on

$$\eta(y)=\int_1^y U_y^{-3}(\tau)\,d\tau=(y^{-2}-1)/2. \quad (13)$$

Physically, η represents the characteristic length scale normal to the porous wall. It thus controls the rate of decay of the rotational wave amplitude as the distance from the porous wall is increased. For example, near the porous wall, $\eta(1)=O(1)$, and so is the rotational wavelength. However, as the core is approached, $\eta(0)\to\infty$, and the rotational wavelength becomes infinitesimally small.

At this point, one may substitute Eq. (12) back into Eqs. (8) and (4). At the outset, one obtains

$$\tilde{u}_z^M(y,z,t)=-(\varepsilon_w/\gamma)\,iy\sin(k_m yz)\exp[-(1-2\sigma^2)\xi\eta - i(\ln y+3\xi\sigma^2\eta)/\sigma-ik_m t]+O(\varepsilon). \quad (14)$$

From continuity, the normal component $\tilde{u}_y$ can be determined. Thus using $\partial\tilde{u}_y/\partial y=-\partial\tilde{u}_z/\partial z$, one gets

$$\tilde{u}_y^M(y,z,t)=-(\varepsilon_w/\gamma)\,M_b y^3\cos(k_m yz)\exp[-(1-2\sigma^2)\xi\eta - i(\ln y+3\xi\sigma^2\eta)/\sigma-ik_m t]+O(\varepsilon), \quad (15)$$

where $\tilde{u}_y/\tilde{u}_z=O(M_b)$ is confirmed. Superimposing acoustic and vortical fields gives the total oscillatory velocity. Writing at $O(M_b)$ renders, at length,

$$u_z^M(y,z,t)=(\varepsilon_w/\gamma)\,i\exp(-ik_m t)\{\sin(k_m z)-y\sin(k_m yz) \times\exp[-(1-2\sigma^2)\xi\eta-i(\ln y+3\xi\sigma^2\eta)/\sigma]\}, \quad (16)$$

whose real part is

$$u_z^M(y,z,t)=\frac{\varepsilon_w}{\gamma}\left[\overbrace{\sin(k_m z)\sin(k_m t)}^{\text{acoustic part}} - \overbrace{\underbrace{y\sin(k_m yz)\exp(-\zeta)}_{\text{vortical amplitude}}\underbrace{\sin(k_m t+\Phi)}_{\text{wave propagation}}}^{\text{vortical part}}\right], \quad (17)$$

where

$$\zeta=\xi(1-2\sigma^2)\,\eta, \quad \Phi=(\ln y+3\xi\sigma^2\eta)/\sigma. \tag{18}$$

C. Exact solution

The separated momentum equation (9) can be expressed as:

$$R^{-1}Y_n''+yY_n'+[i\,\mathrm{Sr}-(2n+2)]Y_n=0; \qquad R\equiv \mathrm{Re}_k/\mathrm{Sr}=V_bH/\nu_0. \tag{19}$$

In addition to Sr, Eq. (19) appears to be controlled by the cross-flow Reynolds number R.

1. Liouville–Green transformation

An exact solution to Eq. (9) seems possible if a transformation can be implemented in a manner to eliminate all variable coefficients from the differential equation. To that end, a Liouville–Green transformation may be attempted (cf. Nayfeh,[2] pp. 364–366). This requires setting

$$X=\phi(y), \quad F(X)=\psi(y)Y_n(y), \tag{20}$$

where $\phi(y)$ and $\psi(y)$ must be determined in a manner to transform Eq. (19) into a simpler equation in $F(X)$ that exhibits an exact solution. Starting with Eq. (20), derivatives become

$$Y_n'=-\frac{\psi'}{\psi^2}F+\frac{1}{\psi}\frac{dF}{dX}\frac{dX}{dy}=-\frac{\psi'}{\psi^2}F+\frac{\phi'}{\psi}\frac{dF}{dX}, \tag{21}$$

$$Y_n''=\frac{\phi'^2}{\psi}\frac{d^2F}{dX^2}+\left(\frac{\phi''}{\psi}-\frac{2\phi'\psi'}{\psi^2}\right)\frac{dF}{dX}-\left(\frac{\psi''}{\psi^2}-\frac{2\psi'^2}{\psi^3}\right)F. \tag{22}$$

Substitution into Eq. (19) gives

$$\frac{d^2F}{dX^2}+\frac{1}{\phi'^2}\left(\phi''-\frac{2\phi'\psi'}{\psi}+yR\phi'\right)\frac{dF}{dX} +\frac{1}{\phi'^2}\left\{\left(-\frac{\psi''}{\psi}+\frac{2\psi'^2}{\psi^2}-\frac{yR\psi'}{\psi}\right) +R[i\,\mathrm{Sr}-(2n+2)]\right\}F=0. \tag{23}$$

At this juncture, functions ϕ and ψ must be chosen so that dominant parts of the transformed equation have constant coefficients. For example, the coefficient of dF/dX can be forced to be zero by setting

$$\phi''-\frac{2\phi'\psi'}{\psi}+yR\phi'=0, \text{ or } \frac{\psi'}{\psi}=\frac{\phi''}{2\phi'}+\frac{yR}{2}. \tag{24}$$

This algebraic maneuver leads to an expression for ψ that transforms Eq. (23) into an equation that can be solved exactly. As a matter of fact, direct integration of Eq. (24) gives $\psi=K_0\sqrt{\phi'}\exp(Ry^2/4)$, where K_0 is an arbitrary constant. This reduces Eq. (23) into

$$d^2F/dX^2+\{(R/\phi'^2)[i\,\mathrm{Sr}-(2n+2)]+\delta\}F=0, \qquad \delta=(1/\phi'^2)(-\psi''/\psi+2\psi'^2/\psi^2-yR\psi'/\psi). \tag{25}$$

The first derivative is hence eliminated. Next, one imposes

$$(R/\phi'^2)[i\,\mathrm{Sr}-(2n+2)]=\mathrm{Const}, \tag{26}$$

so that $\phi'=\sqrt{R}$. As one sets $K_0=R^{-1/4}$, $X=\phi(y)=y\sqrt{R}$, and $\psi(y)=\exp(Ry^2/4)$. The transformed equation becomes

$$\frac{d^2F}{dX^2}+(p+\tfrac{1}{2}-\tfrac{1}{4}X^2)F=0, \quad p\equiv -3-2n+i\,\mathrm{Sr}. \tag{27}$$

Likewise, the two physical boundary conditions given by Eqs. (10) and (11) translate into

$$F(\sqrt{R})=\exp(R/4), \quad dF(0)/dX=0. \tag{28}$$

2. Exact solution

Equation (27) possesses a standard solution that is expressible in terms of the parabolic cylinder function $D_p(X)$:

$$F(X)=C_1D_p(X)+C_2D_p(-X). \tag{29}$$

Since $\mathrm{Re}(p)<0$, formula 9.241.2 in Gradshtein and Ryzhik[3] (cf. p. 1092) can be used for $D_p(X)$. Accordingly,

$$D_p(X)=[\Gamma(-p)]^{-1}\exp(-\tfrac{1}{4}X^2)\times\int_0^\infty \tau^{-p-1}\exp(-\tau X-\tfrac{1}{2}\tau^2)\,d\tau, \tag{30}$$

where Γ is Euler's Integral of the second kind. Careful application of boundary conditions gives, after some effort,

$$F'(0)=-2^{-1/2(1+p)}\Gamma[\tfrac{1}{2}(1-p)](C_1-C_2)/\Gamma(-p)=0, \tag{31}$$

$$C_1=C_2=2^{p/2}\exp(\tfrac{1}{2}R)\Gamma(-p)/[\Gamma(-\tfrac{1}{2}p)\Phi(-\tfrac{1}{2}p,\tfrac{1}{2},\tfrac{1}{2}R)], \tag{32}$$

where Φ is the confluent hypergeometric function given by

$$\Phi(a,b;x)=1+\frac{a}{b}\frac{x}{1!}+\frac{a(a+1)}{b(b+1)}\frac{x^2}{2!}+\frac{a(a+1)(a+2)}{b(b+1)(b+2)}\frac{x^3}{3!}+\cdots. \tag{33}$$

Using the superscript E for ''exact,'' we write

$$Y_n^E(y)=\exp[\tfrac{1}{2}R(1-y^2)]\Phi(-\tfrac{1}{2}p,\tfrac{1}{2},\tfrac{1}{2}Ry^2)/\Phi(-\tfrac{1}{2}p,\tfrac{1}{2},\tfrac{1}{2}R). \tag{34}$$

The remaining equations follow precisely Eqs. (14)–(16) via Eq. (8). For example,

$$\tilde{u}_z^E(y,z)=-\frac{\varepsilon_w}{\gamma}i\exp(Ry^2\eta-ik_mt)\times\sum_{n=0}^{\infty}\frac{(-1)^n(k_mz)^{2n+1}\Phi(-\frac{1}{2}p,\frac{1}{2},\frac{1}{2}Ry^2)}{(2n+1)!\,\Phi(-\frac{1}{2}p,\frac{1}{2},\frac{1}{2}R)}, \tag{35}$$

and

TABLE I. Exact and asymptotic predictions for Sr=50, $\mathrm{Re}_k=10^6$, and $n=0$.

y	Y_n^E Eq. (34) Exact sol.	Y_n^M Eq. (12) Multi-scale	Y_n^W Eq. (40) WKBJ sol.	$\left\|\frac{Y_n^E-Y_n^M}{Y_n^E}\right\|$ %	$\left\|\frac{Y_n^E-Y_n^W}{Y_n^E}\right\|$ %
0.25	0.024 493 5	0.024 242 2	0.023 989 0	1.0300	2.0600
0.30	−0.043 164 5	−0.042 664 3	−0.041 790 0	1.1600	3.1800
0.35	−0.045 410 8	−0.046 011 2	−0.047 679 7	1.3200	5.0000
0.40	−0.026 949 6	−0.027 597 5	−0.029 787 6	2.4000	10.500
0.45	−0.094 223 1	−0.094 654 6	−0.096 500 0	0.4580	2.4200
0.50	−0.206 560 1	−0.206 475 5	−0.206 225 2	0.0410	0.1620
0.55	0.009 566 5	0.009 970 2	0.012 231 6	4.2200	27.900
0.60	0.296 660 9	0.296 503 4	0.295 630 4	0.0531	0.3470
0.65	−0.347 997 5	−0.348 100 1	−0.348 949 3	0.0295	0.2740
0.70	0.240 239 7	0.240 423 3	0.241 935 7	0.0764	0.7060
0.75	−0.129 288 0	−0.129 459 7	−0.130 970 6	0.1330	1.3000
0.80	0.097 999 9	0.098 135 3	0.099 419 1	0.1380	1.4500
0.85	−0.188 407 6	−0.188 502 6	−0.189 477 8	0.0504	0.5680
0.90	0.420 399 6	0.420 452 5	0.421 044 2	0.0126	0.1530
0.95	−0.751 123 7	−0.751 139 1	−0.751 333 2	0.0020	0.0279
1.00	1.000 000 0	1.000 000 0	1.000 000 0	0.0000	0.0000

$$u_z^E(y,z,t)=\frac{\varepsilon_w}{\gamma} i\exp(-ik_m t)\left[\sin(k_m z)-\exp(Ry^2\eta)\times\sum_{n=0}^{\infty}\frac{(-1)^n(k_m z)^{2n+1}\Phi(-\frac{1}{2}p,\frac{1}{2},\frac{1}{2}Ry^2)}{(2n+1)!\,\Phi(-\frac{1}{2}p,\frac{1}{2},\frac{1}{2}R)}\right]. \quad (36)$$

The real part of u_z^E may now be compared to the multiple-scale solution given by Eq. (17).

D. Standard WKBJ solution

In seeking a standard asymptotic solution for Eq. (9), it should be noted that two cases must be considered depending on the order of the Stouhal number.

1. Standard outer expansion

For $\mathrm{Sr}=O(1)$, $y=O(1)$, and one must resolve, at leading order, the outer solution Y_n^o from

$$\sigma y Y_n^{o\prime}+[i-(2n+2)\sigma]Y_n^o=0,$$
$$Y_n^o(1)=1, \text{ or } Y_n^o(y)=y^{2n+2}\exp(-i\,\mathrm{Sr}\ln y). \quad (37)$$

On the one hand, the y^{2n+2} factor in Y_n^o decays rapidly as $y\to 0$. As a result, the remaining boundary condition at the origin is automatically satisfied by the first derivative. This obviates the need for an inner solution of this order. On the other hand, the exponential term in Y_n^o represents an oscillatory behavior that is rapid for $\mathrm{Sr}>10$. Since Sr can be large in practice, rapid oscillations that occur on a shorter scale preclude the possibility of a uniformly valid solution. This becomes apparent in the expression for the first order correction when the outer solution is written at $O(\varepsilon^2)$:

$$Y_n^o(y)=y^{2n+2}\exp(-i\,\mathrm{Sr}\ln y)\{1+\varepsilon\,\mathrm{Sr}[-\mathrm{Sr}^2+2n(2n+1)-i(4n+1)\mathrm{Sr}](y^{-2}-1)/2\}. \quad (38)$$

In fact, since the correction term comprises a part of $O(\varepsilon\,\mathrm{Sr}^3)$, nonuniformity can be expected for large Sr. At the outset, a WKBJ expansion may be called upon.

2. The WKBJ expression

For $\mathrm{Sr}>10$, rapid oscillations occur on a short scale, and there is a slow drift on the scale $y=O(1)$. The leading order equation that provides the WKBJ ansatz is

$$yY_n'+i\,\mathrm{Sr}\,Y_n=0, \quad Y_n(1)=1, \text{or } Y_n(y)=\exp(-i\,\mathrm{Sr}\ln y). \quad (39)$$

This suggests posing $Y_n^W(y)=g(y)\exp(-i\,\mathrm{Sr}\ln y)$ and substituting back into Eq. (9). The emerging formulation can be obtained at $O(\varepsilon\,\mathrm{Sr}^2)$:

$$Y_n^W(y)=y^{2n+2}\exp(-\xi\eta-i\,\mathrm{Sr}\ln y). \quad (40)$$

Clearly, the standard WKBJ solution Y_n^W matches the leading order term of Eq. (12). This confirms the validity of the former multiple-scale solution Y_n^M. Note that both asymptotic and exact formulations depend exponentially on the function η.

E. Comparison

For $\mathrm{Re}_k=10^6$ and $\mathrm{Sr}=50$, exact, multiple-scale, and WKBJ predictions are displayed in Table I. Included are the percentage deviations of asymptotic solutions from Y_n^E. In some entries, it is gratifying to note that Y_n^M and Y_n^W match Y_n^E in several decimal places. This agreement is typical. Naturally, the accuracy of asymptotic predictions deteriorates for smaller values of Re_k.

III. CONCLUSIONS

This work extends the preceding article[1] by presenting an exact solution for the temporal field inside a porous cavity. The exact solution serves a dual purpose. Despite its relative complexity by comparison to the asymptotic formulations, it provides accurate predictions over a far broader range of physical parameters. It also serves as a benchmark for validating other possible, approximate solutions. For example, its exponentially decaying argument is shown to depend on a characteristic length scale that also appears in the

asymptotic formulations. This spatial scale can be ascribed to the nonlinear scaling constitution arising in this problem. It therefore explains the need for an unconventional scaling transformation in the multiple-scale expansion used in the preceding article.[1] The standard WKBJ analysis introduced here is also confirmatory. In hindsight, the establishment of an exact solution for the problem at hand does not undermine the usefulness of asymptotic formulations. The latter have the advantage of being expressible in simple finite forms that clearly display the dependence on physical agents. They hence remain quite practical over a substantial range of parameters corresponding to a Strouhal number in excess of 10 and a kinetic Reynolds number in excess of 1000.

[1] J. Majdalani, ''Vortical and acoustical mode coupling inside a two-dimensional cavity with transpiring walls,'' J. Acoust. Soc. Am. **106**, 46–56 (1999).
[2] A. H. Nayfeh, *Introduction to Perturbation Techniques* (Wiley, New York, 1993).
[3] I. S. Gradshtein and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. (Academic, Boston, 1994).

Fundamental azimuthal modes of a constricted annular resonator: Theory and measurement

Ralph T. Muehleisen[a)]
Civil, Environmental, and Architectural Engineering, University of Colorado, 428 CUB, Boulder, Colorado 80309

Anthony A. Atchley
Graduate Program in Acoustics, The Pennsylvania State University, PO Box 30, State College, Pennsylvania 16804

(Received 3 February 2000; revised 1 October 2000; accepted 9 November 2000)

The fundamental azimuthal modes of a constricted annular resonator are investigated. It is found that a given mode of an unconstricted resonator splits into two separate modes in the constricted resonator. One mode is of a higher frequency and has a pressure antinode centered in the constricted region. The other mode is of a lower frequency and has a pressure node centered in the constricted region. The resonance frequency of the higher-frequency modes increases linearly with a decrease in the constricted to unconstricted area ratio, whereas the lower frequency drops nonlinearly. Measurements and theory match to within 0.5% when end corrections and thermo-viscous losses are included in the system model. It was found that end correction impedances derived by mode-matching techniques were the only ones accurate enough to match the measurements and computation to within the error bounds. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1337958]

PACS numbers: 43.20.Ks, 43.20.Mv [ANN]

I. INTRODUCTION

The annular prime mover and refrigerator are topics of current interest in thermoacoustics.[1,2] In order to begin research on annular prime movers, a thorough knowledge of the acoustics of an annular thermoacoustic system is required. A good basis for this research is the analysis of a constricted annular resonator. While research on the annular resonator is found in many areas such as laser optics,[3] plasma physics,[4] microwave engineering,[5] investigations on the acoustics of an annular resonator are not as widespread. Lawrenson *et al.*[6] analyzed an annular resonator with driven walls, but the resonator did not include constrictions. Denardo and Alkov[7] and Denardo and Bernard[8] analyzed acoustic resonators with constrictions, but the resonators were not annular. The purpose of this paper is to present the results of theoretical and experimental investigations of the fundamental modes of a constricted annular resonator. In Sec. II we provide a modal solution for the unconstricted resonator and develop an equivalent ''unwrapped'' system for the plane wave modes. In Sec. III we add a constriction into the system and develop a characteristic equation for determination of the resonance frequencies, assuming continuity of pressure and volume velocity as the boundary condition at the edges of the constriction. It is shown that a given mode will split into two modes one of which is at a slightly lower frequency and the second at a slightly higher frequency. Measurements of the mode splitting shows the simple theory predicts the higher-frequency resonance well but the lower-frequency resonance poorly. The model is then redeveloped including end corrections and thermo-viscous losses. Section IV discusses experimental validation of the corrected model. The experimental and computational methods are described and computational predictions and measured results are presented.

II. UNCONSTRICTED ANNULAR RESONATOR

Consider an rigid unconstricted annular resonator with a rectangular cross section as shown in Fig. 1. The resonator has an inner radius a, an outer radius b, and a height h. Ignoring any thermo-viscous effects, low-amplitude, time-harmonic, acoustic oscillations in the resonator are described by the linear Helmholtz equation

$$\nabla^2 p + k^2 p = 0, \tag{1}$$

where p is the acoustic pressure, and k is the acoustic wave number given by $k=\omega/c$, where ω is the frequency of oscillation, and c is the adiabatic speed of sound.

A general solution to the acoustic pressure wave equation in cylindrical coordinates can be written as

$$p=\sum_m [A_r J_m(k_r r)+B_r Y_m(k_r r)][A_z \cos(k_z z) + B_z \sin(k_z z)][A_\theta e^{-jm\theta}+B_\theta e^{-jm\theta}]e^{j\omega t}, \tag{2}$$

where m is an integer, J_m and Y_m are the Bessel and Neumann functions of order m, respectively and the radial wave number k_r and vertical wave number k_z are related by

$$k_r^2=\frac{\omega^2}{c^2}-k_z^2. \tag{3}$$

A. Rigid wall condition

In the case of a resonator with rigid walls, the acoustic velocity normal to the walls must vanish at the walls. Hence,

a)Electronic mail: ralph.muehleisen@colorado.edu

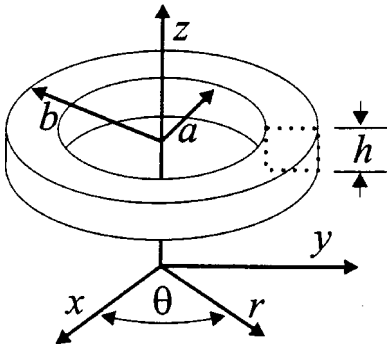


FIG. 1. Geometry of a rectangular cross-section annular resonator with an inner radius a, an outer radius b, and a height h.

the gradient of the pressure normal to the walls must also vanish. Imposing the rigid boundary conditions at the top and bottom walls requires that $B_z=0$ and $k_z=n\pi/h$, where n is any integer. Imposing the rigid boundary conditions on the inner and outer walls of the resonator requires that

$$B_r=\frac{J'_m(k_r a)}{Y'_m(k_r a)}, \tag{4}$$

and

$$J'_m(k_r a)Y'_m(k_r b)-J'_m(k_r b)Y'_m(k_r a)=0, \tag{5}$$

where $J'_m(r)$ and $Y'_m(r)$ are the radial derivatives of $J_m(r)$ and $Y_m(r)$.

The lth value of k_r which satisfies Eq. (5) is denoted k_{ml}. When these values of k_{ml} are used with Eq. (3), the characteristic (resonance) frequencies ω_{mnl} are given by

$$\omega_{mnl}=c\sqrt{(k_{ml})^2+\left(\frac{n\pi}{h}\right)^2}. \tag{6}$$

The complete solution for the pressure within the resonator is then given by

$$p=\sum_{m=-\infty}^{\infty}\sum_{n=0}^{\infty}\sum_{l=-\infty}^{\infty}\left[A_{mnl}e^{jm\theta}+B_{mnl}e^{-jm\theta}\right]\left[J_m(k_{ml}r)+\frac{J'_m(k_{ml}a)}{Y'_m(k_{ml}a)}Y_m(k_{ml}r)\right]\cos\left(\frac{n\pi z}{h}\right)e^{j\omega_{mnl}t}. \tag{7}$$

B. Plane wave equivalent system

For a resonator with an inner radius larger than its width or height $[a>(b-a)$ and $a>h]$ the lowest-frequency modes will have $n=0$ and $l=0$, meaning they are the azimuthal modes of the system. When $n=0$ there is no variation of the acoustic pressure in z and when $l=0$ there is only a small variation of the acoustic pressure in r. Thus, the $l=n=0$ modes closely resemble the plane wave modes of a straight duct and one can approximate the propagation in the annular resonator as plane wave propagation in such a duct. The resonance frequencies of the $l=n=0$ modes will be denoted ω_{m00}. In order to find an equivalent straight duct system, we begin by ''unwrapping'' the annulus as shown in Fig. 2 and determining the length of a straight duct that has the same resonance frequencies. If we consider the duct to extend a length L in the x direction, the boundary conditions at the ends of this equivalent straight duct would be periodic, i.e., the pressure and velocity are the same at $x=L$ as at $x=0$. Such a system has its mth resonance frequency ω'_m at

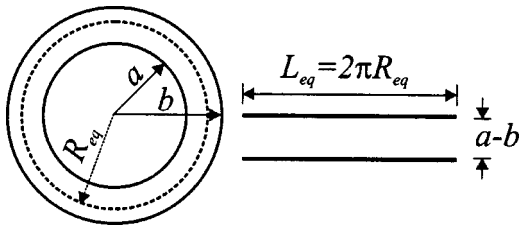


FIG. 2. Illustration of an annular system and its unwrapped equivalent straight duct. The resonator has a plane wave equivalent radius of R_{eq} and an equivalent plane wave length of L_{eq}.

$$\omega'_m=\frac{2\pi mc}{L}. \tag{8}$$

To determine the length of straight duct L_{eq} with the same resonance frequencies as the annular system, we need to have an algebraic expression for the ω_{m00} resonance frequencies of the annular system. One could numerically solve Eq. (5) to determine the eigenvalues, but if the width of the resonant chamber $(b-a)$ is small compared to the radius of the inner wall (a), a more useful analytic result can be determined. Using a Taylor expansion of Eq. (5) about the parameter $\epsilon=(b-a)/a$, Gottlieb[9] determined that to fourth order in ϵ

$$\omega_{m00}=\frac{mc}{a}\sqrt{\left[1-\epsilon+\frac{5}{6}\epsilon^2-\frac{2}{3}\epsilon^3+\frac{16-m^2}{30}\epsilon^4\right]}, \tag{9}$$

which if simplified to first order in ϵ, matches the solution presented by Lawrenson[6] and Rostafinski.[10]

Equating the fundamental resonance frequencies ω'_1 and ω_{100}, the equivalent length of straight duct L_{eq} is given, to fourth order in ϵ

$$L_{\mathrm{eq}}=2\pi a\left(1+\frac{\epsilon}{2}-\frac{\epsilon^2}{24}+\frac{\epsilon^3}{48}-\frac{9\epsilon^4}{640}\right). \tag{10}$$

In order to determine the equivalent length of a section of an annular resonator, one can define an equivalent radius R_{eq} as $L_{\mathrm{eq}}/(2\pi)$. Then, a section of the annular resonator subtending an angle θ_s would have an unwrapped length of $L=\theta_s R_{\mathrm{eq}}$.

Having determined the length of straight duct with the same resonance frequency, we can now define an equivalent straight duct system of length L for plane wave propagation in the annular system. The pressure in the equivalent system is then given by

$$p(x)=Ae^{jkx}+Be^{-jkx}, \tag{11}$$

where $0\leqslant x\leqslant L$. Since the true annular system is periodic, we must also require that $p(0)=p(L)$.

III. CONSTRICTED ANNULAR RESONATOR

The modal solutions of the wave equation within the uniform annular resonator are not unique because there is no particular position in the resonator where nodes must exist—in other words, a degeneracy exists. When a constric-

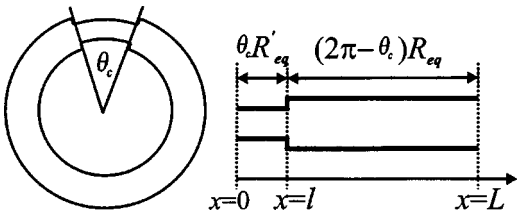


FIG. 3. Illustration of a constricted annular system and its unwrapped equivalent straight duct. The constriction subtends an angle of θ_c of the resonator.

tion is placed in the resonator the degeneracy is split into two distinct modes of different frequencies. One mode, to be denoted the ''high'' mode, has a pressure antinode at the center of the constriction and is of a slightly higher frequency than the same mode in an unconstricted resonator. The other mode, to be denoted the ''low'' mode, has a pressure node in the center of the constriction and is of a lower frequency than the equivalent mode in an unconstricted resonator.

To analyze the constricted annular system consider the system shown in Fig. 3, where a constriction with open area S_1 and angular width θ_c is placed in an annular resonator of open area S_2. The unwrapped unconstricted part of the resonator would have an equivalent length of $L_{\mathrm{eq}}=(2\pi-\theta_c)R_{\mathrm{eq}}$, where R_{eq} is the equivalent radius of the system without the constriction. The constricted part of the resonator has an equivalent length of $l=\theta_c R'_{\mathrm{eq}}$ where R'_{eq} is computed from the inner and outer radius of the constricted area. The system is then unwrapped such that $x=0$ corresponds to one end of the constriction, $x=l=\theta_c R'_{\mathrm{eq}}$ corresponds to the other end, and $x=L=l+(2\pi-\theta_c)R_{\mathrm{eq}}$ corresponds to the end of the unconstricted region. For a first approximation, we will use continuity of pressure and volume velocity as the boundary conditions at the interfaces of the constricted and unconstricted regions.

Looking at the equivalent plane wave system as shown in Fig. 3, the duct is naturally divided into two regions: region 1 is a constriction of length l and region 2 is an unconstricted region of length $L-l$. Suppressing the $e^{j\omega t}$ time dependence and ignoring any thermo-viscous losses, the pressure and volume velocity in both regions can be written in the form

$$p_1=A_1e^{-jkx}+B_1e^{jkx}, \tag{12}$$

$$U_1=\frac{S_1k}{\rho_m c}(A_1e^{-jkx}-B_1e^{jkx}), \tag{13}$$

$$p_2=A_2e^{-jkx}+B_2e^{jkx}, \tag{14}$$

$$U_2=\frac{S_2k}{\rho_m c}(A_2e^{-jkx}-B_2e^{jkx}), \tag{15}$$

where k is the wave number and S_1 and S_2 are the cross-sectional duct area in the constricted and unconstricted regions, respectively. The boundary conditions at $x=0$, $x=l$, $x=L$ can then be written

$$U_2(L)=U_1(0), \tag{16}$$

$$p_2(L)=p_1(0), \tag{17}$$

$$U_2(l)=U_1(l), \tag{18}$$

$$p_2(l)=p_1(l). \tag{19}$$

Solving the simultaneous equations results in the eigenvalue equation

$$\left(1+\frac{S_1}{S_2}\right)^2\cos(kL)-\left(1-\frac{S_1}{S_2}\right)^2\cos(kL-2kl)=4\frac{S_1}{S_2}. \tag{20}$$

Once the eigenvalues kL are determined, the resonance wave numbers and frequencies are known.

A. Interpretation

When a constriction is placed within the annular resonator, the edge of the constriction creates an acoustic reference point and a boundary condition which removes the modal degeneracy and creates two distinct modes. By symmetry, one mode must have a pressure node at the constriction center and the other must have a pressure antinode at the constriction center.

For the high mode with a pressure antinode at the constriction center, the reduced area causes an increase in the acoustic density oscillation in the region compared to an unconstricted system. An increase in acoustic density acts like a decrease in compliance of the constriction, giving rise to an increase in resonance frequency for that mode.

For the low mode with a pressure node at the constriction center, the local pressure changes little. However, the reduced area of the constriction increases the inertance of the constriction, which increases the velocity in the region. The enhanced inertial effects lead to a decrease in the resonance frequency for that mode.

As will be shown in Sec. IV, the predicted resonance frequencies of the high mode match measurements quite well, whereas the predicted resonance frequencies of the low mode do not, especially for the shorter constrictions. The discrepancy can be removed by including thermo-viscous losses throughout the system and including end corrections at the edges of the constriction region.

B. End effects at constrictions

When a straight duct is interrupted with a small constriction, the constriction is often modeled by a lumped inertance.[11] When the constriction is longer, the constriction is often analyzed as a duct itself with appropriate boundary conditions at the junctions between the main duct and the constriction. The boundary conditions usually used are continuity of volume velocity, which is required by conservation of mass, and the continuity of plane wave acoustic pressure.[11] However, a more careful analysis[12] shows that pressure is *not* necessarily continuous and a more appropriate boundary condition which includes an acoustic impedance of the junction, Z_J, is needed. The pressure discontinuity at the junction is proportional to the product of the acoustic volume velocity and the junction impedance.

For a constricted annular resonator, the low mode has a velocity antinode centered on the constriction and consequently will have a high acoustic velocity at the edges of the

constriction. Having a high velocity means a large pressure discontinuity will exist at the boundary and hence an accurate model needs to include the junction impedance. In contrast, the high mode has a velocity node at the constriction center and a low acoustic velocity at the constriction edges. Because of the low velocity, a much smaller pressure discontinuity will exist and a model neglecting the acoustic junction impedance can still be quite accurate.

There are several methods for estimating the plane wave acoustic impedance of the junction. Bolt[13] developed an inertance correction for circular duct orifices which can be applied by replacing the circular duct radii with the rectangular duct hydraulic radii. Miles[14] and Morse[15] both developed equivalent junction inertances for rectangular duct area changes through the use of conformal mapping. Since the conformal mapping is only derived for a two-dimensional area change, the correction must be applied twice—once for the change in area in the *y* direction and once for the change in area in the *z* direction. Kergomard[16] and Muehleisen[17] developed general junction impedances using higher-order mode-matching techniques which are applicable to three-dimensional problems and ducts of various shapes.

It was found that both the orifice inertance correction and the conformal mapping correction underestimated the discontinuity impedance for major constrictions. In order to match the computations to the measurements within the error bounds of the computations, the inertance correction developed by the mode-matching technique had to be used. In the mode-matching method, the pressure and velocity field in both the constricted and unconstricted region is rewritten as a sum of fundamental and evanescent higher-order modes. Continuity of particle velocity and pressure, including the evanescent modes, is imposed at the discontinuity, resulting in a set of coupled equations. The equations can then be solved to determine the plane wave pressure discontinuity across the discontinuity from which a junction impedance is derived.

C. Thermo-viscous losses

In addition to edge corrections, one can also investigate thermo-viscous losses in the system. We will include these losses in our theory through the use of complex wave numbers.

Because of viscosity, the tangential as well as the normal velocity of the fluid must vanish at the wall. This leads to a viscous boundary layer. Because greater viscous losses occur in regions of high velocity, viscous losses will be more important for the low mode where the velocity in the constricted region is high.

When an fluid with an acoustic wave is in contact with a solid object, the temperature oscillations accompanying pressure oscillations in the fluid will cause heat to be exchanged between the fluid and the wall, creating a thermal boundary layer. Because greater thermal losses occur in region of high pressure oscillation, thermal losses will be more important for the high mode where the pressure in the constricted region is high.

The thermal and viscous penetration depths are the key length scales for thermal and viscous losses in the system. The thermal penetration depth, roughly the distance that heat can diffuse in a time $1/\omega$, is given by $\delta_\kappa = \sqrt{2\kappa/(\omega\rho_m C_p)}$, where κ is the thermal conductivity of the gas, ρ_m is the mean density of the gas, ω is the frequency of acoustic oscillation, and C_p is the isobaric heat capacity per unit mass. The viscous penetration depth, roughly the distance over which momentum can diffuse in a time $1/\omega$, is given by $\delta_\nu = \sqrt{2\mu/(\omega\rho_m)}$, where μ is the fluid viscosity. The other important parameter is the hydraulic radius of the duct r_h, which is the ratio of the area of the duct to the perimeter of the duct.

The boundary layer form of Rott's general linear wave equation,[18] which includes thermal and viscous losses, is

$$\left[1+(\gamma-1)(1-j)\frac{\delta_\kappa}{2r_h}\right]p+\frac{c^2}{\omega^2}\left[1-(1-j)\frac{\delta_\nu}{2r_h}\right]\frac{d^2p}{dx^2}=0, \tag{21}$$

where $\gamma = C_p/C_v$ is the ratio of isobaric to isochoric specific heats.

For a straight duct, the plane wave solution to Eq. (21) can be given in the standard form of

$$p(x)=[Ae^{j\hat{k}x}+Be^{-j\hat{k}x}]e^{j\omega t}, \tag{22}$$

where the complex wave number $\hat{k}$ is given by

$$\hat{k}=\frac{\omega}{c}\sqrt{\frac{2r_h+(\gamma-1)\delta_\kappa(1+j)}{2r_h-\delta_\nu(1+j)}}\approx k_0-j\alpha, \tag{23}$$

where the propagating wave number k_0 and attenuation coefficient α are given by

$$k_0=\frac{\omega}{c}\left[1+\frac{\delta_\nu+(\gamma-1)\delta_\kappa}{4r_h}\right], \quad \alpha=\frac{\omega}{c}\left[\frac{\delta_\nu+(\gamma-1)\delta_k}{4r_h}\right]. \tag{24}$$

From Eq. (24) one can see an interesting effect of thermo-viscous losses: an increase in the propagating wave number k_0 which could also be viewed a decrease in the speed of propagation *c*.

D. Corrected equations

Adding in end corrections and thermo-viscous losses, Eqs. (12)–(15) and (16)–(19) can be rewritten. The corrected pressure and volume velocity equations in each region are

$$p_1=A_1e^{-j\hat{k}_1x}+B_1e^{j\hat{k}_1x}, \tag{25}$$

$$p_2=A_2e^{-j\hat{k}_2x}+B_2e^{j\hat{k}_2x}, \tag{26}$$

$$U_1=\frac{S_1\hat{k}_1}{\rho_m c}(A_1e^{-j\hat{k}_1x}-B_1e^{j\hat{k}_1x}), \tag{27}$$

$$U_2=\frac{S_2\hat{k}_2}{\rho_m c}(A_2e^{-j\hat{k}_2x}-B_2e^{j\hat{k}_2x}), \tag{28}$$

where $\hat{k}_1$ and $\hat{k}_2$ are the complex wave numbers and S_1 and S_2 are the sectional duct area in the constricted and unconstricted regions, respectively. Because the areas of the regions are different, the hydraulic radii in the two regions are different, leading to different thermo-viscous losses in the

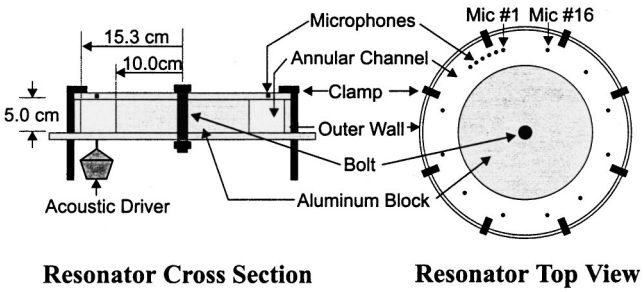


FIG. 4. Cross section and top view of the annular resonator construction used in taking measurements.

regions and the need for two separate complex wave numbers.

The corrected boundary conditions at $x=0$, $x=l$, $x=L$ are then

$$U_2(L)=U_1(0), \tag{29}$$

$$p_2(L)=p_1(0)+Z_J U_1(0), \tag{30}$$

$$U_1(l)=U_2(l), \tag{31}$$

$$p_1(l)=p_2(l)+Z_J U_1(l), \tag{32}$$

where Z_J is the acoustical impedance of the constriction junction.

The above equations can be written in a matrix form as $\tilde{M}\overline{C}=0$. A solution to the equations exists only at a distinct set of complex frequencies ω such that the determinant of the matrix $\tilde{M}$ is zero. The real part of the complex frequency is the resonance frequency of the system, whereas the imaginary part is related to the overall losses in the system.

IV. EXPERIMENTAL VALIDATION

A. Setup

To validate the theory, an experimental annular resonator was constructed. The basic design of the resonator is shown in Fig. 4. The resonator has an inner radius of 10.00 ± 0.01 cm, an outer radius of 15.30 ± 0.02 cm, and a height of 5.00 ± 0.01 cm. The dimensions of the device were chosen to give fundamental frequencies in the 300- to 500-Hz range and to ensure that the cutoff frequency of the first higher order mode was well above this range. The outer walls and bottom of the resonator are made of 1.0-cm-thick aluminum. The inner part of the resonator was solid aluminum. The resonator top was made of 1.0-cm-thick clear acrylic in order to allow for visual sighting of the constriction elements. The acrylic top was clamped to the resonator using a 1-in. nut and bolt at the center of the resonator and eight clamps symmetrically placed along the outer edges. Mounted in the acrylic top were 16 Panasonic WM-60AY electret microphone elements. Six of the microphones were located in a 30-deg region where the constriction will be located and the remaining ten microphones were mounted at 30-deg intervals around the rest of the resonator. The error in the microphone position was estimated to be less than ± 0.25 deg. The microphones were calibrated relative to one another by the use of a small chamber with a dynamic loudspeaker that was placed over the microphones. An HP 35670 dynamic signal analyzer was used to generate a swept sine signal and measure the microphone output when excited by the chamber. The output of each microphone was measured and calibration files for the response of each microphone relative the first microphone were generated. Through the use of the calibration files, the output of each microphone could be phase- and amplitude corrected relative to the first microphone in the system.

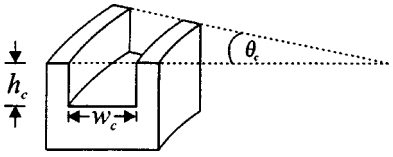


FIG. 5. Geometry of the PVC constrictions placed within the annular resonator.

In order to excite the resonator, a generic midrange compression driver was connected to a small hole in the bottom of the resonator through a thin tube. Holes were placed at positions of 45, 90, and 180 from the zero point of the system. Holes that were not being used were sealed with a small insert to maintain the rigid wall boundary condition.

The HP 35670 was also used in swept sine mode to generate a frequency response curve for the resonator. The output of the analyzer was fed through a power amplifier to the compression driver. A signal level in the range of 1–10 W was fed to the driver to ensure linear operation. The pole-zero curve fitting function of the analyzer was then used to estimate the resonance frequency. The accuracy of the resonance frequency measurement is estimated to be ± 0.2 Hz. Once the resonance frequency was determined the system was driven at the resonance frequency to measure mode shapes.

The resonator constrictions were machined from PVC plastic to a form shown in Fig. 5. The open area is at top of the constriction to allow the pressure within the constriction to be measured by the microphones. Measurements for the constrictions used are shown in Table I, where the channel width and height are denoted w_c and h_c, respectively, the channel area is denote S_c, and the area of the unconstricted region is denoted S_u. Constrictions were machined with an angular width of 22.53 ± 0.06, 11.30 ± 0.05, and 5.65 ± 0.02 deg, which correspond to straight duct constriction lengths of $l \approx L_{\mathrm{eq}}/16$, $l \approx L_{\mathrm{eq}}/32$, and $l \approx L_{\mathrm{eq}}/64$, where L_{eq} is the plane wave equivalent length of the unconstricted annulus. To reduce losses in the crevices where the top meets the rest of the resonator and where the constrictions were in-

TABLE I. Annular resonator constriction dimensions.

h_c and w_c (cm)	S_c (cm^2)	S_c/S_u
4.245	18.017	0.6812
3.630	13.177	0.4982
2.800	7.840	0.2964
2.290	5.244	0.1983
1.610	2.591	0.0980

TABLE II. Resonance frequencies of the low and high modes.

Angle (°)	Area ratio	Low mode frequencies		High mode frequencies	
		Measured (Hz)	Calculated (Hz)	Measured (Hz)	Calculated (Hz)
22.53	0.0990	306.8±0.2	306.1±1.6	459.5±0.2	459.0±0.9
22.53	0.1983	347.0±0.2	345.7±1.5	457.0±0.2	455.9±0.8
22.53	0.2964	372.4±0.2	371.4±1.3	454.0±0.2	452.9±0.8
22.53	0.4982	404.0±0.2	403.5±1.0	447.8±0.2	447.0±0.8
22.53	0.6812	420.2±0.2	419.3±1.0	442.2±0.2	441.8±0.8
22.53	1.0000	434.0±0.2	434.0±0.9	434.0±0.2	434.0±0.8
11.30	0.0990	326.5±0.2	326.4±1.6	446.6±0.2	445.8±0.9
11.30	0.1983	366.6±0.2	365.9±1.4	445.3±0.2	444.4±0.8
11.30	0.2964	389.3±0.2	388.5±1.2	443.7±0.2	442.9±0.7
11.30	0.4982	413.9±0.2	413.8±0.9	440.8±0.2	440.1±0.7
11.30	0.6812	425.7±0.2	424.7±0.8	438.1±0.2	437.6±0.7
5.65	0.0990	346.2±0.2	345.9±1.6	440.1±0.2	439.4±0.8
5.65	0.1983	382.7±0.2	381.2±1.3	439.7±0.2	438.7±0.8
5.65	0.2964	401.0±0.2	399.9±1.1	439.0±0.2	438.0±0.7
5.65	0.4982	420.5±0.2	419.9±0.8	437.6±0.2	436.6±0.7
5.65	0.6812	428.5±0.2	427.7±0.8	436.0±0.2	435.4±0.7

serted in the resonator, the surfaces in contact were sealed with vacuum grease.

The resonance frequencies were measured at different times and thus with the air and apparatus at different temperatures. The primary effect of temperature is to modify the speed of sound in the resonator. In order to better compare the measurements to predictions and create a consistent set of data, the measured resonance frequencies were adjusted to the expected values if the measurements were all taken at a 20 °C temperature. Since the speed of sound in an ideal gas is proportional to the square root of the temperature of the gas,[12] the resonance frequency was adjusted to the 293 K (20 °C) equivalent by the formula

$$f_{293\,\mathrm{K}} = f_T \sqrt{\frac{293.15}{T}}, \tag{33}$$

where f_T is the resonance frequency measured at the temperature T and T is the temperature of the resonator in kelvin.

The measurements were made in 1 bar air. Within the resonator, different modes were excited to different amplitudes, but the range of the maximum acoustic amplitudes was 0.2 to 1 Pa. The maximum acoustic amplitudes were well within the linear range of the Panasonic microphones.

B. Measurement and computational results

Resonance frequency measurements and computations, including end effects and thermo-viscous losses for each of the measurement configurations, are shown in Table II.

The computational error in the tables was determined by a Monte Carlo analysis where the inputs to the model (the resonator dimensions, the constriction position and dimensions, and the measurement temperature) were randomly varied within the error limits of the measurements. The mean value of 1000 runs was taken to be the theoretical value and the error in the computation was taken to be the maximum deviation from the mean. As can be seen in the table, the error bounds of the computation are much larger than the error in the actual measurement of the resonance frequency.

The comparison between the theoretical calculations without end corrections or thermo-viscous losses and the measurements is shown in Fig. 6. As one can see, the resonance frequencies of the high mode rise linearly as the system becomes more constricted. The resonance frequencies of the low mode decreases, but in a nonlinear fashion. One can also clearly see the simple theory was able to predict the resonance frequencies of the high mode quite well, but did not predict the frequencies of the low mode well at all. A plot of the theoretical calculations including end corrections and thermo-viscous losses compared to the measurements is shown in Fig. 7. Compared to Fig. 6, which did not include end corrections or thermo-viscous losses, one can see the significant improvement.

Table III shows the effects of thermo-viscous losses and various end corrections on the computational results on the computed resonance frequency of the low mode with a measured resonance frequency of 306.8 Hz. The table shows the

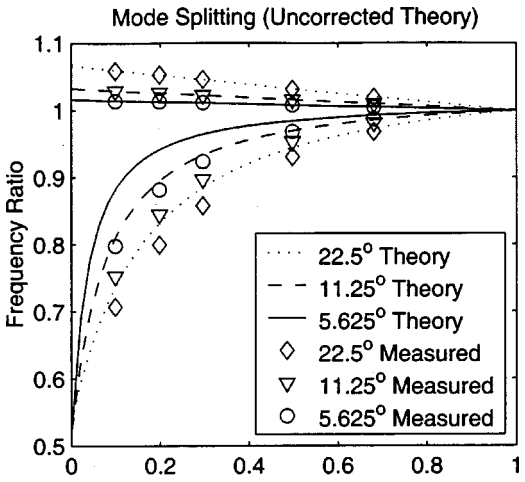


FIG. 6. Mode splitting for constrictions 22.5°, 11.25°, and 5.625° in length in an annulus where thermo-viscous losses and end corrections are ignored. The open symbols are from measurements. The ratio of constricted to unconstricted resonance frequencies is plotted as a function of the ratio of constricted area to unconstricted area.

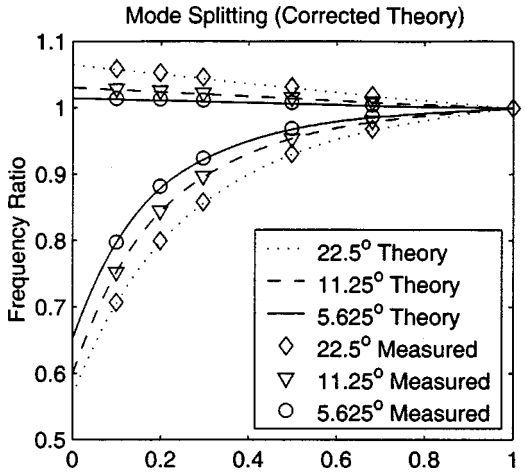


FIG. 7. Mode splitting for constrictions 22.5°, 11.25°, and 5.625° in length in an annulus when thermo-viscous losses and end corrections are included. The ratio of constricted to unconstricted resonance frequencies is plotted as a function of the ratio of constricted area to unconstricted area.

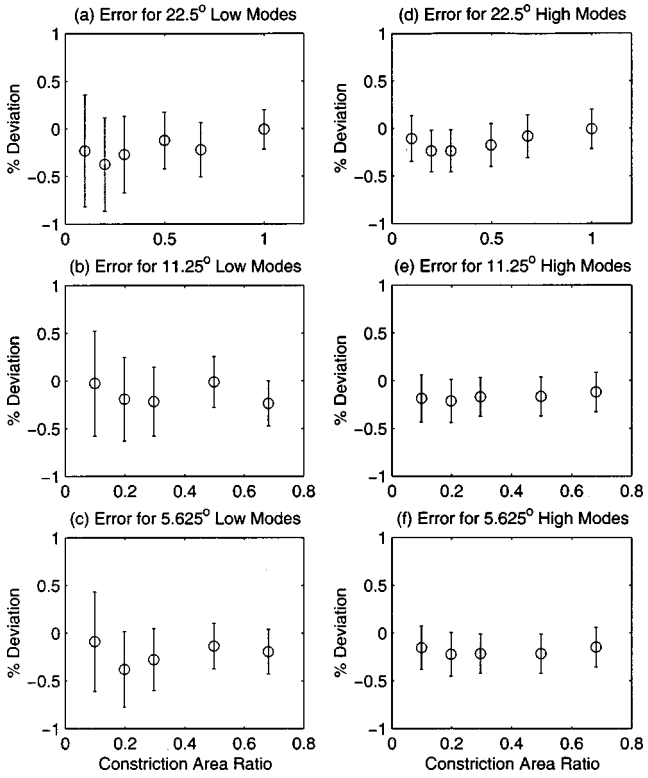


FIG. 8. Percent deviation between the calculated and measured resonance frequencies for the low mode with (a) 22.5°; (c) 11.25°; and (e) 5.625° constrictions and the high mode with (b) 22.5°; (d) 11.25°; and (f) 5.625° constrictions plotted as a function of the constriction ratio.

computed resonance frequency, and the percent deviation from the measured resonance frequency. As one can see from the table, the error in the computation using no corrections, the orifice correction, or the conformal mapping correction are much larger than when using higher-order mode-matching correction. However, you will also notice that the error without including thermo-viscous losses with the mode-matching correction is more than twice that when thermo-viscous losses were included, thus indicating the need to include the thermo-viscous losses in the model.

Because of the high acoustic velocity in the constriction for the low mode, the end correction and thermo-viscous losses are more important in computation of the resonance frequency for the low mode than for the high mode. The effect of including end corrections in the computations is much greater than including thermo-viscous losses, but in order to match the computations to the measurements within the error bounds of the computation, both the end correction and thermo-viscous losses needed to be included within the computation.

A plot of the deviation of the errors is shown in Fig. 8. In this plot the computational and measurement errors were added to generate the error bars shown in the figure. The small error bounds of the deviation show the accuracy and control under which the experiment was made. Nearly all the measurements and their error bounds are within the error bounds of the computation; thus, inclusion of crevice losses from the junction of the resonator and resonator top and the resonator and constrictions in the model was deemed unnecessary.

Figure 9 is a plot of the mode shape for the high and low modes for a 22.5°-wide constriction with a constricted area ratio of 0.1. The measurements were taken at a temperature of 22.5 °C with resonance frequencies of 308.1 and 461.5 Hz for the low and high mode, respectively. The position of the edge of the constriction at 22.5° is denoted by the dotted line and the position of the driver in the resonator at 56.25° is denoted by the dashed-dotted line. The pressure amplitude has been normalized to unity.

To compare the results of the computed mode with the measured mode, the computation was scaled so that the measured data point with the largest amplitude matched the computed curve. As can be seen in the figure, the computed and measured mode shapes match extremely well. The only places with a noticeable deviation are for the low mode at a position just past the end of the constriction and near where the acoustic driver was placed. In both cases evanescent

TABLE III. Effects of corrections on resonance frequency of a low mode with $f_0 = 306.8$ Hz.

Correction	Computed frequency (Hz)	Deviation (%)
Mode-matching impedance with thermo-viscous losses	306.1	−0.23
Mode-matching impedance without thermo-viscous losses	308.1	+0.42
Conformal mapping impedance with thermo-viscous losses	312.3	+1.8
Orifice impedance with thermo-viscous losses	314.7	+2.6
No correction	317.7	+3.5

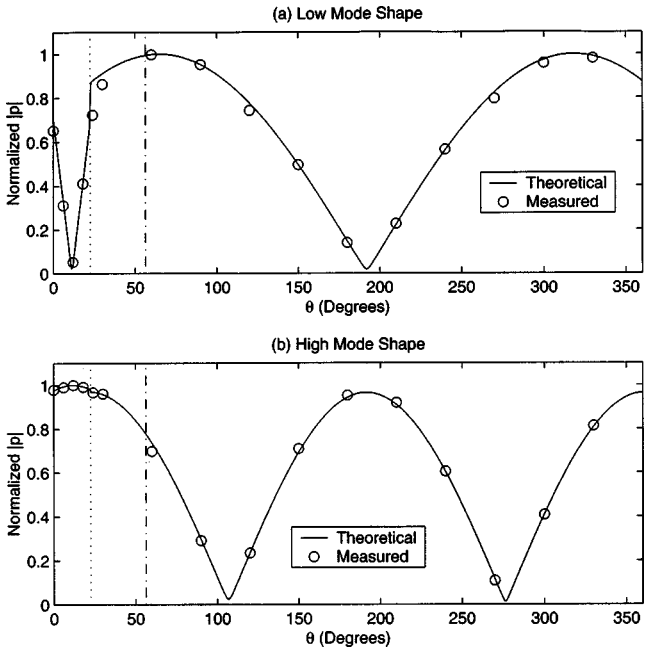


FIG. 9. Mode shapes for the (a) low mode and (b) high mode of an annulus with a 22.5° constriction with a 0.099 area ratio. The edge of the constriction at 22.5° is marked by the dotted line. The driver at 56.25° in the resonator is marked by the dashed-dotted line.

modes contribute to the measured pressure but are not included in the computation (except through the plane wave end correction). Because of the scaling involved, error bounds on the measurement were not estimated. These two plots are typical of the results obtained for plots of other mode shapes as well.

The pressure node of the low mode and pressure antinode of the high mode in the constriction center is clearly evident. As one can see, the constriction modifies the shape of high mode very little, whereas constriction affects the shape of the low mode a large amount.

V. CONCLUSIONS

We have investigated the low-frequency modes of constricted annular resonators. After showing how an annular system may be unwrapped into an equivalent straight duct, the equivalent constricted straight system is analyzed. A given mode in the annular system is shown to split into two separate modes in the constricted system: a higher frequency mode with a pressure antinode in the constriction center and a lower frequency mode with a pressure node in the constriction center. Ignoring thermo-viscous losses and end effects at the constriction edges, a characteristic eigenvalue equation can be developed from which the resonance frequencies can be determined. It is shown that the equation accurately predicts the higher-frequency mode resonance frequencies but not the lower-frequency mode resonance frequencies.

The system was reanalyzed including thermo-viscous losses and end effects and is shown to predict the resonance frequencies to within 0.5%. However, in order to properly match the data, the end correction impedances had to be derived from higher-order mode-matching techniques. End corrections derived from conformal mapping and circular duct orifice corrections were not accurate enough to match computations and measurements within the error bounds of the measurement and computation.

Measurements for a number of constriction sizes are compared to computations and the two are shown to match extremely well. It is shown that the resonance frequencies of the higher frequency mode rise linearly as the ratio of constricted to unconstricted area ratio decreases, whereas the frequencies of the lower-frequency mode drop nonlinearly. The deviation between the computations and the measurements show that the annular system can be unwrapped but that proper prediction requires the use of thermo-viscous losses and end effects at the constriction edges.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Y. S. Choe in taking the measurements and of H. T. Lin and J. D. Maynard in development of the computational models. The authors also thank the Office of Naval Research and the American Society for Engineering Education for their financial support.

[1] H. T. Lin, Ph.D. thesis, Naval Postgraduate School, 1997.
[2] G. W. Swift, ''Thermoacoustic engines,'' J. Acoust. Soc. Am. **84**, 1145–1180 (1988).
[3] K.-Y. Ng and R. Warnock, ''Reactive impedance of a smooth toroidal chamber below the resonance region,'' Phys. Rev. D **40**, 231 (1989).
[4] S. Takamura and T. Mori, ''Wavefield structures in a toroidal plasma at the frequency range of Alfvén resonance,'' Jpn. J. Appl. Phys., Part 1 **28**, 116 (1989).
[5] M. Y. Yegorov and M. A. Below, ''Properties of a toroidal dielectric resonator,'' Telecommun. Radio Eng. (Engl. Transl.) **44**, 87–90 (1989).
[6] C. C. Lawrenson, L. D. Lafleur, and F. D. Shields, ''The solution for the propagation of sound in a toroidal waveguide with driven walls (the acoustitron),'' J. Acoust. Soc. Am. **103**, 1253–1260 (1998).
[7] B. Denardo and S. Alkov, ''Acoustic resonators with variable nonuniformity,'' Am. J. Phys. **62**, 315–321 (1994).
[8] B. Denardo and M. Bernard, ''Design and measurements of variably nonuniform acoustic resonators,'' Am. J. Phys. **64**, 745–751 (1996).
[9] H. P. W. Gottlieb, ''On the lowest radial frequencies of a rigid-walled narrow toroidal duct,'' J. Acoust. Soc. Am. **105**, 2036–2038 (1999).
[10] W. Rostafinski, ''On the propagation of waves of acoustic frequencies in curved ducts,'' J. Acoust. Soc. Am. **52**, 11–15 (1972).
[11] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982).
[12] A. D. Pierce, *Acoustics—An Introduction into Its Physical Principles and Applications*, 1989 ed. (Acoustical Society of America, Woodbury, NY, 1989).
[13] R. H. Bolt, S. Labate, and K. U. Ingard, ''Acoustic reactance of small circular orifices,'' J. Acoust. Soc. Am. **21**, 94 (1949).
[14] J. W. Miles, ''The analysis of plane discontinuities in cylindrical tubes. II,'' J. Acoust. Soc. Am. **17**, 272–284 (1946).
[15] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1968).
[16] J. Kergomard, A. Garcia, G. Tagui, and J. P. Dalmont, ''The analogous acoustical impedance for discontinuities and constrictions of circular cross section,'' J. Sound Vib. **129**, 457–475 (1989).
[17] R. T. Muehleisen and A. A. Atchley, ''Lumped impedance of a planar discontinuity in an acoustic waveguide,'' in *Proceedings of the 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America*, edited by P. K. Kuhl and L. E. Crum (Acoustical Society of America, Woodbury, NY, 1998), pp. 2831–2832.
[18] N. Rott, ''Damped and thermally driven acoustic oscillations in wide and narrow tubes,'' Z. Angew. Math. Phys. **20**, 230 (1969).

Second-harmonic generation in sound beams reflected from, and transmitted through, immersed elastic solids

B. J. Landsberger[a] and M. F. Hamilton
Department of Mechanical Engineering, The University of Texas at Austin, Austin, Texas 78712-1063

(Received 10 February 2000; accepted for publication 22 August 2000)

Second-harmonic generation in sound beams reflected from and transmitted through thick isotropic elastic solids is investigated experimentally. Measurements of diffraction patterns are compared with a theoretical model based on integral solutions for harmonic generation in sound beams. The solutions are connected by classical linear theory for reflection and transmission at fluid–solid interfaces. Nonspecular phenomena associated with rapid phase variations near critical angles are accurately described. The principal restriction is that the solid is sufficiently thick that internal reflections may be ignored. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1318777]

PACS numbers: 43.25.Dc, 43.25.Jh, 43.35.Cg, 43.35.Zc [DEC]

I. INTRODUCTION

Ultrasound is widely used in nondestructive evaluation to determine the elastic properties of materials. Such techniques are normally based on linear acoustical theory for transmission through and reflection from the materials. For isotropic solids they permit measurement of the two second-order elastic moduli, e.g., Young's modulus and Poisson's ratio. At higher wave amplitudes, where finite-amplitude effects such as second-harmonic generation in the material may be detected, information is provided about the third-order elastic moduli. The purpose of the present article is to construct an accurate theoretical model for second-harmonic generation associated with a directional sound beam reflected from or transmitted through a thick isotropic elastic solid.

The model is based on angular spectrum theory. Decomposition of a sound beam incident on an elastic solid in terms of its angular spectrum permits inclusion of plane-wave reflection and transmission coefficients for the fluid–solid interfaces. This approach was introduced by Schoch[1] and used subsequently by Brekhovskikh[2] in asymptotic analyses describing nonspecular effects associated with reflection of a small-signal sound beam incident on a solid near the Rayleigh angle. Ngoc and Mayer[3] avoided the approximations employed by Schoch and Brekhovskikh by using numerical techniques to evaluate the Fourier integral over the angular spectrum of the reflected beam. Numerical evaluation of the integral was subsequently employed by many authors to model both reflection and transmission phenomena associated with small-signal sound beams incident on plates. A number of related investigations is discussed by Chimenti *et al.*[4]

Angular spectrum methodology was introduced in nonlinear acoustics by Alais and Hennion[5] to describe sum- and difference-frequency generation by a parametric array in an unbounded fluid. Integral representations of the secondary field obtained in this way, both with and without use of the parabolic approximation, have been investigated in detail by the Tjøttas.[6,7] The model presented below combines the angular spectrum formulations developed for harmonic generation in fluids with those developed for small-signal reflection from and transmission through an isotropic solid. The principal assumptions are that the solid is sufficiently thick that multiple internal reflections may be ignored, and that the incident sound field in the fluid is a directional beam.

The paper is organized as follows. In Sec. II, small-signal relations are used to determine the reflection and transmission coefficients, formulate the source function, and incorporate effects of absorption in order to model the primary beams. In Sec. III, integral expressions in terms of the angular spectra are developed for second-harmonic generation in the fluid and the solid. These solutions are used to obtain expressions for the second harmonic in both the reflected and transmitted fields. In Sec. IV, measurements are compared with the theory and demonstrate the accuracy of the approach. The main results are contained in the dissertation by one of the authors.[8]

II. LINEAR THEORY

The configuration under consideration is shown in Fig. 1. A homogeneous, isotropic, elastic solid is immersed in a homogeneous fluid. The solid is a plate of thickness h that is sufficiently large that multiple reflections within the solid may be ignored for the purpose of studying reflection and transmission of tone bursts. The sound beam is radiated at angular frequency ω by a source a distance d_i away from the solid. There is no restriction on the angle formed by the axis of the incident sound beam and plane of the fluid–solid interface. Propagation of the beam in both the fluid and the solid is accompanied by second-harmonic generation. We wish to predict the sound pressures at frequencies ω and 2ω throughout the fluid on both sides of the solid.

The present section outlines the linear theory and introduces the notation used to describe the wave fields in the fluid and the solid at the source frequency ω. Reflection and transmission coefficients for the fluid–solid interfaces are presented, diffraction is taken into account using angular spectrum theory, and absorption is included *ad hoc*. The lin-

[a]Present address: Caterpillar, Inc., Technical Center, Bldg. G, P.O. Box 1875, Peoria, IL 61656-1875.

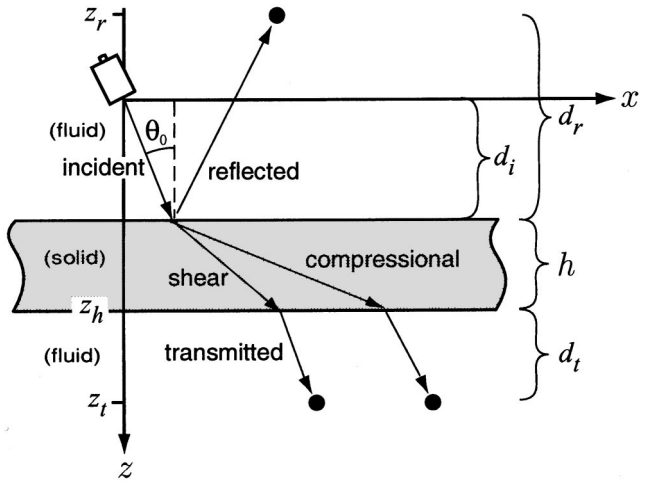


FIG. 1. Coordinates defining positions of the source and receivers in the fluid with respect to the solid, and ray paths for the longitudinal (compressional) and shear (transverse) waves in the solid.

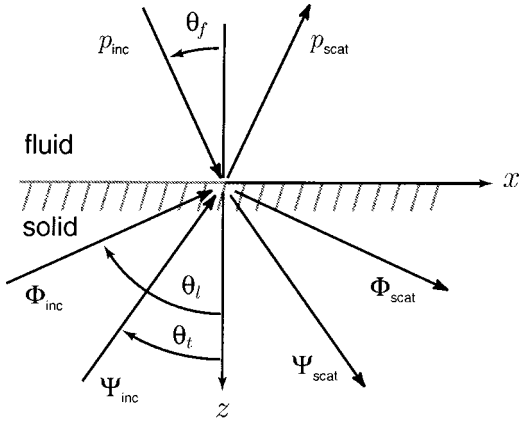


FIG. 2. Wave potentials used in the scattering matrix for the reflection and transmission coefficients.

ear theory is required in Sec. III to develop and connect the quasilinear solutions for the fields at 2ω.

A. Plane-wave reflection and transmission

We begin by considering small-signal propagation in lossless media. The sound pressure p in the fluid satisfies the wave equation

$$\nabla^2 p = \frac{1}{c_f^2}\frac{\partial^2 p}{\partial t^2}, \tag{1}$$

where $c_f=(B_f/\rho_f)^{1/2}$ is the sound speed, B_f the adiabatic bulk modulus, and ρ_f the ambient density. The elastic wave field in the solid is described by the particle displacement vector $\mathbf{u}$, the Helmholtz decomposition of which is

$$\mathbf{u}=\nabla\phi+\nabla\times\boldsymbol{\psi}. \tag{2}$$

The potential functions satisfy the wave equations

$$\nabla^2\phi=\frac{1}{c_l^2}\frac{\partial^2\phi}{\partial t^2}, \quad \nabla^2\boldsymbol{\psi}=\frac{1}{c_t^2}\frac{\partial^2\boldsymbol{\psi}}{\partial t^2}, \tag{3}$$

where $c_l=[(B_s+\frac{4}{3}\mu_s)/\rho_s]^{1/2}$ is the longitudinal (compressional) wave speed, $c_t=(\mu_s/\rho_s)^{1/2}$ the transverse (shear) wave speed, B_s the bulk modulus, μ_s the shear modulus, and ρ_s the density. The sound pressure is related to the displacement potential $\tilde{\phi}$ in the fluid by

$$p=-\rho_f\frac{\partial^2\tilde{\phi}}{\partial t^2}. \tag{4}$$

We introduce here the conventions and notation for the wave numbers and the reflection and transmission coefficients. Consider the plane-wave system sketched in Fig. 2, where the fluid occupies the upper half-space, the solid occupies the lower half-space, and all possible incident, transmitted, and reflected waves are shown. Ideal conditions are assumed at the interface, which is taken to be traction-free. The coordinate system has been adjusted such that the propagation directions of the incident waves and the resulting scattered (i.e., reflected and transmitted) waves are in the xz plane. We consider waves with angular frequency ω and express the plane-wave solutions in the form

$$p=(p_{\text{inc}}e^{iK_f z}+p_{\text{scat}}e^{-iK_f z})e^{i(\kappa x-\omega t)}, \tag{5}$$

$$\phi=(\phi_{\text{inc}}e^{-iK_l z}+\phi_{\text{scat}}e^{iK_l z})e^{i(\kappa x-\omega t)}, \tag{6}$$

$$\psi=(\psi_{\text{inc}}e^{-iK_t z}+\psi_{\text{scat}}e^{iK_t z})e^{i(\kappa x-\omega t)}, \tag{7}$$

where ψ is the y component of $\boldsymbol{\psi}$ (i.e., vertical polarization is assumed), and

$$K_{f,l,t}=\sqrt{k_{f,l,t}^2-\kappa^2}, \quad \kappa=k_{f,l,t}\sin\theta_{f,l,t}. \tag{8}$$

The signs are chosen such that $\mathrm{Re}\,K\geqslant 0$ and $\mathrm{Im}\,K\geqslant 0$. The relation for κ follows from the kinematic constraint that all field components must exhibit the same periodic spatial variation along the interface. Specifically, κ is the wave number in the plane of the interface, in this case the x component of each wave vector, with $k_{f,l,t}=\omega/c_{f,l,t}$ and with all corresponding angles defined as in Fig. 2.

The scattered wave amplitudes are related to the incident wave amplitudes by

$$\begin{pmatrix}\tilde{\phi}_{\text{scat}}\\ \phi_{\text{scat}}\\ \psi_{\text{scat}}\end{pmatrix}=\begin{pmatrix}R_{ll} & \tilde{T}_{ll} & \tilde{T}_{tl}\\ T_{ll} & \tilde{R}_{ll} & \tilde{R}_{tl}\\ T_{lt} & \tilde{R}_{lt} & \tilde{R}_{tt}\end{pmatrix}\begin{pmatrix}\tilde{\phi}_{\text{inc}}\\ \phi_{\text{inc}}\\ \psi_{\text{inc}}\end{pmatrix}, \tag{9}$$

where we have

$$p=\omega^2\rho_f\tilde{\phi} \tag{10}$$

from Eq. (4). Expressions for the matrix elements are given by Eqs. (4.2.8)–(4.2.15) of Brekhovskikh and Godin,[9] with the shear modulus set equal to zero for the fluid half-space (see also their Secs. 4.2.2 and 4.2.3). Each coefficient is a function of κ or, alternatively, any of the angles $\theta_{f,l,t}$ via the second of Eqs. (8).

B. Angular spectra and absorption

Diffraction effects are taken into account by representing the field in terms of its angular spectrum. Because the fluid–solid interfaces shall be taken to be planes perpendicular to the z axis, we employ the following spatial Fourier transform pair:

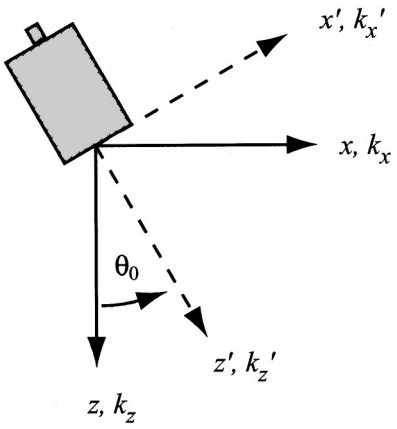


FIG. 3. Coordinates defining rotation of the source with respect to the plane parallel to the fluid–solid interface.

$$\hat{f}(\boldsymbol{\kappa}) = \mathcal{F}\{f(\mathbf{x})\} = \int f(\mathbf{x}) e^{-i\boldsymbol{\kappa}\cdot\mathbf{x}} d\mathbf{x}, \tag{11}$$

$$f(\mathbf{x}) = \mathcal{F}^{-1}\{\hat{f}(\boldsymbol{\kappa})\} = \int \hat{f}(\boldsymbol{\kappa}) e^{i\boldsymbol{\kappa}\cdot\mathbf{x}} \frac{d\boldsymbol{\kappa}}{(2\pi)^2}, \tag{12}$$

where $\mathbf{x}=(x,y)$ and $\boldsymbol{\kappa}=(k_x,k_y)$ are the coordinates and wave numbers, respectively, in planes parallel to the interfaces, and we have employed the notation $d\mathbf{x}=dxdy$ and $d\boldsymbol{\kappa}=dk_x dk_y$. With the sound pressure expressed as

$$p(\mathbf{x},z,t) = \tfrac{1}{2} P(\mathbf{x},z) e^{-i\omega t} + \text{c.c.}, \tag{13}$$

and if the source spectrum is given in the plane $z=0$ by

$$\hat{P}(\boldsymbol{\kappa},0) = \mathcal{F}\{P(\mathbf{x},0)\}, \tag{14}$$

then the solution of Eq. (1) for the sound pressure in a plane $z>0$ is[10]

$$P(\mathbf{x},z) = \mathcal{F}^{-1}\{\hat{P}(\boldsymbol{\kappa},0) e^{iK_f z}\}, \quad K_f = \sqrt{k_f^2 - |\boldsymbol{\kappa}|^2}. \tag{15}$$

Several frequently encountered source functions, e.g., Gaussian, circular piston, and rectangular piston, admit simple analytic expressions for $\hat{P}(\boldsymbol{\kappa},0)$.

We wish to accommodate source functions that are defined in planes that may not be perpendicular to the z axis. Such a situation is depicted in Fig. 3, where the source function is considered known in an $x'y$ plane that is inclined at an angle θ_0 with respect to the xy plane. The transformations between the xz and $x'y'$ coordinate (and wave number) systems associated with rotation by angle θ_0 about the y axis are given by

$$\begin{pmatrix} \xi' \\ \zeta' \end{pmatrix} = \begin{pmatrix} \cos\theta_0 & -\sin\theta_0 \\ \sin\theta_0 & \cos\theta_0 \end{pmatrix} \begin{pmatrix} \xi \\ \zeta \end{pmatrix}, \tag{16}$$

where $(\xi,\zeta)=(x,z)$ for the coordinate transformation, $(\xi,\zeta)=(k_x,k_z)$ for the wave number transformation, and with $y=y'$, $k_y=k_y'$, and $k_z=K$. Given a field distribution $f(\mathbf{x}',0)$ in the (rotated) plane $z'=0$ and its corresponding angular spectrum $\hat{f}(\boldsymbol{\kappa}',0)$, substitution of the coordinate transformations in the Fourier integrals yields

$$\hat{f}(\boldsymbol{\kappa},z=0) = \frac{|k_z'|}{|k_z|} \hat{f}(\boldsymbol{\kappa}',z'=0) \tag{17}$$

for the angular spectrum in the plane $z=0$.

To include absorption, we assume that the waves in the fluid and the solid form reasonably directional beams that propagate along axes determined nominally by the ray paths indicated in Fig. 2. For example, the distance propagated by the incident sound beam from the source to the first interface in Fig. 1 is nominally $d_i/\cos\theta_0$, where d_i is the distance traveled in the z direction. If the absorption coefficient for the fluid is α_f, the attenuation experienced by the incident sound beam is approximately $e^{-\alpha_f d_i/\cos\theta_0}$. Because the angular spectra are propagated in the z direction, attenuation is taken into account by rendering the wave numbers $K=k_z$ complex. Each is replaced with the expression

$$\tilde{K} = K + i\alpha_z, \quad \alpha_z = \alpha/\cos\theta_z, \tag{18}$$

where α is the absorption coefficient appropriate for the medium (α_f for the sound wave in the fluid, α_l or α_t for the longitudinal or transverse wave in the solid, respectively), and θ_z is the angle formed by the z axis and the propagation direction of the given plane wave in the angular spectrum, i.e., $\cos\theta_z = K/k$.

C. Reflection and transmission of a beam

The sound field incident on the solid is given by Eq. (15). We consider the reflected sound field in the receiver plane $z=z_r$, a distance d_r away from the interface (see Fig. 1). The source plane $z=0$ is distance d_i from the interface, and the spectrum of the incident beam in that plane is $\hat{P}(\boldsymbol{\kappa},0)$ [obtained using Eq. (17) if the source is rotated]. The reflected pressure field in the plane of the receiver is

$$P_r(\mathbf{x},z_r) = \mathcal{F}^{-1}\{R_{ll}(\boldsymbol{\kappa}) \hat{P}(\boldsymbol{\kappa},0) e^{iK_f(d_i+d_r)}\}, \tag{19}$$

where d_i+d_r is the total propagation distance along the z axis from the source to the interface and back to the receiver. Since the reflection coefficient R_{ll} depends on the magnitude $\kappa=|\boldsymbol{\kappa}|$, the second of Eqs. (8) permits R_{ll} to be expressed as a function of any of the angles in Fig. 2.

We consider now the transmitted field in the receiver plane $z=z_t$, a distance d_t below a solid of thickness h (see Fig. 1). There are two contributions to this field, one (P_t^{long}) due to the longitudinal wave in the solid and the other (P_t^{tran}) due to the transverse wave. Multiple reflections within the solid are not considered. The total transmitted field is the sum of these two contributions,

$$P_t(\mathbf{x},z_t) = P_t^{\text{long}}(\mathbf{x},z_t) + P_t^{\text{tran}}(\mathbf{x},z_t), \tag{20}$$

where

$$P_t^{\text{long}}(\mathbf{x},z_t) = \mathcal{F}^{-1}\{T_{ll}(\boldsymbol{\kappa}) \tilde{T}_{ll}(\boldsymbol{\kappa}) \hat{P}(\boldsymbol{\kappa},0) e^{iK_f(d_i+d_t)+iK_l h}\}, \tag{21}$$

$$P_t^{\text{tran}}(\mathbf{x},z_t) = \mathcal{F}^{-1}\{T_{lt}(\boldsymbol{\kappa}) \tilde{T}_{tl}(\boldsymbol{\kappa}) \hat{P}(\boldsymbol{\kappa},0) e^{iK_f(d_i+d_t)+iK_t h}\}. \tag{22}$$

In the expression for P_t^{tran}, the coefficient T_{lt} accounts for conversion of the sound wave in the fluid at the upper interface into a transverse wave in the solid, and $\tilde{T}_{tl}$ accounts for conversion of the transverse wave in the solid at the lower interface back into a sound wave in the fluid. When the axis of the incident sound beam is nearly normal to the interface ($\theta_0 \simeq 0$), excitation of the transverse wave is very inefficient

$(T_{lt}\simeq 0)$ and therefore $P\simeq P_t^{\text{long}}$. For angles of incidence in the neighborhood of the critical angle where the longitudinal wave in the solid becomes evanescent yet the transverse wave is excited, one obtains $P\simeq P_t^{\text{tran}}$.

III. QUASILINEAR THEORY

The results presented in Sec. II for the fields at the source frequency ω are used below to construct solutions for the fields at the second-harmonic frequency 2ω. The integrals for second-harmonic generation in the fluid and solid are developed independently and then connected at the interfaces by the reflection and transmission coefficients.

A. Wave interaction in the fluid

A solution for a lossless fluid is developed first, and then Eqs. (18) are used to include absorption *ad hoc*. Specifically, we begin with the following second-order wave equation for an ideal fluid:[11]

$$\left(\nabla^2-\frac{1}{c_f^2}\frac{\partial^2}{\partial t^2}\right)p=-\frac{\beta_f}{\rho_f c_f^4}\frac{\partial^2 p^2}{\partial t^2}-\left(\nabla^2+\frac{1}{c_f^2}\frac{\partial^2}{\partial t^2}\right)\mathcal{L}, \tag{23}$$

where $\mathcal{L}=\frac{1}{2}(\rho_f\mathbf{v}\cdot\mathbf{v}-p^2/\rho_f c_f^2)$ is a Lagrangian density, $\mathbf{v}$ is the particle velocity vector, and β_f is the coefficient of nonlinearity. For liquids, one normally expresses the coefficient of nonlinearity in the form

$$\beta_f=1+B/2A, \tag{24}$$

where B/A is referred to as the parameter of nonlinearity.[12] Henceforth in this section we shall suppress the subscript f unless there is likelihood of confusion with expressions used later for solids.

We let $p=p_1+p_2$, where $|p_1|\gg|p_2|$, p_1 is the sound pressure at the primary frequency ω, p_2 the pressure at the nonlinearly generated second-harmonic frequency 2ω, and we use notation as in Eq. (13):

$$p_n(\mathbf{x},z,t)=\tfrac{1}{2}P_n(\mathbf{x},z)e^{-in\omega t}+\text{c.c.},\quad n=1,2. \tag{25}$$

The source condition is taken to be

$$p(\mathbf{x},z_0,t)=\tfrac{1}{2}[P_1(\mathbf{x},z_0)e^{-i\omega t}+P_2(\mathbf{x},z_0)e^{-i2\omega t}]+\text{c.c.}, \tag{26}$$

which corresponds to radiation at frequencies ω and 2ω from the plane $z=z_0$. The purpose of taking $P_2(\mathbf{x},z_0)\neq 0$ is to allow for second-harmonic generation by the primary beam incident on this plane.

Equation (23) is solved via successive approximations. In the first approximation the right-hand side is set to zero, and the solution of the resulting linear equation for the primary beam is denoted p_1. By way of Eq. (25) we have

$$(\nabla^2+k^2)P_1=0, \tag{27}$$

where $k=\omega/c_f$. Using Eqs. (12) and (15), we express the primary beam at $z>z_0$ as a superposition of plane wave modes,

$$P_1(\mathbf{x},z)=\mathcal{F}^{-1}\{\hat{P}_1(\boldsymbol{\kappa},z_0)e^{iK_1\Delta z}\}=\int\hat{P}_1(\boldsymbol{\kappa},z_0)e^{i\mathbf{k}\cdot\mathbf{r}}\frac{d\boldsymbol{\kappa}}{(2\pi)^2}, \tag{28}$$

where

$$K_n=\sqrt{(nk)^2-|\boldsymbol{\kappa}|^2},\quad n=1,2, \tag{29}$$

$\mathbf{k}=(\boldsymbol{\kappa},K_1)$, $\mathbf{r}=(\mathbf{x},\Delta z)$, and $\Delta z=z-z_0$.

Now separate P_2 as follows:

$$P_2=P_2^h+P_2^p, \tag{30}$$

where

$$P_2^h(\mathbf{x},z_0)=P_2(\mathbf{x},z_0), \tag{31}$$

$$P_2^p(\mathbf{x},z_0)=0. \tag{32}$$

The function P_2^h is the homogeneous solution of the equation

$$(\nabla^2+4k^2)P_2^h=0 \tag{33}$$

that satisfies the boundary condition in Eq. (31), and therefore

$$P_2^h(\mathbf{x},z)=\mathcal{F}^{-1}\{\hat{P}_2(\boldsymbol{\kappa},z_0)e^{iK_2\Delta z}\}, \tag{34}$$

where K_2 is given by Eq. (29) with $n=1$. Next substitute p_2 in the left-hand side of Eq. (23), with $P_2=P_2^p$. On the right-hand side substitute p_1, use the first-order relation $\nabla p=-\rho(\partial\mathbf{v}/\partial t)$ to eliminate $\mathbf{v}$ in favor of p, and finally substitute Eq. (28) to obtain

$$(\nabla^2+4k^2)P_2^p=\frac{2k^2}{\rho c^2}\int\int\tilde{\beta}(\mathbf{k}',\mathbf{k}'')e^{i(\mathbf{k}'+\mathbf{k}'')\cdot\mathbf{r}}\times\hat{P}_1(\boldsymbol{\kappa}',z_0)\hat{P}_1(\boldsymbol{\kappa}'',z_0)\frac{d\boldsymbol{\kappa}'\,d\boldsymbol{\kappa}''}{(2\pi)^4}, \tag{35}$$

where $|\mathbf{k}'|=|\mathbf{k}''|=k$. If $\Theta(\mathbf{k}',\mathbf{k}'')$ designates the angle between a given pair of plane-wave components in the primary beam, such that $\mathbf{k}'\cdot\mathbf{k}''=k^2\cos\Theta$, then one obtains

$$\tilde{\beta}(\Theta)=\beta-2\sin^2(\Theta/2)+\sin^4(\Theta/2). \tag{36}$$

Equation (36) is an augmented coefficient of nonlinearity that indicates the strength with which any given pair of plane-wave modes in the primary beam interact to generate a second harmonic.

The right-hand side of Eq. (35) represents the forcing function as a superposition of plane-wave modes having wave vectors $\mathbf{k}'+\mathbf{k}''$ and phase speeds $2\omega/|\mathbf{k}'+\mathbf{k}''|$. The latter differ from the sound speed c unless the given pair of plane waves propagate in directions that are collinear and thus $|\mathbf{k}'+\mathbf{k}''|=2\omega/c$. The phase mismatch at large interaction angles produces inefficient nonlinear interactions. When the radiation forms a directional sound beam, the main contributions to the solution of Eq. (35) result from nearly collinear interactions, for which Θ is small and the relatively weak angular dependence in Eq. (36) may be ignored. Measurements in an acoustic waveguide at large interaction angles have demonstrated both the dependence of $\tilde{\beta}$ on Θ and the corresponding effects of phase incoherence for the case of sum- and difference-frequency generation.[13,14] We shall henceforth consider only directional sound beams and replace $\tilde{\beta}(\mathbf{k}',\mathbf{k}'')$ in Eq. (35) with the constant β. This approximation is equivalent to setting $\mathcal{L}=0$ in Eq. (23).[11]

To solve Eq. (35) with $\tilde{\beta}$ replaced by β, begin by taking the Fourier transform of both sides and evaluating the resulting integrals over $\mathbf{x}$ and $\boldsymbol{\kappa}''$ to obtain

$$\left(\frac{\partial^2}{\partial z^2}+K_2^2\right)\hat{P}_2^p=\frac{2\beta k^2}{\rho c^2}\int e^{i(K_a+K_b)\Delta z} \times \hat{P}_1(\boldsymbol{\kappa}',z_0)\hat{P}_1(\boldsymbol{\kappa}-\boldsymbol{\kappa}',z_0)\frac{d\boldsymbol{\kappa}'}{(2\pi)^2}. \tag{37}$$

where

$$K_a=\sqrt{k^2-|\boldsymbol{\kappa}'|^2}, \quad K_b=\sqrt{k^2-|\boldsymbol{\kappa}-\boldsymbol{\kappa}'|^2}. \tag{38}$$

The solution of Eq. (37) that satisfies Eq. (32) is

$$\hat{P}_2^p(\boldsymbol{\kappa},z)=\int Q_f(\boldsymbol{\kappa},\boldsymbol{\kappa}',\Delta z) \times \hat{P}_1(\boldsymbol{\kappa}',z_0)\hat{P}_1(\boldsymbol{\kappa}-\boldsymbol{\kappa}',z_0)\frac{d\boldsymbol{\kappa}'}{(2\pi)^2}, \tag{39}$$

where

$$Q_f(\boldsymbol{\kappa},\boldsymbol{\kappa}',\Delta z)=\frac{2\beta k^2}{\rho c^2}\left(\frac{e^{i(K_a+K_b)\Delta z}-e^{iK_2\Delta z}}{K_2^2-(K_a+K_b)^2}\right). \tag{40}$$

The desired pressure at frequency 2ω is given by the inverse transform of $\hat{P}_2^h+\hat{P}_2^p$,

$$P_2(\mathbf{x},z)=\mathcal{F}^{-1}\{\hat{P}_2(\boldsymbol{\kappa},z_0)e^{iK_2\Delta z}+\hat{P}_2^p(\boldsymbol{\kappa},z)\}. \tag{41}$$

Evaluation of the second-harmonic field in any plane is thus determined completely by calculations in the source plane.

Equations (39)–(41) correspond to Eqs. (25) and (26) of Alais and Hennion,[5] and more closely to Eqs. (7)–(9) of Naze Tjøtta and Tjøtta,[7] when the latter results are evaluated at the sum frequency and minor scaling is introduced for application to second-harmonic generation. Alais and Hennion employ a parabolic approximation not used here, although this approximation affects their result in a relatively minor way. Our solution appears initially to be equivalent to the one presented by the Tjøttas, but theirs is expressed in terms of an augmented pressure variable that transforms Eq. (23) into the Westervelt equation[11] [Eq. (23) with $\mathcal{L}=0$]. Our solution is thus formally the same as that of the Tjøttas insofar as ignoring the angular dependence in Eq. (36) is equivalent to setting $\mathcal{L}=0$ in Eq. (23). The two approaches have different implications for the source condition on the second harmonic, but this distinction is not significant in the present investigation. Moreover, it is straightforward to retain the angular dependence in Eq. (36) by substituting $\tilde{\beta}$ for β in Eq. (40).

To include absorption on the basis of Eq. (18), recall that we impose the condition that the radiation form a reasonably directional sound beam. With α_1 and α_2 taken to be the absorption coefficients at the fundamental and second-harmonic frequencies, the following complex forms of $K_{a,b}$ and K_2 may be substituted in Eq. (40):

$$\tilde{K}_{a,b}=K_{a,b}+i\alpha_{a,b}, \quad \tilde{K}_2=K_2+i\alpha_c, \tag{42}$$

where

$$\alpha_{a,b}=\alpha_1/\cos\theta_{a,b}=(k/K_{a,b})\alpha_1, \quad \alpha_c=\alpha_2/\cos\theta_c=(2k/K_2)\alpha_2. \tag{43}$$

In terms of the notation used in Eq. (39) there is collinear interaction for $\boldsymbol{\kappa}'=\boldsymbol{\kappa}-\boldsymbol{\kappa}'$, i.e., for $\boldsymbol{\kappa}'=\frac{1}{2}\boldsymbol{\kappa}$. In this case $K_a=K_b=\frac{1}{2}K_2$ and Eq. (40) reduces to

$$Q_f=\frac{\beta k}{i2\rho c^2}\left(\frac{e^{-2\alpha_1\Delta z/\cos\theta}-e^{-\alpha_2\Delta z/\cos\theta}}{(\alpha_2-2\alpha_1)/\cos\theta}\right)e^{iK_2\Delta z}, \tag{44}$$

where terms quadratic in $\alpha_{1,2}$ have been neglected in the denominator. With $\alpha_{1,2}=0$, Eq. (44) reduces to

$$Q_f=\frac{\beta k\Delta z}{i2\rho c^2}e^{iK_2\Delta z}, \tag{45}$$

which describes resonant growth of the second harmonic.

B. Wave interaction in the solid

The purpose of the present section is to develop a simplified model that accounts for the dominant elastic wave interactions that result from a directional sound beam that propagates into an immersed elastic half-space. When the sound beam in the fluid is transmitted into the solid, mode conversion takes place, giving rise to transverse as well as longitudinal waves. The efficiency of second-harmonic generation in the longitudinal and transverse modes of propagation may be assessed by considering the resonance condition for plane-wave interaction,

$$|\mathbf{k}'_{l,t}+\mathbf{k}''_{l,t}|=2k_{l,t}, \tag{46}$$

where $|\mathbf{k}_{l,t}|=\omega/c_{l,t}$ is the wave number of the longitudinal or transverse primary wave. Equation (46) indicates the possibility of six interactions associated with the permutations of the incides, three corresponding to generation of a longitudinal wave at 2ω and three to generation of a transverse wave. We shall employ the notation, e.g., $LT\rightarrow L$ to indicate second-harmonic generation in the longitudinal mode $(2k_l)$ by one primary wave in the longitudinal mode $(\mathbf{k}'_l)$ and the other in the transverse mode $(\mathbf{k}''_t)$. The nominal value $c_l/c_t=2$ is used for the purpose of discussing resonance angles that satisfy Eq. (46) subject to the definition $\cos\Theta=\mathbf{k}'\cdot\mathbf{k}''/k'k''$.

For the three interactions that generate a longitudinal second-harmonic component, the resonance angles are $\Theta=0$ for $LL\rightarrow L$, $\Theta=104.5°$ for $LT\rightarrow L$, and $\Theta=120°$ for $TT\rightarrow L$. Only the first of these interactions, $LL\rightarrow L$, is significant. The other two involve widely separated components in the angular spectrum, and these interactions are negligible for the geometry under consideration, namely, a directional sound beam incident on the solid. For generation of a transverse second-harmonic component, the resonance angle is $\Theta=0$ for $TT\rightarrow T$, but no angle satisfies the resonance condition for either $TL\rightarrow T$ or $LL\rightarrow T$. The condition $\Theta=0$ for $TT\rightarrow T$ corresponds to generation of a second harmonic in the transverse mode by a single plane wave in the transverse mode, but this coupling does not occur at quadratic order in the equations of motion.[15] (Zabolotskaya[16] derived a wave equation retaining terms through cubic order and discusses mechanisms by which second-harmonic generation in

transverse-wave beams may occur.) Therefore, of the six possible interactions of interest in the present investigation, only the one involving exclusively longitudinal modes is significant. Resonance conditions for harmonic generation in isotropic solids are discussed in greater detail elsewhere.[17–19]

In view of these observations, an appropriate longitudinal-wave equation in terms of the particle displacement potential is

$$\left(\nabla^2-\frac{1}{c_l^2}\frac{\partial^2}{\partial t^2}\right)\phi=\frac{\beta_s}{c_l^4}\left(\frac{\partial^2\phi}{\partial t^2}\right)^2, \tag{47}$$

where

$$\beta_s=-\left(\frac{3}{2}+\frac{\mathcal{A}+3\mathcal{B}+\mathcal{C}}{\rho_s c_l^2}\right), \tag{48}$$

and $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are Landau's third-order elastic constants.[17] For a fluid these constants become[20] $\mathcal{A}=0$, $\mathcal{B}=-\rho_f c_f^2$, and $\mathcal{C}=\frac{1}{2}(1-B/A)\rho_f c_f^2$, for which Eq. (48) reduces to Eq. (24). For plane-wave propagation along the z axis, substitute the first-order relation $\partial^2\phi/\partial t^2=c_l^2\partial^2\phi/\partial z^2$ in the right-hand side of Eq. (47), introduce the particle displacement $u=\partial\phi/\partial z$, and thus recover the wave equation derived by Gol'dberg,[15]

$$\frac{\partial^2 u}{\partial z^2}-\frac{1}{c_l^2}\frac{\partial^2 u}{\partial t^2}=2\beta_s\frac{\partial u}{\partial z}\frac{\partial^2 u}{\partial z^2}. \tag{49}$$

When the retarded time $\tau=t-z/c_l$ is introduced in Eq. (47), the parabolic approximation is employed, and the result is expressed in terms of the particle velocity $v=\partial^2\phi/\partial z\partial\tau$, Zabolotskaya's paraxial wave equation for sound beams is recovered,[16]

$$\frac{\partial^2 v}{\partial z\partial\tau}-\frac{c_l}{2}\nabla_\perp^2 v=\frac{\beta_s}{2c_l^2}\frac{\partial^2 v^2}{\partial\tau^2}, \tag{50}$$

where $\nabla_\perp^2=\partial^2/\partial x^2+\partial^2/\partial y^2$. Equation (47) is thus expected to provide an accurate description of second-harmonic generation in longitudinal-wave beams by components of the angular spectrum whose directions of propagation are similar. As these small-angle interactions are the strongest in the spectrum of the primary beam, the entire second-harmonic field should be accurately modeled by Eq. (47).

Use of Eq. (47) is further motivated by the accuracy with which the terms depending on Θ in Eq. (36) may be ignored for a directional sound beam in a fluid. Note that if the linear Eq. (4) is used to eliminate ϕ in favor of p in Eq. (47), Eq. (23) is obtained with $\mathcal{L}=0$. The terms containing $\mathcal{L}$ are precisely those giving rise to the dependence on Θ in Eq. (36), and these corrections are negligible for directional beams.[11] Equation (47) is thus analogous to the Westervelt equation for fluids and should therefore yield accurate results for directional longitudinal-wave beams in isotropic solids.

We call attention to the fact that the elastic wave equations at second order are derived in Lagrangian coordinates.[15,16] We are ignoring the distinction between Eulerian and Lagrangian coordinates because the correction terms describe local effects that are negligible in comparison with the cumulative nonlinear effects associated with propagation of directional beams. Similarly, in Secs. II C and II D we use linear theory for the reflection and transmission coefficients to establish boundary conditions for the second-harmonic component, because the error at second order is again a local effect that may be ignored in comparison with the cumulative nonlinear effects. Discussions of distinctions between Eulerian and Lagrangian coordinates and linearization of boundary conditions in nonlinear acoustics are provided by Blackstock.[21]

The procedure for solving Eq. (47) follows precisely the method described in Sec. III A. Here we write $\phi=\phi_1+\phi_2$, where

$$\phi_n(\mathbf{x},z,t)=\tfrac{1}{2}\Phi_n(\mathbf{x},z)e^{-in\omega t}+\text{c.c.},\quad n=1,2, \tag{51}$$

with

$$\phi(\mathbf{x},z_0,t)=\tfrac{1}{2}[\Phi_1(\mathbf{x},z_0)e^{-i\omega t}+\Phi_2(\mathbf{x},z_0)e^{-i2\omega t}]+\text{c.c.} \tag{52}$$

From Eq. (47) we have

$$(\nabla^2+k_l^2)\Phi_1=0, \tag{53}$$

$$(\nabla^2+4k_l^2)\Phi_2=\tfrac{1}{2}\beta_s k_l^4\Phi_1^2, \tag{54}$$

the solutions of which are

$$\Phi_1(\mathbf{x},z)=\mathcal{F}^{-1}\{\hat{\Phi}_1(\boldsymbol{\kappa},z_0)e^{iK_1\Delta z}\}, \tag{55}$$

$$\Phi_2(\mathbf{x},z)=\mathcal{F}^{-1}\{\hat{\Phi}_2(\boldsymbol{\kappa},z_0)e^{iK_2\Delta z}+\hat{\Phi}_2^p(\boldsymbol{\kappa},z)\}, \tag{56}$$

where

$$\hat{\Phi}_2^p(\boldsymbol{\kappa},z)=\int Q_s(\boldsymbol{\kappa},\boldsymbol{\kappa}',\Delta z)\times\hat{\Phi}_1(\boldsymbol{\kappa}',z_0)\hat{\Phi}_1(\boldsymbol{\kappa}-\boldsymbol{\kappa}',z_0)\frac{d\boldsymbol{\kappa}'}{(2\pi)^2} \tag{57}$$

and

$$Q_s(\boldsymbol{\kappa},\boldsymbol{\kappa}',\Delta z)=\frac{\beta_s k_l^4}{2}\left(\frac{e^{i(K_a+K_b)\Delta z}-e^{iK_2\Delta z}}{K_2^2-(K_a+K_b)^2}\right). \tag{58}$$

The definitions of $K_{1,2}$ and $K_{a,b}$ are the same as in Eqs. (29) and (38) but with $k=k_l$. Absorption is included as in Eqs. (42).

C. Reflection from an elastic half-space

The solution for this case does not depend on harmonic generation inside the solid and can be constructed from the results in Sec. III A. As in Sec. II C, the reflected field is desired in the receiver plane $z=z_r$ a distance d_r away from the interface. Contributions from both the incident and reflected primary beams must be taken into account in the evaluation of the second-harmonic field. The angular spectra for the incident sound beam are

$$\hat{P}_{1i}(\boldsymbol{\kappa},z)=\hat{P}_1(\boldsymbol{\kappa},0)e^{iK_1 z}, \tag{59}$$

$$\hat{P}_{2i}(\boldsymbol{\kappa},z)=\int Q_f(\boldsymbol{\kappa},\boldsymbol{\kappa}',z)\times\hat{P}_1(\boldsymbol{\kappa}',0)\hat{P}_1(\boldsymbol{\kappa}-\boldsymbol{\kappa}',0)\frac{d\boldsymbol{\kappa}'}{(2\pi)^2}, \tag{60}$$

where $\hat{P}_1(\boldsymbol{\kappa},0)$ is the angular spectrum of the sound source at $z=0$ (we assume there to be no 2ω radiation from the source). The spectra immediately after reflection at $z=d_i$ are given by

$$\hat{P}_{1r}(\boldsymbol{\kappa},d_i)=R_{ll}(\boldsymbol{\kappa})\hat{P}_1(\boldsymbol{\kappa},0)e^{iK_1 d_i}, \tag{61}$$

$$\hat{P}_{2r}(\boldsymbol{\kappa},d_i)=R_{ll}(\boldsymbol{\kappa})\hat{P}_{2i}(\boldsymbol{\kappa},d_i). \tag{62}$$

These source conditions correspond to the spectra of P_1 and P_2 in Eq. (26). The solution at $z=z_r$ is thus provided by Eq. (41),

$$P_{2r}(\mathbf{x},z_r)=\mathcal{F}^{-1}\{\hat{P}_{2r}^{\mathrm{inc}}(\boldsymbol{\kappa},z_r)+\hat{P}_{2r}^{\mathrm{refl}}(\boldsymbol{\kappa},z_r)\}, \tag{63}$$

where

$$\hat{P}_{2r}^{\mathrm{inc}}(\boldsymbol{\kappa},z_r)=R_{ll}(\boldsymbol{\kappa})\hat{P}_{2i}(\boldsymbol{\kappa},d_i)e^{iK_2 d_r} \tag{64}$$

is the contribution due to second-harmonic generation in the incident primary beam, and

$$\hat{P}_{2r}^{\mathrm{refl}}(\boldsymbol{\kappa},z_r)=\int Q_f(\boldsymbol{\kappa},\boldsymbol{\kappa}',d_r) \times\hat{P}_{1r}(\boldsymbol{\kappa}',d_i)\hat{P}_{1r}(\boldsymbol{\kappa}-\boldsymbol{\kappa}',d_i)\frac{d\boldsymbol{\kappa}'}{(2\pi)^2} \tag{65}$$

is the contribution due to second-harmonic generation in the reflected primary beam.

D. Transmission through an immersed elastic solid

The desired sound field is in the receiver plane $z=z_t$, a distance d_t below a solid of thickness h. The angular spectrum of the sound beam incident on the solid is given again by Eqs. (59) and (60). (We must now distinguish between quantities for sound in the fluid and longitudinal waves in the solid with the notation f and l, respectively.) The angular spectra of the displacement potentials in the solid at the first interface, $z=d_i$, are given by

$$\hat{\Phi}_1(\boldsymbol{\kappa},d_i)=T_{ll}(\boldsymbol{\kappa})\frac{\hat{P}_1(\boldsymbol{\kappa},0)}{\omega^2\rho_f}e^{iK_{1f}d_i}, \tag{66}$$

$$\hat{\Phi}_2(\boldsymbol{\kappa},d_i)=T_{ll}(\boldsymbol{\kappa})\frac{\hat{P}_{2i}(\boldsymbol{\kappa},d_i)}{4\omega^2\rho_f}, \tag{67}$$

where the factors of $n^2\omega^2\rho_f$ $(n=1,2)$ in the denominators follow from using Eq. (10) to convert sound pressure into displacement potential in order to apply the transmission coefficients. The displacement potentials in the solid at the second interface $(z_h=d_i+h)$ are

$$\hat{\Phi}_1(\boldsymbol{\kappa},z_h)=T_{ll}(\boldsymbol{\kappa})\frac{\hat{P}_1(\boldsymbol{\kappa},0)}{\omega^2\rho_f}e^{i(K_{1f}d_i+K_{1l}h)}, \tag{68}$$

$$\hat{\Phi}_2(\mathbf{x},z_h)=T_{ll}(\boldsymbol{\kappa})\frac{\hat{P}_{2i}(\boldsymbol{\kappa},d_i)}{4\omega^2\rho_f}e^{iK_{2l}h}+\hat{\Phi}_2^p(\boldsymbol{\kappa},z_h), \tag{69}$$

where Eq. (69) follows from Eq. (56), with

TABLE I. Properties of the media used in the experiments.

Medium	ρ (kg/m^3)	c_l (m/s)	c_t (m/s)	β ⋯	h (cm)
Water	998	1486	⋯	3.5	⋯
Aluminum	2727	6381	3150	8.5	6
Acrylic	1182	2760	1415	10	10

$$\hat{\Phi}_2^p(\boldsymbol{\kappa},z_h)=\int Q_s(\boldsymbol{\kappa},\boldsymbol{\kappa}',h) \times\hat{\Phi}_1(\boldsymbol{\kappa}',d_i)\hat{\Phi}_1(\boldsymbol{\kappa}-\boldsymbol{\kappa}',d_i)\frac{d\boldsymbol{\kappa}'}{(2\pi)^2}. \tag{70}$$

The spectrum of the second harmonic in the fluid at z_h is obtained by multiplying Eq. (69) by $4\omega^2\rho_f\tilde{T}_{ll}$,

$$\hat{P}_{2t}(\boldsymbol{\kappa},z_h)=T_{ll}(\boldsymbol{\kappa})\tilde{T}_{ll}(\boldsymbol{\kappa})\hat{P}_{2i}(\boldsymbol{\kappa},d_i)e^{iK_{2f}h} +4\omega^2\rho_f\tilde{T}_{ll}(\boldsymbol{\kappa})\hat{\Phi}_2^p(\boldsymbol{\kappa},z_h). \tag{71}$$

The transmitted pressure field at 2ω is separated into its homogeneous and particular solutions, such that Eq. (41) may be used to obtain

$$P_{2t}(\mathbf{x},z_t)=\mathcal{F}^{-1}\{\hat{P}_{2t}(\boldsymbol{\kappa},z_h)e^{iK_{2f}d_t}+\hat{P}_{2t}^p(\boldsymbol{\kappa},z_t)\}, \tag{72}$$

where

$$\hat{P}_{2t}^p(\boldsymbol{\kappa},z_t)=\int Q_f(\boldsymbol{\kappa},\boldsymbol{\kappa}',d_t) \times\hat{P}_{1t}(\boldsymbol{\kappa}',z_h)\hat{P}_{1t}(\boldsymbol{\kappa}-\boldsymbol{\kappa}',z_h)\frac{d\boldsymbol{\kappa}'}{(2\pi)^2}. \tag{73}$$

When the primary beam transmitted into the fluid on the far side of the solid is so weak as to generate a negligible contribution to the second-harmonic field, the term $\hat{P}_{2t}^p$ may be ignored.

IV. EXPERIMENTS

Experimental investigations of reflection and transmission were conducted with thick plates of aluminum (thickness $h=6$ cm) and acrylic (thickness $h=10$ cm) immersed in distilled water in the configuration shown in Fig. 1. Properties of the three media are provided in Table I. For the water, the temperature of which was 21 ± 1 °C, standard values reported in the literature are given.[12,22] The densities and wave speeds for the solids were measured directly. The value of β_s for aluminum, which is not required in the simulations reported below but is provided for reference, is obtained by substituting in Eq. (48) measured values of the third-order elastic constants reported by Smith *et al.*[23] The value of β_s for the acrylic was deduced from the transmission experiment described in Sec. IV B.

The source transducer was made by Panametrics and radiated at frequency $f=1$ MHz from an unfocused circular piezoelectric element of effective radius $a=1.21$ cm. The collimation length of the beam in water is thus $\frac{1}{2}k_f a^2=31$ cm. Taking p_0 to characterize an effective uniform sound pressure on the face of the source, and letting θ_0 designate the angle in the xz plane between the incident beam

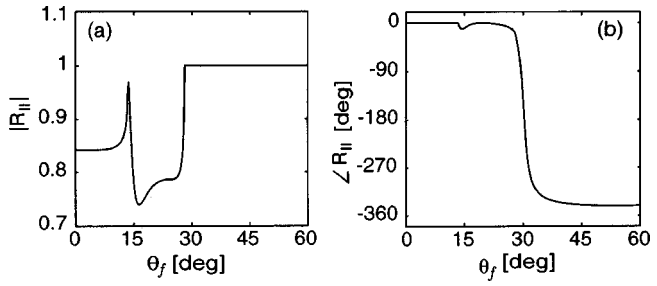


FIG. 4. (a) Magnitude and (b) phase of the reflection coefficient R_{ll} for a water–aluminum interface.

axis and the normal to the interface (see Figs. 1 and 3), we express the source condition in the rotated (primed) coordinate system of Eq. (16) as $P_1'(\mathbf{x}',0)=p_0$ for $|\mathbf{x}'|\leqslant a$ and $P_1'(\mathbf{x}',0)=0$ for $|\mathbf{x}'|>a$. The angular spectrum in the plane $z'=0$ is thus

$$\hat{P}_1'(\boldsymbol{\kappa}',0)=\pi a^2 p_0 \frac{2J_1(\kappa' a)}{\kappa' a}, \tag{74}$$

where $\kappa'=|\boldsymbol{\kappa}'|$, and J_1 is the first-order Bessel function of the first kind. Equation (74) is substituted in Eq. (17) to provide the source condition $\hat{P}_1(\boldsymbol{\kappa},0)$ in the plane $z=0$, which is used as the starting point for all calculations reported below. As indicated in Fig. 1, the center of the source coincides with the origin of the coordinate system, $(x,y,z)=(0,0,0)$.

Measurements of the sound pressure in the water at the fundamental (f) and second-harmonic ($2f$) frequencies were made with a Marconi bilaminar PVDF membrane hydrophone possessing an active element approximately 1 mm in diameter. The positioning apparatus, electronics, and signal processing have been described elsewhere.[8,24] Comparisons of linear theory with beam patterns measured in the free field determined the value of the radius a given above, and comparisons with pressure amplitudes determined p_0. Continuous radiation was simulated with 24-cycle tone bursts, for which the corresponding pulse lengths are approximately 4 cm in the water, 15 cm in the aluminum, and 7 cm in the acrylic.

A. Reflection

Experiments on reflection were conducted with the aluminum. The source was centered a distance $d_i=10$ cm away from the solid, and measurements of the reflected field were made in a plane parallel to the interface a distance $d_r=15$ cm away. Incident beam angles in the neighborhood of the Rayleigh angle, $\theta_R=30.4°$, were investigated because of the pronounced nonspecular effects associated with rapid variations in the reflection coefficient R_{ll}. Figure 4(a) shows that $|R_{ll}|=1$ for angles greater than $\theta_f=29.6°$, the critical incidence angle at which the transverse wave in the aluminum becomes evanescent (the longitudinal wave becomes evanescent at $\theta_f=13.5°$). The phase variation observed in Fig. 4(b) at $\theta_f\simeq 30°$ produces the nonspecular reflection phenomena that have traditionally been the focus of considerable interest.[1–4,25] Note that the coefficients in Eq. (9) are not functions of frequency, and therefore the same coefficient R_{ll} applies to both the f and $2f$ components in the beam. Because of the oblique angles of incidence in this experiment, the wave reflected from the far side of the aluminum plate did not interfere with the wave reflected from the first liquid–solid interface. The experiment thus models reflection from an elastic half-space.

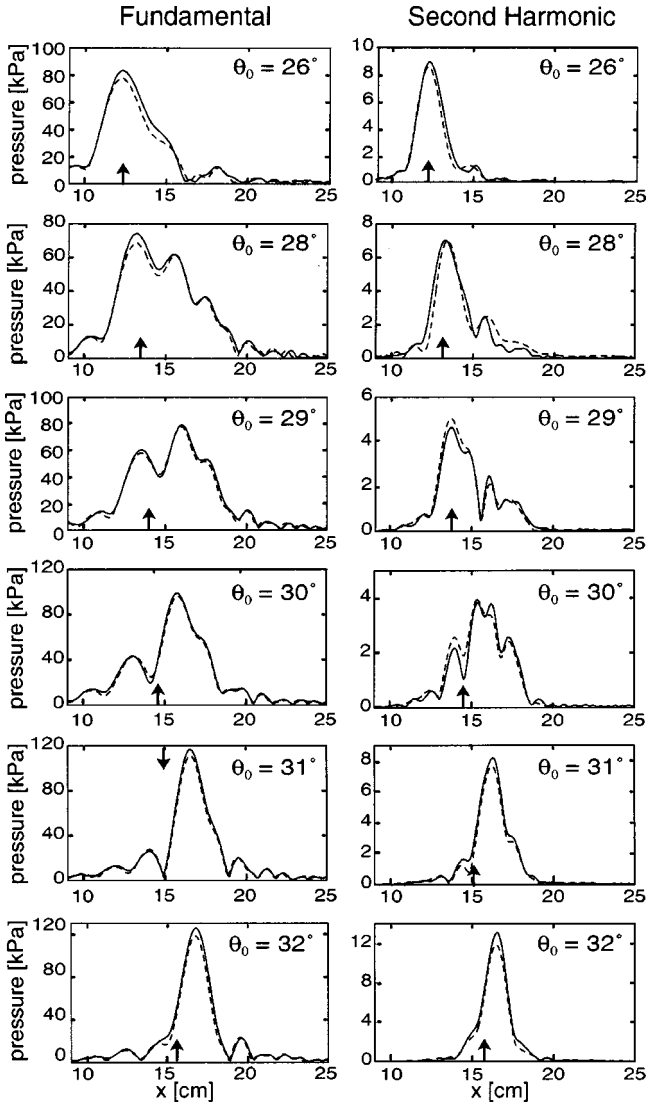


FIG. 5. Comparison of measurements (solid lines) and predictions (dashed lines) for the pressures of the fundamental (left column) and second-harmonic (right column) components in a sound beam reflected from aluminum in water ($d_i=10$, $d_r=15$ cm). The Rayleigh angle is $\theta_R=30.4°$. Arrowheads indicate locations of the beam axis predicted by ray theory.

Figure 5 shows measurements (solid lines) and predictions (dashed lines) of the reflected pressure field along the line $(x,y,z)=(x,0,z_r)$ at the fundamental (left column) and second-harmonic (right column) frequencies for a source pressure of $p_0=100$ kPa and $26°\leqslant\theta_0\leqslant 32°$. The theory in the left column is obtained from Eq. (19), and in the right column from Eq. (63). No curve fitting was employed. All calculations are based exclusively on the parameter values in Table I, the source condition in Eq. (74), and geometrical factors. Thermoviscous absorption in the water was negligible for the path lengths involved and was not included in the calculations. Quantitative agreement between theory and

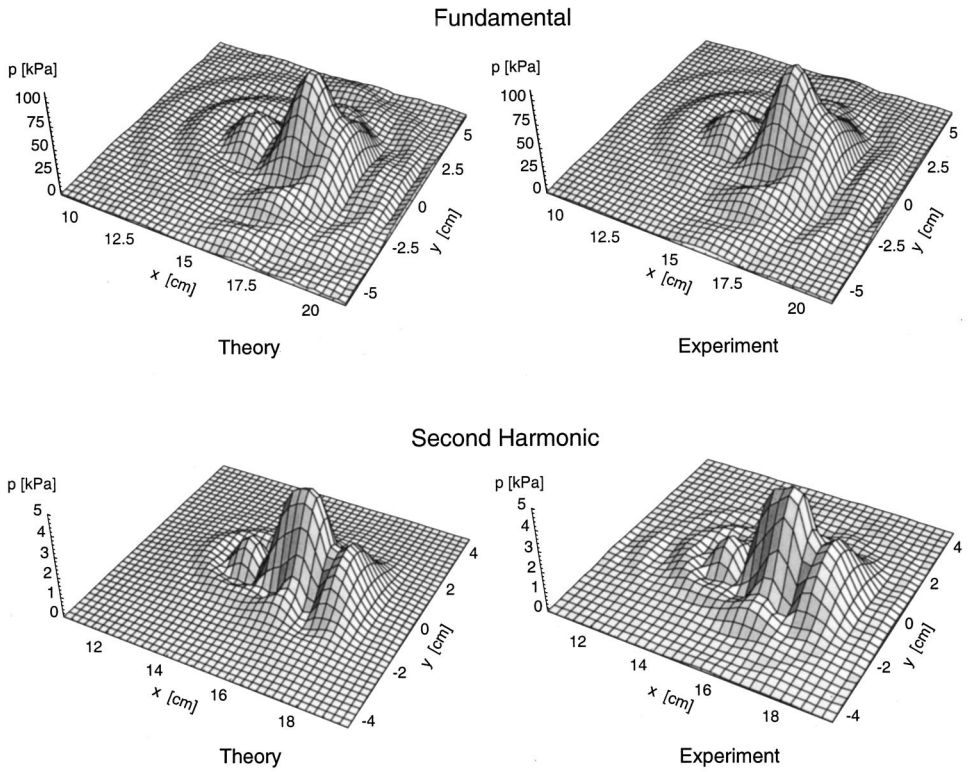


FIG. 6. Planar scans (vertical axis is sound pressure) corresponding to Fig. 5, with the incident beam axis at the Rayleigh angle $\theta_R = 30.4°$.

experiment is excellent. The arrowheads in Fig. 5 identify the coordinate

$$x_{\text{ray}} = (d_i + d_r)\tan\theta_0, \tag{75}$$

where ray theory predicts the reflected beam axis to be located.

We consider first the small-signal results in the left column. Linear acoustical phenomena associated with reflection near the Rayleigh angle are well understood and shall not be discussed extensively here. For a recent discussion see Chimenti *et al.*,[4] who also present quantitative comparisons of theory and experiment, although for Gaussian rather than uniform-piston beams, and with theory for two-dimensional rather than three-dimensional propagation. In Fig. 5 it is observed that the effective beam center shifts ahead of the prediction by ray theory as the Rayleigh angle is approached. A distinct minimum appears at $x \simeq x_{\text{ray}}$ for $\theta_0 \simeq 30°$ because of interference between the specularly reflected component and reradiation by the leaky Rayleigh wave. The pressure distributions in the plane $z = z_r$ that are shown in Fig. 6 for the fundamental demonstrate clearly both the interference phenomenon at the Rayleigh angle and the full three-dimensional validity of the theoretical model. They also show that at the Rayleigh angle the null at $x \simeq 15$ cm extends through the entire beam, which explains the dramatic visualization of the interference pattern with Schlieren photography.[3,25]

We consider next the results for the second harmonic, shown in the right column of Fig. 5. The excellent quantitative agreement between theory and experiment, particularly for the intricate field structure near the Rayleigh angle, provides confirmation of the approach used to model harmonic generation in the reflected field. The distributions shown in Fig. 6 for the second harmonic reveal agreement throughout the entire cross section of the beam.

One notable feature observed in the second-harmonic field is the decrease in amplitude to $|P_{2r}|_{\max} \simeq 4$ kPa at $\theta_0 = 30°$, compared with $|P_{2r}|_{\max} \simeq 8$ kPa at $\theta_0 = 26°$ and $|P_{2r}|_{\max} \simeq 12$ kPa at $\theta_0 = 32°$. The amplitude of the primary beam does not vary to such an extent. Equations (63)–(65) provide an explanation for this difference. In Fig. 7, the solid curves are the same calculations of $|P_{2r}|$ [Eq. (63)] that appear in the right column of Fig. 5, the short-dash lines are the $2f$ component $|P_{2r}^{\text{inc}}|$ [Eq. (64)] generated exclusively by the incident primary beam and reflected from the interface, and the long-dash lines are the $2f$ component $|P_{2r}^{\text{refl}}|$ [Eq. (65)] generated exclusively by the reflected primary beam. At $\theta_0 = 30°$ the values of $|P_{2r}^{\text{inc}}|_{\max}$ and $|P_{2r}^{\text{refl}}|_{\max}$ individually are

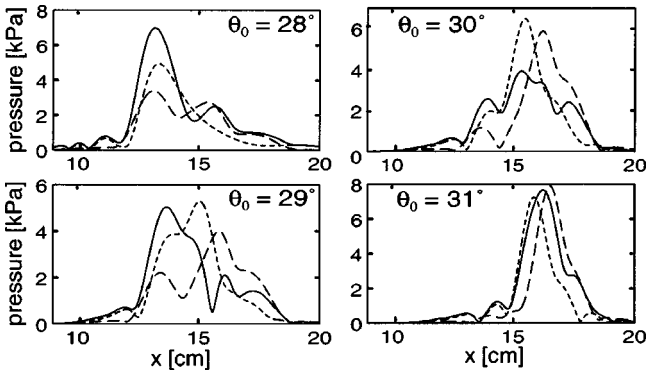


FIG. 7. Decomposition of the predicted second-harmonic pressures in Fig. 5 at various incidence angles. Solid lines are $|P_{2r}|$, short-dash lines are $|P_{2r}^{\text{inc}}|$, and long-dash lines are $|P_{2r}^{\text{refl}}|$.

considerably greater than the value $|P_{2r}|_{\max}=|P_{2r}^{\text{inc}}+P_{2r}^{\text{refl}}|_{\max}$ because of phase interference.

The phase interference occurs because P_1^2, which is the forcing function for second-harmonic generation, and P_2 experience different phase shifts following reflection. Suppose that before reflection, coherent second-harmonic generation has been established, and the incident fields P_{2i} and P_{1i}^2 are in phase. If we write $R_{ll}=|R_{ll}|e^{i\psi}$ for a given plane wave in the angular spectrum, then just after reflection the phase of P_{2r} is advanced by ψ, whereas the phase of P_{1r}^2 is advanced by 2ψ. A net phase difference of ψ is thus introduced, and the amplitude of the second harmonic is diminished for any $\psi\neq 0$. Moreover, the amplitude of the second harmonic will initially decrease with distance following reflection for $90°<\psi<270°$. The effect is strongest for $\psi=180°$, which is a situation that is well understood in connection with reflection from a pressure-release surface.[26,27] Inspection of Fig. 7 indicates that for reflection of a beam from an elastic halfspace, the effect is strongest at the Rayleigh angle. See Ref. 28 for further discussion of the influence that phase of the reflection coefficient has on second-harmonic generation.

B. Transmission

In immersion experiments configured as in Fig. 1 where harmonic generation in the solid is the subject of investigation, the question arises as to whether harmonic generation in the surrounding fluid may dominate measurements of the harmonics in the transmitted field. An estimate of these relative contributions to the transmitted field may be obtained by considering second-harmonic generation in a plane wave that propagates through the solid in a direction normal to the interfaces. For simplicity we ignore absorption in the fluid, but not in the solid. We also ignore harmonic generation in the fluid on the far side of the solid ($z>d_i+h$). The latter assumption is reasonable if the wave is sufficiently weakened by absorption in the solid or transmission through the interfaces, or if the receiver used to measure the transmitted field in the fluid is sufficiently close to the interface.

To investigate plane waves at normal incidence we require the transmission coefficients in Eq. (9) for the first and second interface, respectively, evaluated at $\theta_f=0$,[9]

$$T_{ll}=\frac{2(\rho_f/\rho_s)}{1+Z_f/Z_s},\quad \tilde{T}_{ll}=\frac{2(\rho_s/\rho_f)}{1+Z_s/Z_f}, \tag{76}$$

where $Z_f=\rho_f c_f$ and $Z_s=\rho_s c_s$ are the specific acoustic impedances of the fluid and the solid, respectively (it is convenient here to introduce c_s for the longitudinal wave speed c_l in the solid). Let the incident plane wave in the fluid have pressure amplitude p_0. The amplitude of the second harmonic in the incident field is $|P_{2i}|=|Q_f|p_0^2$, where Q_f is given by Eq. (45) with $\Delta z=d_i$. The amplitude of this wave after linear propagation through the solid is $|P_{2t}^f|=T_{ll}\tilde{T}_{ll}|P_{2i}|e^{-\alpha_2^s h}$, or

$$|P_{2t}^f|=2e^{-\alpha_2^s h}\frac{p_0^2\beta_f k_f d_i}{\rho_f c_f^2}\frac{Z_s/Z_f}{(1+Z_s/Z_f)^2}, \tag{77}$$

where $\alpha_n^s\,(n=1,2)$ designates the absorption coefficient for the longitudinal wave in the solid at frequency $n\omega$. Equation (77) accounts only for second-harmonic generation in the fluid region $z<d_i$.

We now calculate $|P_{2t}^s|$, the transmitted second-harmonic amplitude due to nonlinearity of the solid. The displacement potential amplitude of the incident wave in the fluid is $\Phi_{0f}=p_0/\omega^2\rho_f$, and transmission through the first interface gives $\Phi_{0s}=T_{ll}\Phi_{0f}$ in the solid. Second-harmonic generation in the solid is described by $|\Phi_{2s}|=|Q_s|\Phi_{0s}^2$. Evaluation of Eq. (58) for $\Delta z=h$, normal incidence, and with absorption taken into account yields $|Q_s|=\frac{1}{8}\beta_s k_s^3 h A_s e^{-\alpha_2^s h}$, where

$$A_s=\frac{e^{(\alpha_2^s-2\alpha_1^s)h}-1}{(\alpha_2^s-2\alpha_1^s)h} \tag{78}$$

is an absorption parameter. Note that $A_s\simeq 1$ is obtained for either $\alpha_{1,2}^s\simeq 0$ or $\alpha_2^s\simeq 2\alpha_1^s$. Transmission from the solid into the fluid yields $|\Phi_{2f}|=\tilde{T}_{ll}|\Phi_{2s}|$, for which the corresponding pressure is $|P_{2t}^s|=4\omega^2\rho_f|\Phi_{2f}|$. Combining the above relations, we have

$$|P_{2t}^s|=4A_s e^{-\alpha_2^s h}\frac{p_0^2\beta_s k_s h}{\rho_s c_s^2}\frac{(Z_s/Z_f)^2}{(1+Z_s/Z_f)^3}. \tag{79}$$

The ratio of Eq. (79) to (77) provides the desired measure of second-harmonic generation in the solid relative to that in the fluid,

$$\frac{|P_{2t}^s|}{|P_{2t}^f|}=\frac{2A_s}{1+Z_s/Z_f}\left(\frac{h}{d_i}\right)\left(\frac{\beta_s}{\beta_f}\right)\left(\frac{c_f}{c_s}\right)^2. \tag{80}$$

With $A_s=1$ for the aluminum (negligible absorption), $A_s=0.95$ for the acrylic (based on parameters given in the next paragraph), $h/d_i=2$, and with the values listed in Table I, Eq. (80) yields $|P_{2t}^s|=0.04|P_{2t}^f|$ for the aluminum and $|P_{2t}^s|=0.98|P_{2t}^f|$ for the acrylic. Although diffraction was not taken into account, these numbers suggest that nonlinearity of the aluminum contributes negligibly to the transmitted field in the water, whereas the contribution due to nonlinearity of the acrylic is comparable to that of the water. Moreover, insofar as the values of β_s are similar for the aluminum and the acrylic, the ability to measure second-harmonic generation in a solid using the immersion technique employed here is observed to depend critically on the small-signal properties of the materials.

Transmission experiments on second-harmonic generation were therefore performed with the acrylic. Small-signal tests determined the absorption coefficients for longitudinal waves in the acrylic to be $\alpha_l=12$ Np/m at 1 MHz and $\alpha_l=23$ Np/m at 2 MHz. These coefficients were used in the calculations. The next task was to determine the nonlinearity coefficient β_s. We accomplished this by comparing measurements with theory and considering β_s to be the only adjustable parameter. The experiments were performed using the configuration in Fig. 1. With the source at distance $d_i=5$ cm in front of the solid ($h=10$ cm), beam profiles were measured in a plane at distance $d_t=10$ cm behind the solid, along the line $(x,y,z)=(x,0,z_t)$. The source pressure was $p_0=420$ kPa, and the incident beam axis was rotated from $\theta_0=0$ to $\theta_0=30°$.

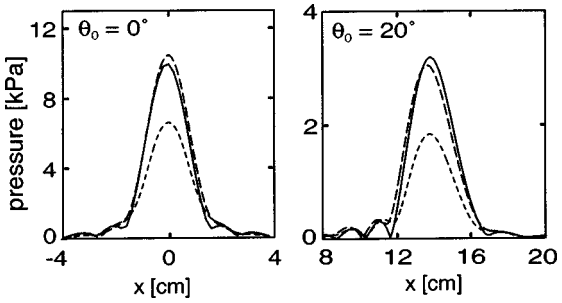


FIG. 8. Comparison of measurements (solid lines) and predictions (short-dash lines for $\beta_s=0$, long-dash lines for $\beta_s=10$) for the pressure of the second harmonic in a sound beam transmitted through a thick acrylic plate in water ($h=10$, $d_i=5$, $d_t=10$ cm).

Beam profiles measured at the second-harmonic frequency are shown as solid lines in Fig. 8 for $\theta_0=0$ and $\theta_0=20°$. The short-dash lines are theory obtained from Eq. (72) with $\beta_s=0$, i.e., with nonlinearity of the acrylic ignored and therefore $Q_s=0$ in Eq. (70). Without nonlinearity of the acrylic taken into account, theory underestimates experiment. The value of β_s was increased until theory matched experiment in the center of the beam, which was achieved with $\beta_s=10$. This is the value reported in Table I, and which was used to evaluate Eq. (72) and generate the long-dash curves in Fig. 8. The results on axis for normal incidence indicate that the contribution due to nonlinearity of the acrylic is similar to that due to nonlinearity of the water, $|P_{2t}^s|\sim|P_{2t}^f|$, which is consistent with the estimate based on Eq. (80).

The solid lines in Fig. 9 comprise the complete set of measurements for the transmitted field at frequencies f and $2f$ for $0°\leqslant\theta_0\leqslant30°$. Here the pressure amplitudes are expressed in dB (re 1 μPa) to emphasize the sidelobe structure. The dashed lines for the primary wave are the theory obtained from Eq. (21), and for the second harmonic they are

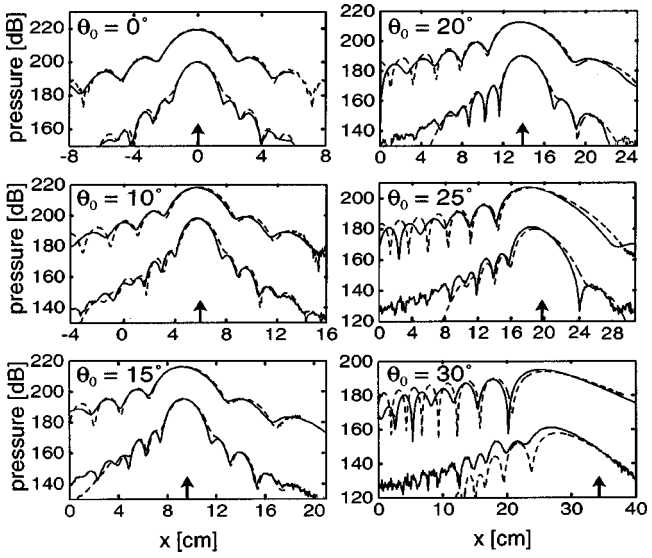


FIG. 9. Comparison of measurements (solid lines) and predictions (dashed lines) for the sound pressure levels (in dB *re* 1 μPa) of the fundamental and second-harmonic components in a sound beam transmitted through a thick acrylic plate in water ($h=10$, $d_i=5$, $d_t=10$ cm). Arrowheads indicate locations of the beam axis predicted by ray theory.

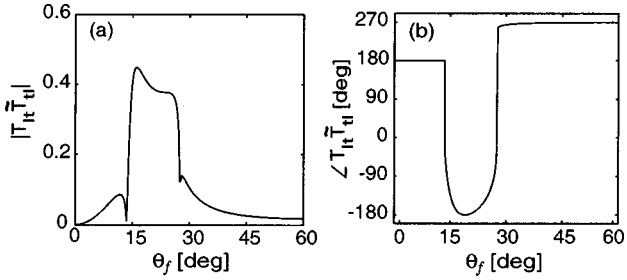


FIG. 10. (a) Magnitude and (b) phase of the composite transmission coefficient $T_{lt}\tilde{T}_{tl}$ for aluminum in water.

the theory obtained from Eq. (72). Apart from selecting $\beta_s=10$ as described above, no curve fitting was employed. Overall agreement between theory and experiment for both the fundamental and second harmonic is very good up to $\theta_0=25°$. At the critical angle $\theta_f=32.6°$, the longitudinal wave in the acrylic becomes evanescent. The coordinate x_{ray} of the transmitted beam axis predicted by ray theory and indicated by the arrowheads in Fig. 9 is

$$x_{\text{ray}}=(d_i+d_t)\tan\theta_0+h\tan\left[\arcsin\left(\frac{c_s}{c_f}\sin\theta_0\right)\right], \qquad (81)$$

where $c_s=c_l$ for the case at hand. Ray theory is an accurate predictor of the location of the main lobe in the diffraction pattern up to $\theta_0=20°$. At $\theta_0=30°$ the angle corresponding to the longitudinal wave in the acrylic is $\theta_l=68°$ (recall Fig. 2). Despite the substantial refraction at this angle, diffraction theory provides a reasonable description of not only the main lobe but also the first few sidelobes in both the fundamental and second-harmonic beam patterns.

We conclude with results for small-signal transmission through aluminum when the principal mode of propagation inside the solid consists of transverse waves. As explained in Sec. III B, a transverse wave field produces a negligible second-harmonic component. In this case we expect to have $|P_t^{\text{long}}|\ll|P_t^{\text{tran}}|$ in Eq. (20). The magnitude and phase of the composite transmission coefficient $T_{lt}\tilde{T}_{tl}$ that appears in Eq. (22) is presented in Fig. 10. At the critical angle $\theta_f=13.5°$ for the longitudinal wave in the aluminum, a notch is observed in the magnitude of $T_{lt}\tilde{T}_{tl}$ and a rapid variation is observed in the phase. When the axis of the incident sound beam in the water forms an angle in the vicinity of $\theta_0=13.5°$ at the first aluminum interface, longitudinal waves in the solid are refracted far away from the transverse waves. The longitudinal wave field in the water on the opposite side of the aluminum, generated by mode conversion of the transverse wave field at the second interface, should be influenced strongly by the rapid variation in the phase of $T_{lt}\tilde{T}_{tl}$. It is this case we examine experimentally to determine whether Eq. (22) accurately describes the transverse wave field in the solid and the corresponding mode conversions at the two interfaces. Equation (21) has already been verified by virtue of the agreement between theory and experiment in Fig. 9.

Experiments were performed at 1 MHz, at which frequency the absorption coefficient for transverse waves in the aluminum was determined to be $\alpha_t=0.05$ Np/m. The value is sufficiently small that absorption was ignored in the cal-

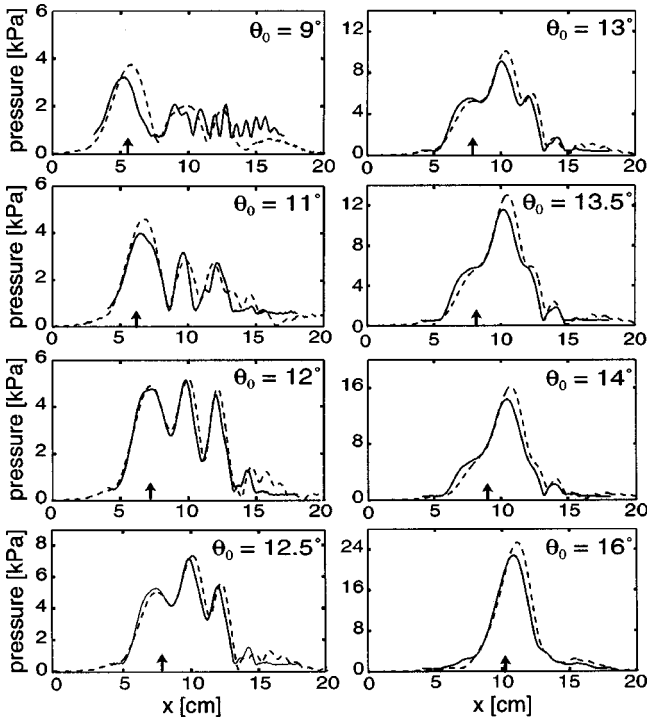


FIG. 11. Comparison of measurements (solid lines) and predictions (dashed lines) for the pressure of the fundamental component in a sound beam transmitted as a transverse wave through a thick aluminum plate in water ($h=6$, $d_i=10$, $d_t=10$ cm). The critical angle for the longitudinal wave in the aluminum is $\theta_f=13.5°$. Arrowheads indicate locations of the beam axis predicted by ray theory.

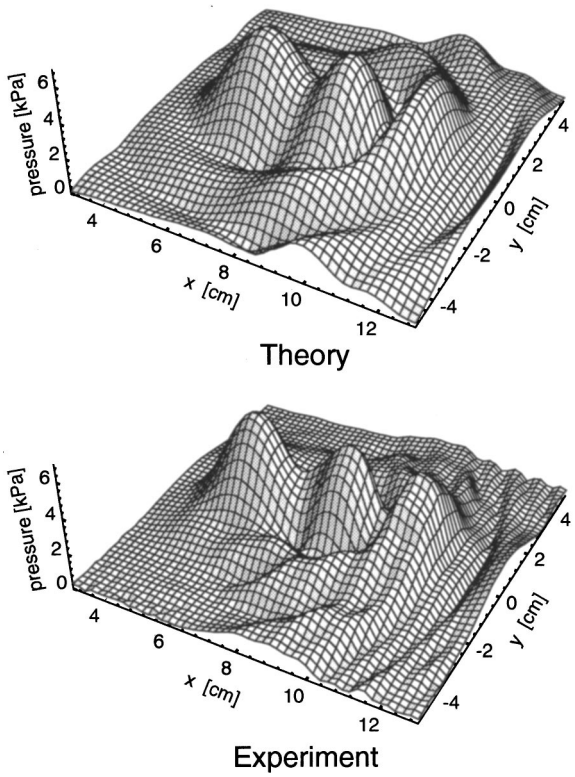


FIG. 12. Planar scans (vertical axis is sound pressure) corresponding to Fig. 11, with the incident beam axis at $\theta_0=12°$ (just below the critical angle at 13.5° for the longitudinal wave).

culations. The source was located $d_i=10$ cm in front of the aluminum plate ($h=6$ cm), and beam profiles in the transmitted field were measured along the line $(x,y,z)=(x,0,z_t)$ a distance $d_t=10$ cm behind the plate. Presented as solid lines in Fig. 11 are measurements with $p_0=72$ kPa and for $9° \leqslant \theta_0 \leqslant 16°$. The dashed lines are calculations obtained from Eq. (22). Except at $\theta_0=9°$, where the signal is very weak, agreement between theory and experiment is very good. Note in particular the interference structure in the beam profiles for $11° \leqslant \theta_0 \leqslant 12.5°$, as the critical angle $\theta_0=13.5°$ is approached and the rapid phase variations in $T_{lt}\tilde{T}_{tl}$ are encountered. Perhaps surprisingly, the interference structure subsides as the angle increases further, up to the critical angle and beyond. The arrowheads designate the locations x_{ray} predicted by ray theory, Eq. (81) with $c_s=c_t$. Not until $\theta_0=16°$ does the center of the main lobe in the beam realign with ray theory. The planar pressure distributions shown in Fig. 12, for $\theta_0=12°$ and the same plane $z=z_t$ as in Fig. 11, demonstrate the accuracy of the model throughout the sound field.

V. CONCLUSION

Comparisons of theory and experiment were presented for second-harmonic generation in sound beams reflected from and transmitted through thick isotropic elastic solids. The theoretical model combines angular spectrum theory for the diffraction and harmonic generation with classical reflection and transmission coefficients for the interfaces. Nonspecular phenomena associated with critical angles are accurately described. A number of the measurements of diffraction patterns appear to be novel: second-harmonic generation near the Rayleigh angle; second-harmonic generation due to oblique transmission through liquid–solid interfaces; transmission of a shear-wave beam near a critical angle. Theory and experiment are in close quantitative agreement.

ACKNOWLEDGMENTS

Discussions with Yu. A. Il'inskii and E. A. Zabolotskaya are gratefully acknowledged. This work was supported by the Office of Naval Research and the National Science Foundation.

[1] A. Schoch, ''Sound reflection, refraction, and bending,'' Ergeb. Exakten. Naturwiss. **23**, 127–134 (1950).
[2] L. M. Brekhovskikh, *Waves in Layered Media*, 2nd ed. (Academic, New York, 1980), Sec. 14.
[3] T. D. K. Ngoc and W. G. Mayer, ''Numerical integration method for reflected beam profiles near Rayleigh angle,'' J. Acoust. Soc. Am. **67**, 1149–1152 (1980).
[4] D. E. Chimenti, J.-G. Zhang, S. Zeroug, and L. B. Felsen, ''Interaction of acoustic beams with fluid-loaded elastic structures,'' J. Acoust. Soc. Am. **95**, 45–59 (1994).
[5] P. Alais and P. Y. Hennion, ''A Fourier theory of the nonlinear interaction of acoustical beams in absorbing fluid. The special case of parametric emission,'' Acustica **43**, 1–11 (1979).
[6] G. S. Garrett, J. Naze Tjøtta, and S. Tjøtta, ''Nearfield of a large acoustic transducer. Part III: General results,'' J. Acoust. Soc. Am. **75**, 769–779 (1984).
[7] J. Naze Tjøtta and S. Tjøtta, ''Interaction of sound waves. Part III: Two real beams,'' J. Acoust. Soc. Am. **83**, 487–495 (1988).
[8] B. J. Landsberger, ''Second harmonic generation in sound beams reflected

from and transmitted through immersed elastic solids,'' Ph.D. dissertation, The University of Texas at Austin, 1997.
[9] L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media I: Plane and Quasi-Plane Waves* (Springer-Verlag, New York, 1990).
[10] J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968).
[11] J. Naze Tjøtta and S. Tjøtta, ''Interaction of sound waves. Part I: Basic equations and plane waves,'' J. Acoust. Soc. Am. **82**, 1425–1428 (1987).
[12] R. T. Beyer, ''The parameter B/A,'' in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, Boston, 1998), Chap. 2.
[13] M. F. Hamilton and J. A. TenCate, ''Sum and difference frequency generation due to noncollinear wave interaction in a rectangular duct,'' J. Acoust. Soc. Am. **81**, 1703–1712 (1987).
[14] M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, ''Dispersion,'' in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, Boston, 1998), Chap. 5, Sec. 3.1.
[15] Z. A. Gol'dberg, ''Interaction of plane longitudinal and transverse elastic waves,'' Sov. Phys. Acoust. **6**, 306–310 (1961).
[16] E. A. Zabolotskaya, ''Sound beams in a nonlinear isotropic solid,'' Sov. Phys. Acoust. **32**, 296–299 (1986).
[17] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, 3rd ed. (Pergamon, New York, 1986), Sec. 26.
[18] G. L. Jones and D. R. Kobett, ''Interaction of elastic waves in an isotropic solid,'' J. Acoust. Soc. Am. **35**, 5–10 (1963).
[19] L. K. Zarembo and V. A. Krasil'nikov, ''Nonlinear phenomena in the propagation of elastic waves in solids,'' Sov. Phys. Usp. **13**, 778–797 (1971).
[20] S. Kostek, B. K. Sinha, and A. N. Norris, ''Third-order elastic constants for an inviscid fluid,'' J. Acoust. Soc. Am. **94**, 3014–3017 (1993).
[21] D. T. Blackstock, ''Propagation of plane sound waves of finite amplitude in nondissipative fluids,'' J. Acoust. Soc. Am. **34**, 9–30 (1962).
[22] L. E. Kinsler, A. F. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982), Eq. (5.22) and p. 462.
[23] R. T. Smith, R. Stern, and R. W. B. Stevens, ''Third-order elastic moduli of polycrystalline metals from ultrasonic velocity measurements,'' J. Acoust. Soc. Am. **40**, 1002–1008 (1966). Data taken for aluminum alloy 2S in Table III with $\mathcal{A}=4\nu_3$, $\mathcal{B}=\nu_2$, $\mathcal{C}=\frac{1}{2}\nu_1$.
[24] M. A. Averkiou and M. F. Hamilton, ''Nonlinear distortion of short pulses radiated by plane and focused circular pistons,'' J. Acoust. Soc. Am. **102**, 2539–2548 (1997).
[25] M. A. Breazeale, L. Adler, and G. W. Scott, ''Interaction of ultrasonic waves incident at the Rayleigh angle onto a liquid-solid interface,'' J. Appl. Phys. **48**, 530–537 (1977).
[26] A. L. Van Buren and M. A. Breazeale, ''Reflection of finite-amplitude ultrasonic waves. II. Propagation,'' J. Acoust. Soc. Am. **44**, 1021–1027 (1968).
[27] T. G. Muir, L. L. Mellenbruch, and J. C. Lockwood, ''Reflection of finite-amplitude waves in a parametric array,'' J. Acoust. Soc. Am. **62**, 271–276 (1977).
[28] I. R. S. Makin, M. A. Averkiou, and M. F. Hamilton, ''Second-harmonic generation in a sound beam reflected and transmitted at a curved interface,'' J. Acoust. Soc. Am. **108**, 1505–1513 (2000).

Observation of nonlinear acoustic effects at isotropic solid–solid interfaces

Jianjun Chen, Wenhua Jiang, and Yongan Shui
The Institute of Acoustics, State Key Laboratory of Modern Acoustics, Nanjing University, Nanjing 210093, People's Republic of China

(Received 9 February 1998; revised 10 September 1999; accepted 12 September 2000)

The second harmonic generation of SV shear waves at isotropic solid–solid interfaces is experimentally studied. The amplitude of shear waves is measured for the interfaces of glass–air, glass–iron, glass–copper, and glass–aluminum. The measured angular relation of amplitude of the second harmonic wave is compared with theory and the agreement is reasonably good. The influence of the physical state of the interface on second harmonic generation is also observed. It is found that the second harmonic generation is sensitive to the interface state. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1333418]

PACS numbers: 43.25.Jh [MAB]

I. INTRODUCTION

A boundary is inevitable because of the limited size of the solid sample. Bonded interfaces also exist in various kinds of composite materials. In addition, some interfaces originate from cracks in fatigued materials. The existence of interfaces has prominent influence on the nonlinear propagation of an acoustic wave in solid materials. Therefore, the nonlinear effects at the interface have attracted much attention.[1–5] Interest also comes from the fact that research in this field may find wide applications in nondestructive evaluation (NDE) of materials. There is considerable effort to provide a nonlinear acoustic method to evaluate the bonding strength of composite materials and nondestructive testing of the degree of fatigue in materials.[6,7]

One important nonlinear interface effect is the generation of the second harmonic shear wave through reflection at the interface.[8] It is known that the second harmonic shear wave will not be generated when a sinusoidal ultrasonic shear wave propagates in an unbounded isotropic solid under the assumption of quadratic nonlinearity.[9] However, the situation will be different when the sinusoidal ultrasonic shear wave encounters an interface of two isotropic solids. In this case the reflected, freely propagating, second harmonic shear wave will be generated as a result of the nonlinearities at the interface if the shear wave is obliquely incident upon the interface.[8]

Some theories have been developed to describe the nonlinear phenomenon for nondissipative isotropic as well as anisotropic solid interfaces.[4,5] A nonlinear reflective coefficient is defined to estimate the efficiency of the second harmonic generation. In this paper the angular dependence of the nonlinear reflective coefficient is measured experimentally for solid–solid and solid–air interfaces. The results are compared with the theory. The influence of the material properties of the refractive medium on the nonlinear reflection is also investigated. Emphasis is put on the peaks of nonlinear reflection coefficients which appear at some incident angles. The experimental observations of the influence of interface state on the nonlinear reflection of the SV wave are also presented.

II. EXPERIMENTAL OBSERVATION ON NONLINEAR GENERATION OF SHEAR WAVE AT ISOTROPIC SOLID–SOLID INTERFACES

As mentioned above, the bulk nonlinearity does not exist for the ultrasonic shear wave as it propagates in an unbounded isotropic solid under the assumption of quadratic nonlinearity. The generation of the second harmonic SV wave through reflection at an interface is an entirely nonlinear interface effect. To understand the experiment presented in this section, the successive approximate theory of the nonlinear reflection of shear wave at an isotropic solid-solid interface is repeated briefly here. The details of the theory can be found in Refs. 4 and 5. Then the experimental results are presented. Some discussions are also given.

A. Brief description of the successive approximation theory of nonlinear reflection

The coordinate system used in the calculation is shown in Fig. 1. The *XY* plane is the interface of two isotropic solids. The *Z* axis is normal to the interface and directed from incident medium (I) to refractive medium (II). A SV shear wave is obliquely incident to the interface in the *XZ* plane. It is assumed that all variables are independent of the *y* coordinate. Under the assumption of successive approximations the first-order reflection is not disturbed by the nonlinearity. Therefore, there will be reflective and refractive longitudinal and SV waves in incident and refractive media due to mode conversion. The propagation of all these waves in the nonlinear media will be governed by the following equations:

$$\ddot{\tilde{u}}_l - C_l^2 \nabla^2 \tilde{u}_l = \frac{1}{\rho_0} \nabla \cdot \mathbf{F}, \tag{1a}$$

$$\ddot{\tilde{\mathbf{u}}}_v - C_v^2 \nabla^2 \tilde{\mathbf{u}}_v = \frac{1}{\rho_0} \nabla \times \mathbf{F}, \tag{1b}$$

where C_l and C_v are the propagation velocities of longitudinal and shear waves, respectively. Here $\tilde{u}_l$ and $\tilde{\mathbf{u}}_v$ are the particle displacements associated with longitudinal and shear waves. They are related to particle displacement vectors $\mathbf{u}$ by

$$\tilde{u}_l = \nabla \cdot \mathbf{u} \quad \text{and} \quad \tilde{\mathbf{u}}_v = \nabla \times \mathbf{u}. \tag{1c}$$

In (1), $\mathbf{F}$ is the nonlinear force vector. Its components can be expressed as

$$\begin{aligned} F_i(\mathbf{u}) = {} & [C_{44} + \tfrac{1}{4}(2C_{155} - C_{112} + C_{123})](u_{l,kk}u_{l,i} + u_{l,kk}u_{i,l} + 2u_{i,lk}u_{l,k}) \\ & + [(C_{11} - C_{44} + \tfrac{1}{4}(2C_{155} + C_{112} - C_{123})](u_{l,ik}u_{l,k} + u_{k,lk}u_{i,l}) + [(C_{11} - 2C_{44} + \tfrac{1}{2}(C_{112} - C_{123})](u_{i,kk}u_{l,l}) \\ & + \tfrac{1}{4}(2C_{155} + C_{112} - C_{123})(u_{k,lk}u_{l,i} + u_{l,ik}u_{k,l}) + \tfrac{1}{2}(C_{112} + C_{123})u_{k,ik}u_{l,l}, \end{aligned} \tag{2}$$

where C_{IJ} and C_{IJK} are second- and third-order elastic constants. To solve these equations by the perturbation method of successive approximation, it is assumed that

$$\tilde{u}_l = \tilde{u}_l^0 + \tilde{u}_l^S + \cdots, \tag{3a}$$

$$\tilde{\mathbf{u}}_v = \tilde{\mathbf{u}}_v^0 + \tilde{\mathbf{u}}_v^S + \cdots, \tag{3b}$$

where $\tilde{u}_l^0$, $\tilde{u}_l^S$, $\tilde{\mathbf{u}}_v^0$, $\tilde{\mathbf{u}}_v^S$ are the fundamental and second harmonic particle displacements associated with longitudinal and shear waves, respectively. Substituting (3) into (1) and equating the terms of the same order, it is found that

$$\ddot{\tilde{u}}_l^0 - C_l^2 \nabla^2 \tilde{u}_l^0 = 0 \quad \text{or} \quad \ddot{u}_l^0 - C_l^2 \nabla^2 u_l^0 = 0, \tag{4a}$$

$$\ddot{\tilde{\mathbf{u}}}_v^0 - C_v^2 \nabla^2 \tilde{\mathbf{u}}_v^0 = 0 \quad \text{or} \quad \ddot{\mathbf{u}}_v^0 - C_v^2 \nabla^2 \mathbf{u}_v^0 = 0, \tag{4b}$$

and

$$\ddot{\tilde{u}}_l^S - C_l^2 \nabla^2 \tilde{u}_l^S = \frac{1}{\rho_0} \nabla \cdot \mathbf{F}', \tag{5a}$$

$$\ddot{\tilde{\mathbf{u}}}_v^S - C_v^2 \nabla^2 \tilde{\mathbf{u}}_v^S = \frac{1}{\rho_0} \nabla \times F', \tag{5b}$$

where u_l^0 and $\mathbf{u}_v^0$ are particle displacements of fundamental longitudinal and shear waves, respectively. Only second-order nonlinear terms are preserved in (2), i.e., $\mathbf{F}' = \mathbf{F}(\mathbf{u}^0)$. The solutions of (4) can be written as

$$u_\alpha^0 = \mathbf{U}_\alpha^0 \exp[j(\omega t - k_\alpha x \sin\theta_\alpha - k_\alpha z \cos\theta_\alpha)], \tag{6}$$

where $k_\alpha = \omega / C_\alpha$ ($\alpha = l$ or v) is the wave number of the α wave mode and θ_α is the angle of the wave propagation direction with respect to the Z axis.

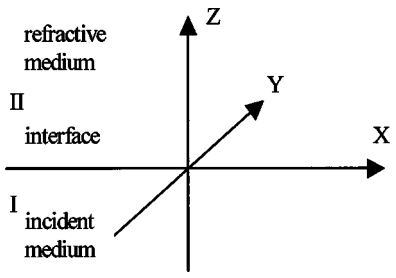


FIG. 1. The coordinate for calculation.

Substituting (6) into the right-hand sides of (5a) and (5b) turns (5a) and (5b) into linear inhomogeneous equations. The inhomogeneous terms are the second-order productions of the fundamental longitudinal and shear waves. This means that the particular solutions of (5a) and (5b) correspond to the second harmonic waves driven by the cross- or self-actions of the fundamental waves. These second harmonic waves are called driving waves as in Refs. 4 and 5 and can be expressed as

$$\begin{aligned} {}_\gamma\tilde{U}_{\alpha\mu}^{(D)} = {} & \frac{\beta_{\alpha\mu} U_\alpha^0 U_\mu^0}{C_\gamma^2 (k_{\alpha\mu}^2 - 4k_\gamma^2)} \exp[j(2\omega t - k_{\alpha\mu} x \sin\theta_{\alpha\mu} \\ & - k_{\alpha\mu} z \cos\theta_{\alpha\mu})] \quad (k_{\alpha\mu} \neq 2k_\gamma), \end{aligned} \tag{7a}$$

$$\begin{aligned} {}_\alpha\tilde{U}_\alpha^{(D)} = {} & \left(\frac{A_\alpha}{\sin\theta_\alpha} x + \frac{\beta_\alpha (U_\alpha^0)^2 - A_\alpha}{\cos\theta_\alpha} z \right) \exp[j2(\omega t \\ & - k_\alpha x \sin\theta_{g\alpha} - k_\alpha z \cos\theta_\alpha)] \quad (k_{\alpha\mu} = 2k_\alpha). \end{aligned} \tag{7b}$$

In the expression, ${}_\gamma\tilde{U}_{\alpha\mu}^{(D)}$ is the γ ($\gamma = l$ or v) mode second harmonic wave driven by the cross-action of the α mode and μ ($\alpha, \mu = l$ or v) mode fundamental waves; $\beta_{\alpha\mu}$ is the nonlinear parameter for the cross-action; and $k_{\alpha\mu}$ is the magnitude of the vector $\mathbf{k}_{\alpha\mu}$ which is the vector sum of two wave vectors $\mathbf{k}_\alpha$ and $\mathbf{k}_\mu$. Here $\theta_{\alpha\mu}$ is the angle between the propagation direction of the α wave mode and that of the μ wave mode. ${}_\alpha\tilde{U}_\alpha^{(D)}$ is the α mode second harmonic wave driven by the self-action of the α mode fundamental wave, A_α is the accumulation constant to be determined, and β_α is the nonlinear parameter for the self-action. It is seen from (7b) that the α mode second harmonic wave generated by the self-action of the α mode fundamental wave exhibits spatially two-dimensional accumulation.[4,5] The general solutions of (5a) and (5b) correspond to the second harmonic waves propagating freely in space and can be written as

$$\tilde{U}_\alpha^F = \tilde{U}_\alpha^S \exp[2j(\omega t - k_\alpha x \sin\theta_\alpha - k_\alpha z \cos\theta_\alpha)]. \tag{8}$$

At this point in the theory, no interface effects have been accounted for. The origin of nonlinear effects at the interface comes from the following fact. All the second harmonic waves mentioned above, both driving (7a) and (7b) and freely propagating (8), will generate second harmonic dis-

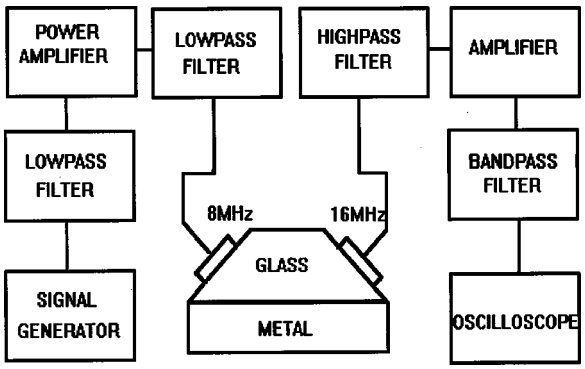


FIG. 2. The setup for measuring the nonlinear reflective SV wave on the isotropic solid–solid interface.

placements and stresses at the interface. Also, all fundamental waves, including incident, reflective and refractive waves, will generate second harmonic stresses at the interface through the nonlinear Hook's law as shown by (2). The resultant stresses $\mathbf{P}_{33}^{S},\mathbf{P}_{13}^{S}$ and displacements $\mathbf{U}_{3}^{S},\mathbf{U}_{1}^{S}$ at the second harmonic frequency should obey continuity conditions at the interface, that is,

$$\begin{bmatrix} U_3^{s,(I)} \\ U_1^{s,(I)} \\ \frac{1}{2}jP_{33}^{s,(I)} \\ \frac{1}{2}jP_{13}^{s,(I)} \end{bmatrix} = \begin{bmatrix} U_3^{s,(II)} \\ U_1^{s,(II)} \\ \frac{1}{2}jP_{33}^{s,(II)} \\ \frac{1}{2}jP_{13}^{s,(II)} \end{bmatrix} \quad \text{at } z=0. \tag{9}$$

From (7a) and (7b) it can be found that the resonant second harmonic driving waves exhibit spatially two-dimensional accumulation. The spatial coordinate x will appear in Eq. (9).

Since the boundary condition should be true for arbitrary x, then (9) is separated into two sets of equations as expressed as

$$\begin{vmatrix} A_{l-}^{(I)} \\ A_{v-}^{(I)} \\ A_{l+}^{(II)} \\ A_{v+}^{(II)} \end{vmatrix} = 0, \tag{10}$$

$$\begin{vmatrix} U_{l-}^{s,(I)} \\ U_{v-}^{s,(I)} \\ U_{l+}^{s,(II)} \\ U_{v+}^{s,(II)} \end{vmatrix} = \begin{vmatrix} R_{LV}^{S} \\ R_{VV}^{S} \\ T_{LV}^{S} \\ T_{VV}^{S} \end{vmatrix} k_v^{(I)} (U_{v+}^{(0)})^2. \tag{11}$$

In the expressions, ''+'' refers to a refractive wave and ''−'' refers to a reflective wave. From (10) the accumulation constants can be determined. From (11) the amplitude of the freely propagating second harmonic waves generated at the interface can be calculated. To estimate the efficiency of the interface generation of the freely propagating second harmonic SV wave, a nonlinear reflection coefficient is defined as

$$R_{VV}^{S}=\frac{U_{v-}^{S}}{k_v^{(I)}(U_{v+}^{(0)})^2}. \tag{12}$$

In the expression, $U_{v+}^{(0)}$ and U_{v-}^{S} are the amplitudes of the incident fundamental and the reflected freely propagating second harmonic SV waves, respectively. $k_v^{(I)}=\omega/C_v^{(I)}$ is the wave number of the SV wave in the incident medium. In the experiment reported in this article R_{vv}^{S} or U_{v-}^{S} is measured.

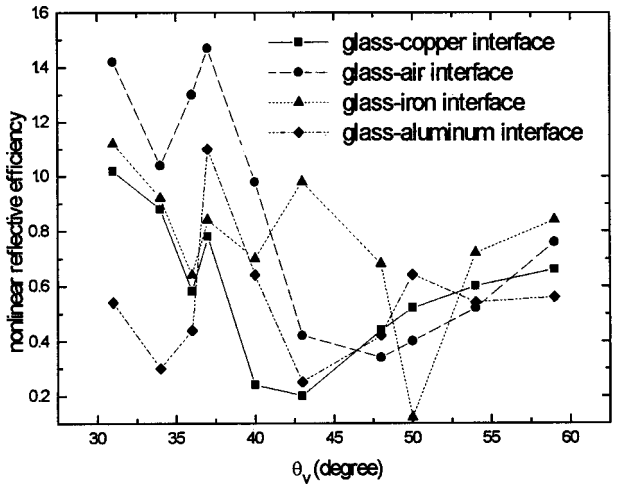


FIG. 3. The experimental results for glass–air, glass–iron, glass–copper, and glass–aluminum interfaces.

B. Experiments

The setup for measurement of the amplitude of reflected freely propagating second harmonic SV wave U_{v-}^{S} is shown in Fig. 2. In the experiment glass is chosen as the incident medium because of its negligible attenuation. Iron, copper, and aluminum are used as refractive media. The nonlinear reflection at the glass–air interface is also investigated. The glass sample is shaped as an optical Dove prism. The incident angle can be changed by changing the angle of its slant plane with respect to its base plane. The transducer is 136°-y cut $LiNbO_3$ plate. The transducers and glass prism as well as glass prism and refractive medium are bonded together with phenyl salicylate (Silo).

The fundamental SV wave is transmitted by a transducer with an 8-MHz center frequency and the reflected second harmonic SV wave is detected by a transducer with 16-MHz center frequency. The receiving transducer still responds to the 8-MHz fundamental wave. Thus, the amplitude of the incident wave can be monitored. Since the transducers are not calibrated, only the voltage proportional to the amplitude is measured in the experiment. The measured results for the interfaces of glass–iron, glass–copper, glass–aluminum, and glass–air are shown in Fig. 3. In the figure the horizontal axis is the incident angle, and the vertical axis is the voltage ratio of the second harmonic to the square of the fundamental wave, which is proportional to the nonlinear reflection coefficient defined by (12). From the figure it is observed that peaks of the second harmonic amplitude appear at some

TABLE I. Material parameters used in calculation of the nonlinear reflection coefficient.

	ρ (10^3 Kg/m^3)	C_{11} (GPa)	C_{44} (GPa)	C_{113} (GPa)	C_{155} (GPa)
Glass	4.5	61.9	22.3	40	90
Iron	7.7	284	84	−340	−760
Copper	8.38	168	37.3	−571	−509
Aluminum	2.77	107.9	26.4	−670	−395

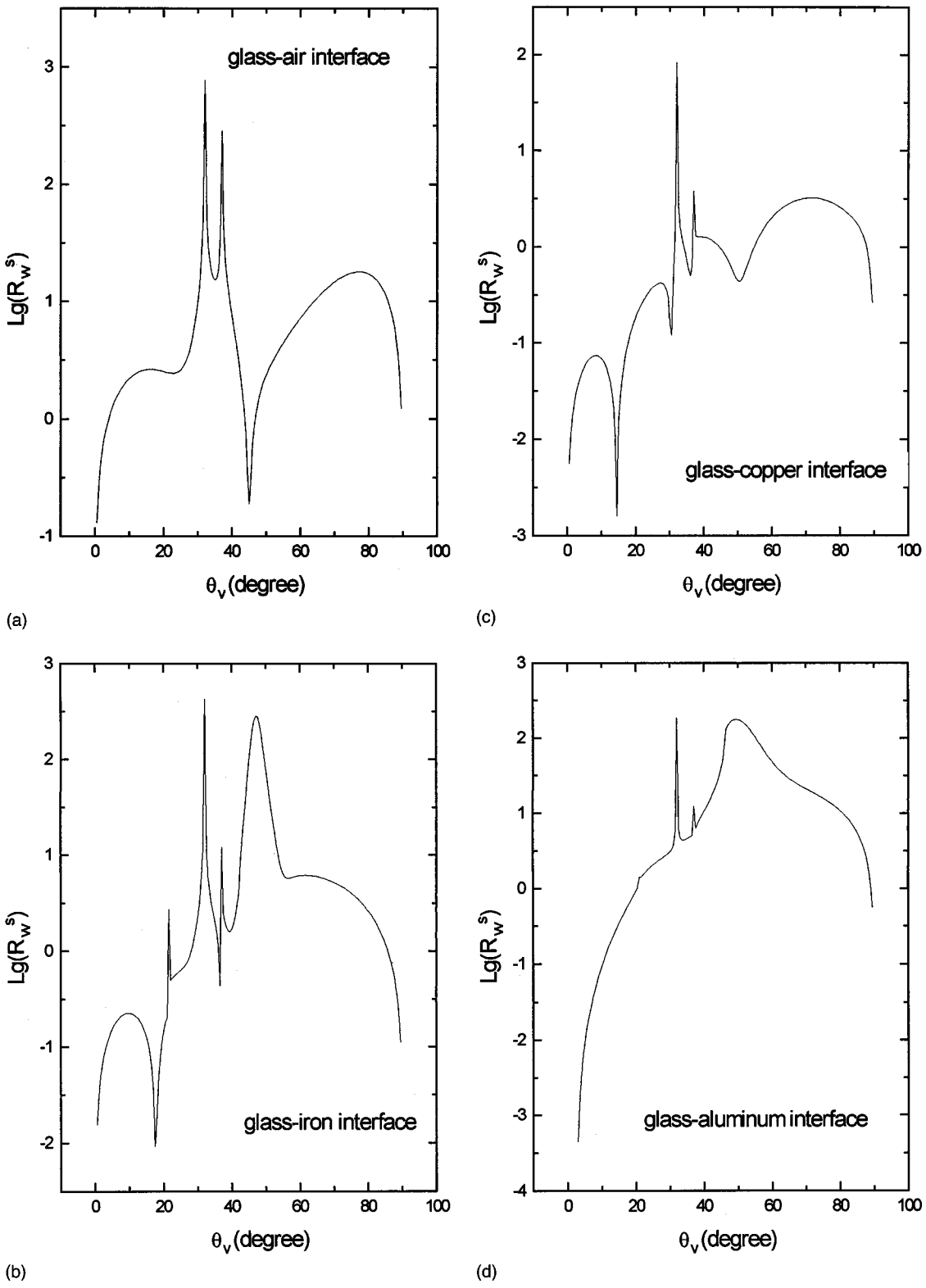


FIG. 4. (a) The calculated result for glass–air interface. (b) The calculated result for glass–iron interface. (c) The calculated result for glass–copper interface. (d) The calculated result for glass–aluminum interface.

incident angles. To understand this, the nonlinear reflection coefficient is calculated based on the theory given in Sec. II A. The material parameters used in the calculation are listed in Table I. The calculated results are shown in Fig. 4. From Fig. 3, it is noticed that a large value appears at the incident angle of 32° for all four interfaces. Comparing the results with Fig. 4, it is found that the incident angle θ_v satisfies the following equation:

$$\sin\theta_v = \sqrt{\frac{9k_l^2 - k_v^2}{8k_v^2}}. \tag{13}$$

Here k_l, k_v are the wave numbers of longitudinal and shear waves in the incident medium. At this incident angle, the resonant condition of cross-action between incident SV wave and reflected longitudinal wave

$$|\mathbf{k}_{VI}^{(I)}+\mathbf{k}_{LR}^{(I)}|=2k_l^{(I)} \tag{14}$$

is satisfied. Here $\mathbf{k}_{\mathrm{VI}}^{(I)}$ and $\mathbf{k}_{\mathrm{LR}}^{(I)}$ are the wave vectors of incident SV and reflected longitudinal waves. Under the resonant condition, the driving wave formed by the cross-action between incident SV wave and reflected longitudinal wave [see Eq. (7b)] will exhibit spatial accumulation.[4,5] This is the reason for the appearance of the maxima.

The calculation also gives peaks which appear near the critical angles of the refractive longitudinal wave, the reflective longitudinal and the refractive SV waves. Since the critical angle of the reflective longitudinal wave is independent of the refractive medium and is equal to 37° for glass, the peaks near the incident angle are observed in the experiment for all four interfaces, as shown in Fig. 3. For the glass–aluminum interface, the latter two critical angles, i.e., those of refractive longitudinal and refractive SV waves, are 21° and 46°, respectively. For the glass–iron interface, these critical angles are 21.5° and 42.4°, respectively. Since there is no glass prism with an angle less than 30° in the present experiment, the experimental observation cannot be conducted for incident angles less than 30°. Hence the peaks expected near the angles of 21° and 21.5° are not observed. For the same reason the peaks near 46° and 42.4° are not observed unambiguously in the experiment. However, the tendency for the peak to appear near these angles can be seen in Fig. 3. For the glass–copper interface, there is no critical angle of the refractive SV wave; the critical angle of the refractive longitudinal wave is 30°. In Fig. 4(c), it is observed that the peak near the critical angle for the refractive longitudinal wave is obscured by the peak at the resonant angle.

The fact that peaks appear in the vicinity of the critical angles has physical reasons. When the incident angle is larger than the critical incident angles, the scattering waves become evanescent. With the increase of angle, the penetration depth of the evanescent waves decreases, and the energy density at the interface increases. On one hand, the high energy density at the interface can cause strong generation of the reflective second harmonic SV wave. On the other hand, the linear reflection and refraction coefficients decrease with the increase of the incident angle. As a result, the amplitudes of these fundamental waves decrease. To balance these two effects, the peaks do not appear at the critical angles exactly, but in the vicinity of the critical angles. For example, for the glass–iron interface the peak near the critical angle of the refractive SV wave does not appear at the angle of 42.4°, but at the angle of 47°.

The calculation shows that the magnitude of the peaks in the vicinity of the critical angles is mainly determined by nonlinear characteristics of the incident or refractive media, and the position of the peaks is determined by linear characteristics of the media.

A numerical simulation can be performed: One keeps the SOE of iron unchanged and increases the TOE of iron by 100 times. The calculated result is shown in Fig. 5(a). Comparing Fig. 5(a) with Fig. 4(b), it is observed that the peaks of the amplitude of the reflective second harmonic wave are in the same location, but the magnitude of the peaks in the vicinity of the critical angles of refractive SV and longitudinal waves increases prominently. The peak value near the critical angle of the reflective longitudinal wave is almost unchanged. The result shows that the magnitude of the peaks

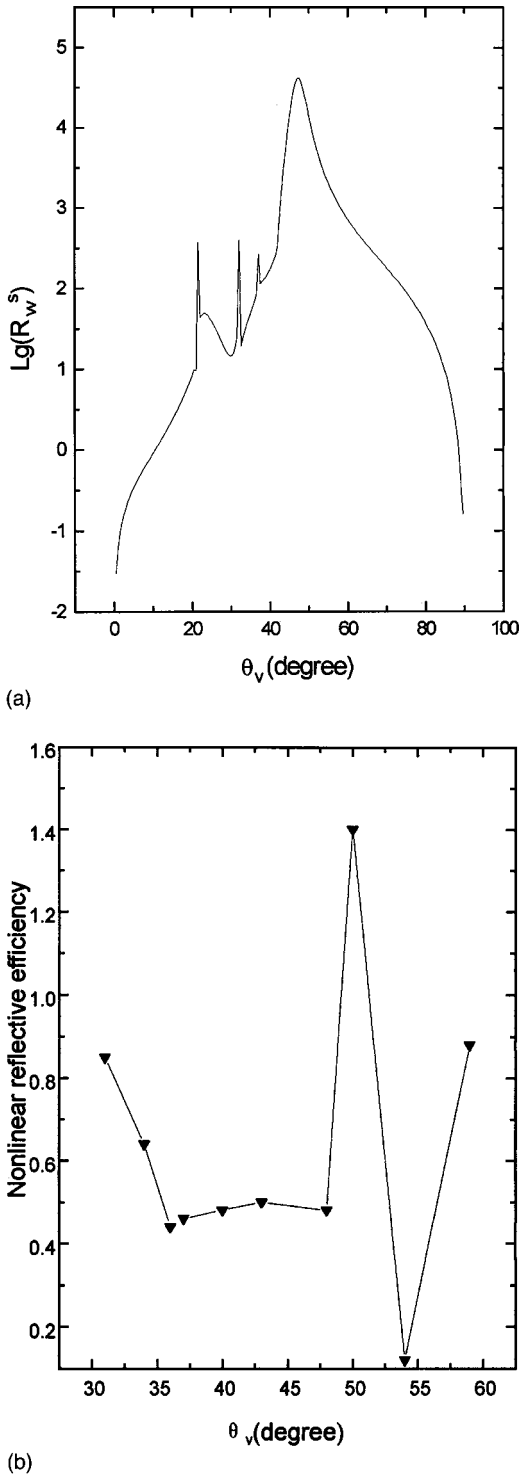


FIG. 5. (a) The calculated result for glass–iron interface when one keeps the SOE of iron unchanged and increases the TOE of iron by 100 times. (b) The experimental results for glass-rock Interface

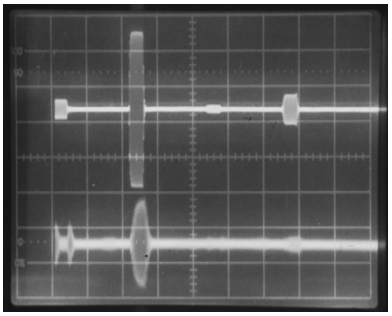


FIG. 6. The observed signals when Silo is recrystalized entirely.

in the vicinity of the critical angles of refractive SV and longitudinal waves mainly depends on the nonlinear characteristics of the refractive medium and the peak value near the critical angle of the reflective longitudinal wave is mainly determined by the nonlinear characteristics of the incident medium. The materials used in this experiment have similar nonlinear properties. So, the magnitude of the peaks does not differ by much. The experiment is repeated with a glass–rock interface. It is reported that rocks exhibit extremely strong nonlinearity.[10] The measured angular dependence of the amplitude of the nonlinear reflective SV wave is shown in Fig. 5(b). It is observed that strong nonlinear reflection appears in some incident angles even though the attenuation of rocks for ultrasonic waves is extremely large.

III. OBSERVATION OF THE INFLUENCE OF INTERFACE STATE ON NONLINEAR REFLECTION OF THE SV WAVE

The generation of freely propagating second harmonic SV waves is an interface effect and is strongly dependent on the linear and nonlinear properties of the media which compose the interface. Thus it is anticipated that the bonding state of the interface will affect the effect. In this section, two experiments are described which demonstrate these influences.

A. The influence of the state of the coupling agent on the nonlinear reflection of the SV wave

The setup for doing the experiment is the same as shown in Fig. 2. The incident medium is glass and the refractive medium is copper. The incident angle is 31°. As before, glass and copper are bonded together with Silo. The bonding procedure is usually as follows. First one puts Silo powder onto the surface of the sample copper and heats it until the Silo

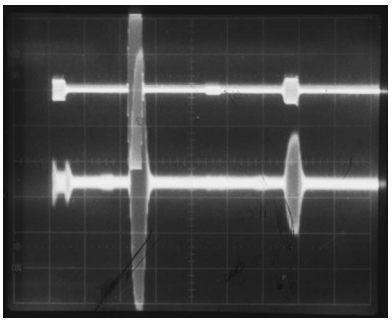


FIG. 7. The observed signals when Silo is melted entirely.

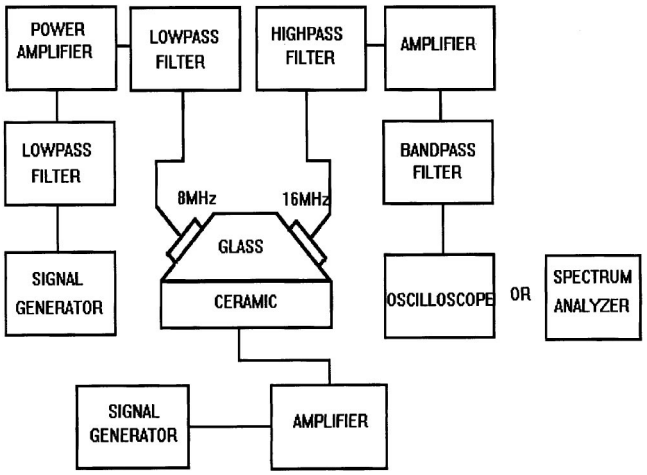


FIG. 8. The setup for measuring the nonlinear reflective SV wave on glass–PZT interface when PZT is vibrating.

melts completely (about 40 °C). Then the glass prism is placed on the copper sample. They are pushed together and cooled until the Silo is recrystallized entirely. In this way a good acoustical bond between glass and copper is usually attained. Under this condition, the observed echo trains of linear and nonlinear reflective SV waves are shown in Fig. 6. In the figure the upper trace stands for the linear reflection (fundamental) and the lower trace for the nonlinear (second harmonic) reflection.

If the assembly of glass–copper is heated by 70 °C water beneath the copper sample, the coupling agent will be heated slowly and melted gradually. During this process, the amplitude of the reflective SV wave is monitored. It is found that the amplitude of the nonlinear reflective SV wave increases. When the coupling agent is melted entirely, the amplitude of the nonlinear reflective SV wave increases to a maximum, as shown in Fig. 7. Comparing Fig. 6 with Fig. 7, it is found that the amplitude of the nonlinear reflective SV wave increases three to four times, while the amplitude of the linear reflective SV wave does not change appreciably. When the hot water is removed and the Silo is cooled and recrystallizes, it restores the original state as given in Fig. 6.

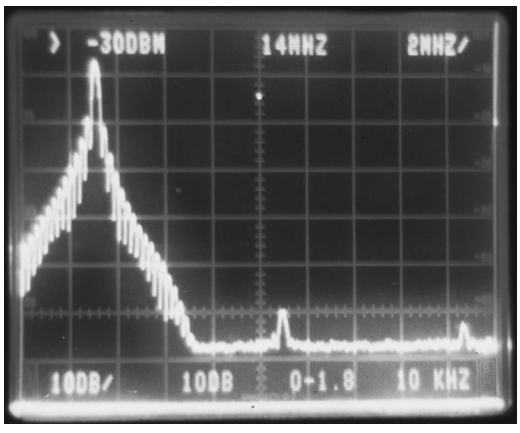


FIG. 9. The spectrum of the reflective fundamental SV wave when the PZT is not vibrating.

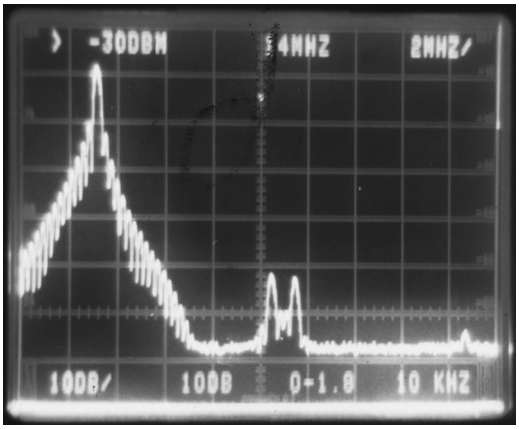


FIG. 10. The spectrum of the reflective fundamental SV wave when the PZT is vibrating.

B. The influence of the vibrating interface on the nonlinear reflection of the SV wave

The glass prism and PZT ceramic plate are bonded together in this experiment. The thickness of the ceramic plate is 0.5 cm, and the corresponding resonant frequency of the thickness-longitudinal vibration is 488.5 KHz. The setup for this experiment is shown in Fig. 8. When the ceramic plate is set to vibrate, it is observed that the nonlinear reflective SV wave becomes amplitude modulated. The modulation frequency is the resonant frequency of the PZT plate.

To show the modulation phenomenon, the signal output from the receiver transducer is analyzed by a frequency spectrum analyzer. When the PZT plate is vibrating it is seen that side lobes of ± 488.5 KHz appear on both sides of the main lobe of the second harmonic frequency; the main lobe has no changes, as shown in Figs. 9 and 10. It is also observed that the degree of modulation increases with the increase of amplitude of the PZT plate. No modulation has been observed for the linear reflective SV wave.

From the two experiments described here, it is found that the nonlinear reflection of the SV wave is more strongly dependent on the bonding state of the interface than the linear reflection of SV waves. The theoretical explanation for the effects has not been developed.

IV. CONCLUSIONS

The interface generation of the second harmonic SV wave is experimentally investigated. Some of the peaks of the nonlinear reflective coefficient R_{VV}^{S} predicted by the successive approximation theory are observed in the experiments. However, the magnitude of the peak is lower than that predicted by theory. The first reason for this may be that the dissipation of the refractive medium is ignored in the theoretical calculation. The second reason may be that the theoretical calculation assumes plane waves, while the experiment is carried out by using an almost collimated beam. The modification to the theory by taking the dissipation of the refractive medium into account is just under way. It will be reported later. The modification to the plane wave has not been done.

In the investigation reported here it is found that the amplitude of the nonlinear reflective SV wave reaches a considerable value at some incident angles. Moreover, the magnitudes of some amplitude peaks mainly depend on the nonlinear properties of the refractive medium. This result suggests developing an interface method to determine the third-order elastic constants of materials, including the materials with high attenuation if the dissipation of medium can be included in the theory. It is also observed in the experiment that the interface state has a pronounced influence on the second harmonic generation at the interface. Therefore this result may find applications in the development of nonlinear acoustic methods for nondestructive evaluation of materials, such as the evaluation of bonding strength of composite materials.

ACKNOWLEDGMENT

This work was supported by NFS of China.

[1] A. L. Van Buren and M. A. Breazeale, ''Reflection of finite-amplitude ultrasonic waves,'' J. Acoust. Soc. Am. **44**, 1014–1020 (1968).
[2] Z. W. Qian, ''Reflection of finite-amplitude sound wave on a plane boundary of half space,'' Sci. Sinica A **25**, 492 (1982).
[3] K. T. Shu and J. H. Ginsberg, ''Oblique reflection of a nonlinear *p* wave from the boundary of an elastic half space,'' J. Acoust. Soc. Am. **89**, 2652–2662 (1991).
[4] S. Zhou and Y. Shui, ''Nonlinear reflection of bulk acoustic wave at the interface,'' J. Appl. Phys. **72**, 5070–5079 (1992).
[5] S. Zhou, W. Jiang, and Y. Shui, ''Nonlinear bulk acoustic waves in anisotropic solids: Propagation generation and reflection,'' J. Appl. Phys. **78**, 39–46 (1995).
[6] A. S. Korotkov and A. M. Sutin, ''Modulation of ultrasound by vibrations in metal constructions with cracks,'' Acoustic Lett. **18**(4), 59–62 (1994).
[7] J. H. Cantrell and W. T. Yost, ''Acoustic harmonic generation and dislocation dynamics of fatigued aluminum alloys,'' Rev. Prog. Quant. Nondestr. Eval. **12**, 2059–2066 (1993).
[8] Y. Mao, Y. Shui, W. Jiang, Z. Lu, and W. Wu, ''Second-harmonic generation of interface waves,'' Appl. Phys. Lett. **55**, 2394–2396 (1989).
[9] R. E. Green, Jr., *Ultrasonic Investigation of Mechanical Properties* (Academic, New York, 1973), pp. 84–86.
[10] V. E. Nazarov, L. A. Ostrovsky, I. A. Soustova, and A. M. Sutin, ''Nonlinear acoustics of microinhomogeneous media,'' Phys. Earth Planet. Inter. **50**, 65 (1988).

Scattering from a partially fluid-filled, elastic-shelled sphere

J. A. Fawcett
Defence Research Establishment Atlantic, PO Box 1012, Dartmouth, Nova Scotia B2Y 3Z7, Canada

(Received 29 June 2000; revised 30 October 2000; accepted 15 November 2000)

In this article a sphere is taken to be partially filled with fluid so that its interior is part fluid and part air. A set of basis of functions, based upon an origin at the fluid/air interface, is used for the interior and a set of basis functions based upon the center of the sphere is used for the shell and exterior of the sphere. These sets of basis functions are coupled at the shell/interior interface and the resulting coupled system of equations solved to yield the scattered field. Numerical computations using this approach are presented for varying amounts of fluid-fill and for varying incident plane waves. [DOI: 10.1121/1.1339827]

PACS numbers: 43.30.Gv, 43.40.Fz [DLB]

I. INTRODUCTION

The modeling of acoustic scattering from an elastic-shelled sphere, either evacuated or filled, has been studied previously by a variety of authors.[1–4] In this article we describe an approach to modeling scattering from a partially filled elastic-shelled sphere. A schematic of a partially filled sphere is shown in Fig. 1. The partially filled interior introduces two interesting effects: (1) the monostatic scattering is no longer symmetric as a function of the angle off vertical and (2) there is an additional flat interface within the interior of the sphere which can cause enhanced scattering.

We start by assuming that the theory for scattering from a sphere with a homogeneous interior is well understood. The case of a partially filled sphere is then considered. The standard spherical harmonic series is assumed for the shell and exterior of the sphere. Another basis set (constructed from spherical harmonics) based with respect to the center of the interior air/water interface is used for the interior. These two basis sets are then coupled through the continuity equations at the shell/interior interface. Since these interface equations have an infinite number of coupled terms, they must be truncated in order to yield a numerically solvable system. Hence, unlike the standard spherical scattering problem it is not possible to write down an explicit numerical solution for the various functional coefficients in terms of simple determinants.[3,4]

From the derived set of scattering equations, numerical computations are performed showing the spectral characteristics for various amounts of fluid (we consider water) and corresponding pulse computations. These computations are done for a plane wave incident along the z axis. We conclude with a computation of the monostatic scattering strength as a function of the off-axis angle for a fixed amount of fill and fixed frequency.

II. THEORY

Let us first consider the case of a plane wave incident along the z axis so that the wavefield solution possesses azimuthal symmetry. We use the spherical coordinate system (R,θ,ϕ) where R is the three-dimensional radius, θ is the angle measured from the z axis and ϕ is the projection of the position vector in the x-y plane. Then we can write for the particle displacement $\vec{u}$ that

$$\vec{u} = -\nabla\Phi + \nabla\times\vec{\Psi}, \tag{1}$$

where Φ is a scalar compressional potential and $\vec{\Psi}$ is the vector shear potential. We will take

$$\vec{\Psi} = \nabla\times(\hat{e}_r\Psi). \tag{2}$$

Our basic building block for both Φ and Ψ is the solution to the Helmholtz equation in spherical coordinates:

$$P_n^m(\cos\theta) f_n(kR) e^{im\phi}, \tag{3}$$

where P_n^m are the Legendre polynomials, $f_n(kR)$ is either a spherical Bessel or Hankel function of order n, and k is the wave number depending upon the medium and the wave type (i.e., compressional or shear). Now for the sphere (shell and a homogeneous interior) and the exterior surrounding fluid, at a fixed azimuthal order m, the unknown potentials can be expressed as a sum over the components with the index $n \geqslant m$, the nth components having the form

$$\Phi_e = a_1^n h_n(k_e^p R) P_n^m(\cos\theta), \tag{4}$$

$$\Phi_{sh} = (a_2^n h_n(k_{sh}^p R) + a_3^n j_n(k_{sh}^p R)) P_n^m(\cos\theta), \tag{5}$$

$$\Psi_{sh} = (a_4^n h_n(k_{sh}^s R) + a_5^n j_n(k_{sh}^s R)) P_n^m(\cos\theta), \tag{6}$$

$$\Phi_{in} = a_6^n j_n(k_{in}^p R) P_n^m(\cos\theta). \tag{7}$$

Here we have used the notation that e denotes exterior, sh denotes shell and in denotes the interior. There are thus six unknown coefficients (a_i^n) and there are eight conditions to be satisfied (four at the outer and four at the inner interface), namely, for example, at the inner shell/interior interface:

$$u_r^{sh} = u_r^{in}, \tag{8}$$

$$\sigma_{rr}^{sh} = \sigma_{rr}^{in}, \tag{9}$$

$$\sigma_{r\theta}^{sh} = 0, \tag{10}$$

$$\sigma_{r\phi}^{sh} = 0. \tag{11}$$

However, some analysis shows that the fourth equation is satisfied if the third equation is satisfied; hence we only use

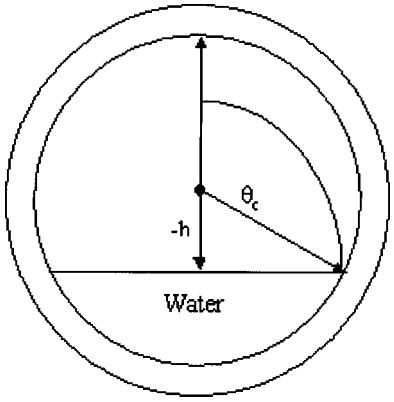


FIG. 1. A schematic of the cross-section of a partially filled shelled sphere.

the first three equations at each interface. At the shell/interior interface (and also at the exterior interface) we can write any of the equations (8)–(11) in the generic form (truncated at $n=N$),

$$\sum_{n=m}^{N} c_n P_n^m(\cos\theta) = \sum_{n=m}^{N} d_n P_n^m(\cos\theta), \tag{12}$$

where the right-hand side may, in fact, be equal to zero. In Eq. (12) the coefficients c_n and d_n represent all the nonazimuthal terms at order (m,n) lumped together, including various combinations of Bessel and Hankel functions and their derivatives and the unknown coefficients a_i^n of Eqs. (5)–(7). The equations for $\sigma_{r\theta}$ in fact involve the derivatives of the Legendre polynomials, but this condition can be integrated with respect to θ to yield the equivalent condition in terms of $P_n^m(\cos\theta)$.

For a shelled sphere with a homogeneous fluid interior, one just equates each term in Eqs. (8)–(10) at the shell/interior interface [which are of the form of Eq. (12)] to yield three of the equations of a 6×6 system for each order n (the other three equations come from the exterior water/shell interface). However, below where we consider the interior not to be homogeneous with respect to θ (i.e., there is an interior water/air interface), we will see that it is no longer possible to just simply equate terms in the series solution.

Let us now consider the interior of the sphere to be partially filled with fluid ($\theta_c<\theta<\pi$) and, as an example, consider in detail Eq. (9). For the unfilled portion we set $\sigma_{rr}=0$. For $\theta_c<\theta<\pi$ the interior is fluid with parameters c_1 and ρ_1. Within this fluid we consider an expansion based upon the center of the fluid/air interface (the coordinate $\tilde{\theta}=\pi/2$ along the interface); in particular, for the mth azimuthal order, we use the expansion for the interior, fluid compressional potential

$$\Phi_f = \sum_{n=m}^{N} a_6^n j_{q(n)}(k\tilde{R}) P_{q(n)}^m(\cos\tilde{\theta}), \tag{13}$$

where $q\equiv m+2(n-m+1)-1$ and the a_6^n are the expansion coefficients [of the form of Eq. (7)] to be determined. The integer q represents the odd indicies after the first index which is $n=m$ (for example, for $m=0$, $q=1,3,5,7$, etc.). Substituting the expression $q\equiv m+2(n-m+1)-1$ into the relation that

$$P_q^m(0)\propto\cos\left(\frac{\pi}{2}(q-m)\right), \tag{14}$$

we find that with this choice of q, $P_q^m(\cos\tilde{\theta})=0$ for $\tilde{\theta}=\pi/2$ and thus the boundary condition $\Phi_f=0$ is automatically satisfied at the fluid/air interface.

In Eq. (13) the coordinates $\tilde{R}$ and $\tilde{\theta}$ represent the radius and the polar angle based upon the origin at the center of the fluid/air interface ($z=z_I$) and are given by

$$\tilde{R}(\theta)\equiv\sqrt{R^2\sin^2\theta+(R\cos\theta-z_I)^2} \tag{15}$$

and

$$\tilde{\theta}(\theta)\equiv\cos^{-1}\left(\frac{R\cos-z_I}{\sqrt{R^2\sin^2\theta+(R\cos\theta-z_I)^2}}\right), \tag{16}$$

where $R\equiv a$ along the shell/interior interface.

Now consider the equation for σ_{rr} at the inner interface. We have that

$$0=\sum_{n=m}^{N}(t(n)_1+t(n)_2+t(n)_2+t(n)_3+t(n)_4)P_n^m(\cos\theta), \tag{17}$$

$$\sum_{n=m}^{N} -\rho_1\omega^2 a_6^n j_{q(n)}(k\tilde{R})P_{q(n)}^m(\cos\tilde{\theta}) = \sum_{n=m}^{N}(t(n)_1+t(n)_2+t(n)_3+t(n)_4)P_n^m(\cos\theta), \tag{18}$$

where Eq. (17) is valid in the angular range $0<\theta<\theta_c$ and Eq. (18) in the range $\theta_c<\theta<\pi$. The functions $t(n)_i$ represent the terms involving the shear and compressional potentials in the shell for the spherical Bessel functions j_n and h_n. We now multiply Eq. (18) by $\sin(\theta)P_\nu^m(\cos\theta)n_\nu$, where n_ν is a normalizing factor, and integrate with respect to θ to obtain

$$(t(\nu)_1+t(\nu)_2+t(\nu)_3+t(\nu)_4)=\sum_{n=m} a_6^n S_{\nu n}, \tag{19}$$

where

$$S_{\nu n}\equiv -\rho_1\omega^2 n_\nu\int_{\theta_c}^{\pi} P_\nu^m(\cos\theta)j_{q(n)}(k\tilde{R}(\theta)) \times P_{q(n)}^m(\cos\tilde{\theta}(\theta))\sin\theta d\theta. \tag{20}$$

We will not list all the coefficients $t(\nu)_i$, their basic form can be found in the elastic equations for a shelled sphere (or as entries in the determinant of the system of equations) (see, for example, Refs. 3 and 4), but as an example we have

$$t(\nu)_2=\left[-2\nu\frac{(\nu+1)}{R}\left(-\frac{1}{R}j_\nu(k_{sh}^s R)+j_\nu'(k_{sh}^s R)\right)\right]a_5^\nu, \tag{21}$$

where R is evaluated at $R=a$ and a_5^ν is the coefficient in Eq. (6).

Similarly, for the radial displacement u_R we have from Eq. (8)

$$\sum_{n=m}^{N}(s(n)_1+s(n)_2+s(n)_3+s(n)_4)P_n^m(\cos\theta)$$

$$=\sum_{n=m}^{N}\frac{\partial}{\partial R}(j_{q(n)}(k\tilde{R})P_{q(n)}^m(\cos\tilde{\theta})),$$

$$\theta_c<\theta<\pi, \tag{22}$$

where $s(n)_i$ are the expressions for the four shell potential terms. Our approach is to multiply both sides of Eq. (22) by $P_k^m(\cos\tilde{\theta}(\theta))\sin(\tilde{\theta}(\theta))\partial\tilde{\theta}/\partial\theta$ and integrate with respect to θ where k corresponds to one of the possible values of q. Then we obtain

$$\sum_{n=m}^{N}(s(n)_1+s(n)_2+s(n)_3+s(n)_4)T_{\nu n}^m=\sum_{n=m}^{N}Q_{\nu n}^m, \tag{23}$$

where

$$T_{\nu n}\equiv\int_{\theta_c}^{\pi}P_{q(\nu)}^m(\cos\tilde{\theta}(\theta))P_n^m(\cos\theta)\sin\tilde{\theta}(\theta)\frac{d\tilde{\theta}}{d\theta}d\theta, \tag{24}$$

$$Q_{\nu n}\equiv\int_{\theta_c}^{\pi}P_{q(\nu)}^m(\cos\tilde{\theta}(\theta))\frac{\partial}{\partial R}[P_{q(n)}^m(\cos\tilde{\theta}(\theta))$$

$$\times j_{q(n)}(k\tilde{R})]\sin\tilde{\theta}(\theta)\frac{d\tilde{\theta}}{d\theta}d\theta. \tag{25}$$

Once again the general form of the coefficients can be found in Refs. 3 and 4, but as an example:

$$s(\nu)_2=\left[-\frac{1}{R}\nu(\nu+1)j_\nu(k_{sh}^sR)\right]a_5^\nu, \tag{26}$$

where R is evaluated at $R=a$.

In summary, for a fixed azimuthal order m, we take N_{in} unknown coefficients for the interior solution and N coefficients for each of the shell potentials and the exterior potential, obtaining a coupled $(5N+N_{in})\times(5N+N_{in})$ system of equations, with two of the continuity conditions at the inner interface providing the coupling mechanism. The condition Eq. (10) reduces to the standard noncoupled equation for the shell coefficients.

We now discuss an efficient method to compute the coupling integrals; the integrals T_{kn} are computed numerically and are independent of frequency. Thus they need to be computed only once if one is doing multiple frequency computations. The other integrals, S_{kn} and Q_{kn}, do depend upon frequency through the term $j_q(k\tilde{R})$ and thus it is advantageous to reexpress these integrands in terms of azimuthal (θ) terms which are frequency independent and frequency terms which do not depend upon θ; then the various azimuthal integrals need to be computed only once for multi-frequency computations. In order to reexpress these integrals we use the expressions for the translation of origins in spherical coordinate systems,[5,6]

$$j_q(k\tilde{R})P_q^m(\cos\tilde{\theta})=\sum_{n=m}^{\infty}C_{q,n}^m j_n(kR)P_n^m(\cos\theta), \tag{27}$$

where

$$C_{q,n}^m=(2n+1)\sum_{p=|q-n|}^{q+n}w_p(2p+1)(-i)^p j_p(kh)$$

$$\times i^{q-n}(-1)^m, \tag{28}$$

h is the vertical separation between the two origins and w_p is the product of Wigner-3j coefficients,

$$w_p=\begin{pmatrix}q & n & p\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}q & n & p\\ m & -m & 0\end{pmatrix}\sqrt{\frac{(q-m)!(n+m)!}{(q+m)!(n-m)!}}. \tag{29}$$

References 5 and 6 use this type of representation for scattering from a pair of spheres; Lim[7] has also used the Wigner-3j representation to simplify his integrals for scattering in a multilayered acoustic medium. If the vertical offset of the center of the sphere from the origin is negative, then the result of Eq. (28) is multiplied by $(-1)^{q+n}$. Using these relationships in the definitions of Eq. (25) the various angular projection integrals can be expressed in terms of integrals which are independent of frequency. Thus, once again, these integrals are computed once and then repeatedly used in multiple-frequency computations.

III. NUMERICAL EXAMPLES

For the first example, a plane wave is incident on a sphere of radius 0.25 m with a 6-mm steel shell. The plane wave travels vertically along the z axis. Due to the azimuthal symmetry of the problem it is only necessary to compute the $m=0$ term of the solution. The level of the water fill in the sphere is varied as a fraction of the inner radius $R_a=0.244$ m. The plane wave is incident from the bottom so that it is initially incident upon the water in the sphere before encountering the water/air interface. Numerically, we investigated using different numbers of spherical harmonics in the computations. For the exterior and shell harmonics, the formula $N=1.5ka$ was used with a minimum number of eight harmonics used. This means that until approximately 5 kHz, 8 harmonics are used and this is increased linearly to 15 harmonics at the frequency of 10 kHz. For the interior basis functions it was found that for negative values of h (for those cases where the fraction of water fill is less than one-half) it was advantageous to use $N_{in}=N/2$ functions (one could use $N_{in}=N$, but for some frequencies there seemed to be numerically unstable results due to poor conditioning of the system of equations). It makes physical sense that less basis functions should be used for the case of small fractional fill since in this case the length of the water/shell boundary is relatively small. For values of fill greater or equal to one-half, the full number N was used (at a fraction of one-half, essentially the same results are obtained using either $N_{in}=N$ or $N_{in}=N/2$). As a check of the numerical accuracy of the method, we can consider the case of $h=R_a$ in which case the sphere is completely full of water. In Fig. 2 we show the resulting spectra as computed using the coupled-mode approach (solid) and using a standard harmonic expansion[3,4] for a fluid-filled sphere (dashed). In this computation a receiver is located in the backscatter direction, 10 m from the

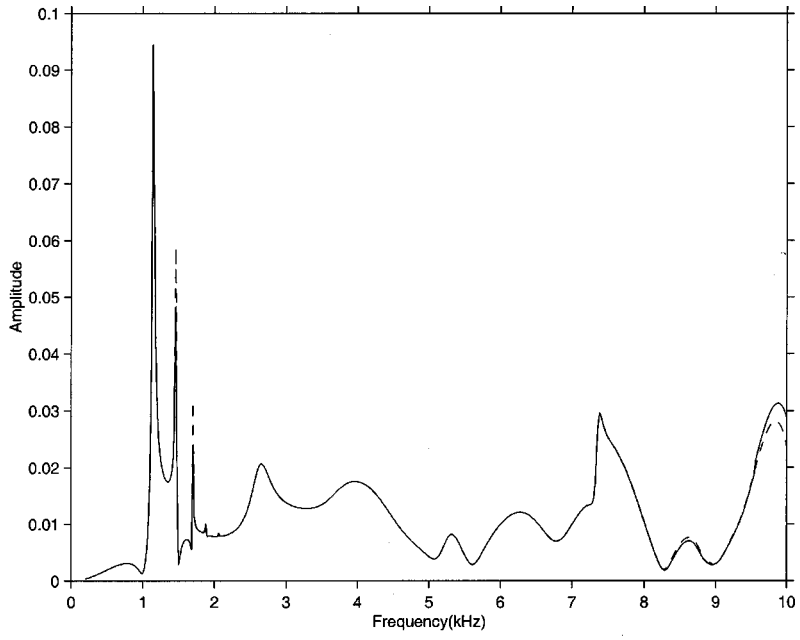


FIG. 2. A comparison between the exact (dashed) and computed (solid) spectra for the totally water-filled sphere.

center of the sphere. As can be seen the agreement is excellent. This is rather remarkable as the interior expansion for the coupled-expansion method is based upon an origin at the very top of the sphere. In Fig. 3 we show the resulting spectra for the backscattered energy starting with $h=-0.75R_a$ and ending with $h=0.75R_a$ in steps of $0.25R_a$. The first curve corresponds to the interior being almost empty and the last case to it being almost full. As the fill approaches the one-half fraction, it can be seen that the spectral levels for the higher frequencies increases significantly; this is due to the growing size of the water/air interface within the sphere. There is also a shifting of the resonance peaks in the 1–2 kHz range. At the $h=0.75R_a$ value a resonance peak can be seen at about $f=800$ Hz; although not shown in Fig. 3 this feature can be tracked for fill values less than this (e.g., $h=0.6R_a$ and starts to disappear as the sphere is filled more).

In Fig. 4 the corresponding time series for an incident 4-kHz Ricker pulse are shown. The incident pulse is still a plane wave but with a Ricker pulse shape in the time domain. The time series $p(t)$ is computed from the multi-frequency spectrum $S(\omega)$ using Fourier synthesis; in particular,

$$p(t)=2\Re\left\{\sum_{\omega_k} S(\omega)R(\omega)\Delta\omega e^{i\omega_k t}\right\}, \tag{30}$$

and $R(\omega)$ is the source spectrum,

$$R(\omega)\equiv\omega^2\exp(-\omega^2/\omega_c^2)\sqrt{\frac{2\pi}{\omega_c}}, \tag{31}$$

where in our example $\omega_c=2\pi 4000$. An important feature in Fig. 4 is that the strongest reflection comes from energy propagating in the sphere, hitting the water/air interface and

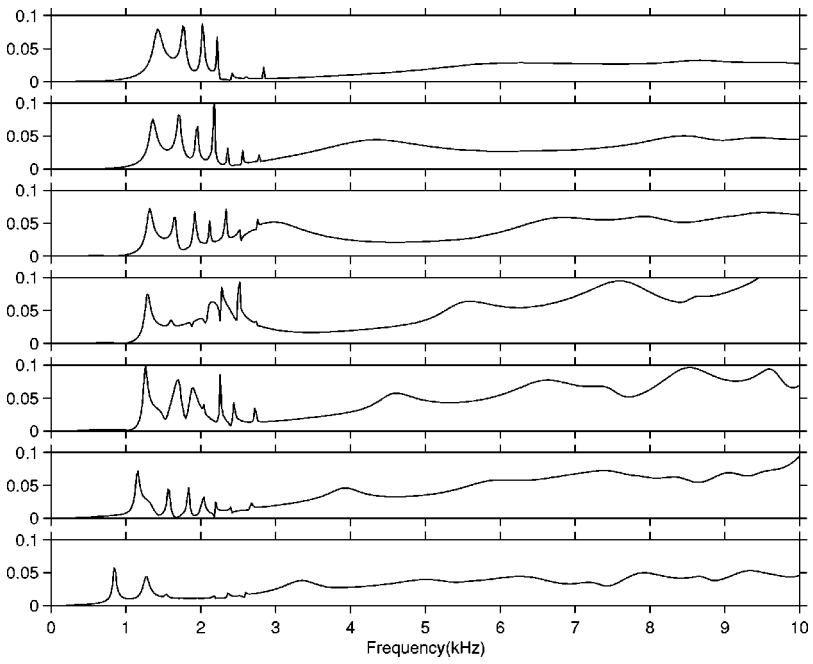


FIG. 3. The backscattered spectra as a function of fluid-fill: top curve is $h=-0.75R_a$ and bottom curve is $h=0.75R_a$ (varying from nearly empty to nearly full in steps of $0.25R_a$). The wave is incident from below (i.e., strikes water interior first).

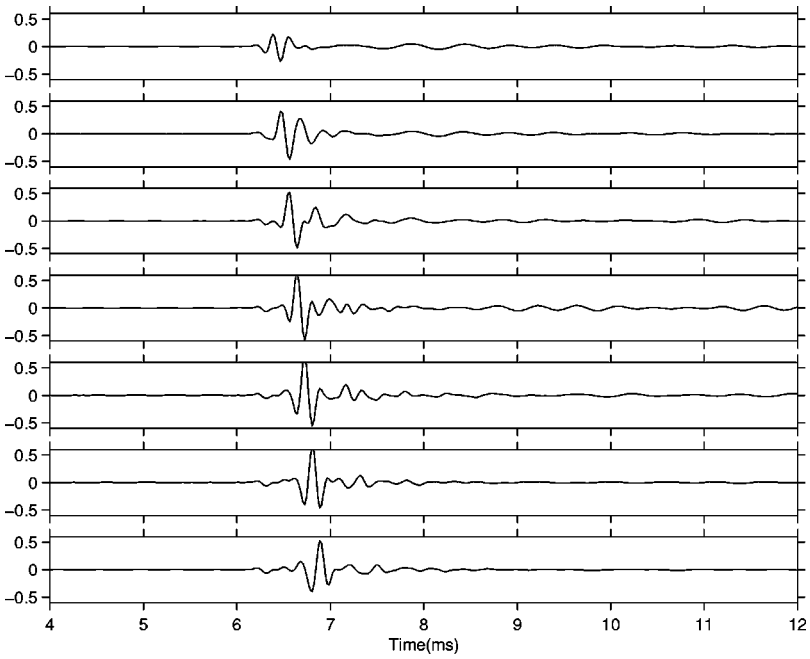


FIG. 4. The time series corresponding to the spectra of Fig. 3 for an incident 4-kHz Ricker pulse.

reflecting out. The amplitude of this reflection grows until $h=0$ or $h=0.25R_a$. This is reasonable as in this case the surface area of the water/air interface increases until $h=0$.

These computations are now repeated but for the plane wave incident from above, thus encountering the air-backed shell first. This is significantly different from the previous case in the sense that the incident wave cannot initially propagate in the interior of the sphere. The spectra for the various levels of fill and the corresponding time series are shown in Figs. 5 and 6. The top curves are for little fill ($h=-0.75R_a$) or equivalently for a large fraction of air. There is a noticeable variation in the spectra and time series as the amount of fill is changed, but there is no very large reflection from the water/air interface within the sphere as there was for the first set of curves (Figs. 2 and 3).

Finally, we consider a fixed frequency of 4 kHz, a fixed amount of fill ($h=0$ or the sphere is half-filled) and the monostatic scattering angle is varied; in this case it is necessary to compute the solution for varying values of m. For a sphere, totally fluid-filled or totally air-filled the scattering plot should be independent of the polar angle. However, as can be seen in Fig. 7, the polar plot for the half-filled case has a strong angular dependence.

IV. SUMMARY

It has been shown that the acoustic field scattered by a partially filled elastic-shelled sphere can be computed using a dual set of basis functions for the interior of the sphere and the shell and exterior of the sphere. The computations

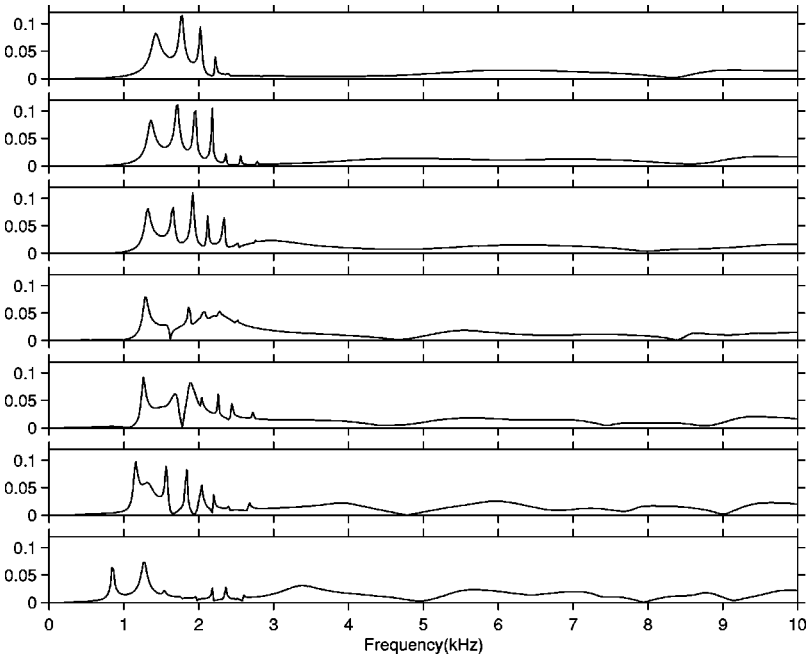


FIG. 5. The backscattered spectra as a function of fluid-fill: top curve is $h=-0.75R_a$ and bottom curve is $h=0.75R_a$ (varying from nearly empty to nearly full in steps of $0.25R_a$). The wave is incident from above (i.e., strikes air interior first).

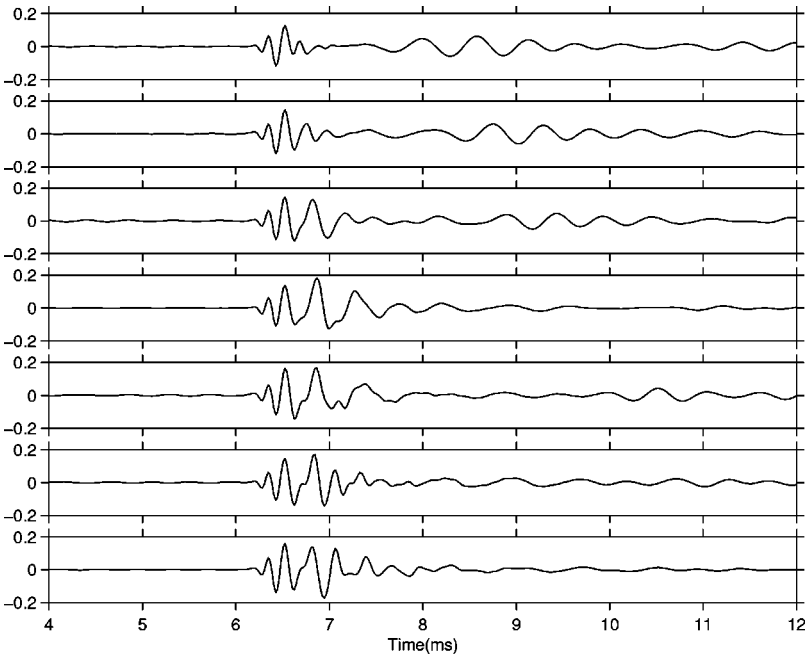


FIG. 6. The time series corresponding to the spectra of Fig. 5 for an incident 4-kHz Ricker pulse.

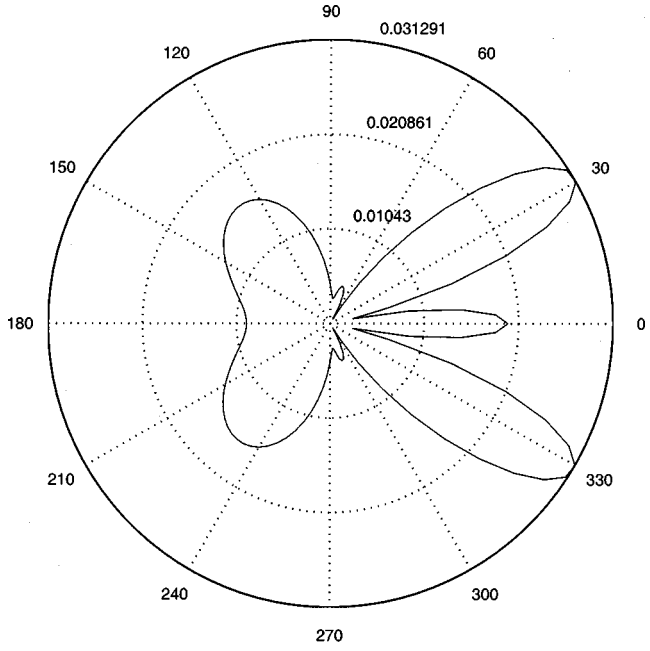


FIG. 7. A polar plot of monostatic backscattering level (receiver is 10 m from the sphere center) as a function of polar angle for a frequency of 4 kHz and the sphere is half-filled.

showed that partial fill could increase the target strength of the sphere for higher frequencies as the interior fluid/air interface became an increasingly strong scattering surface for a plane wave incident upon the fluid-fill. When the plane wave is incident upon the air-filled side, this effect is not so evident as the incident wave cannot propagate through the first part of the interior. The computations also showed how partial fill of the sphere makes the target scattering strength significantly angularly dependent.

[1] R. Hickling, ''Analysis of echos from a hollow metallic sphere in water,'' J. Acoust. Soc. Am. **36**, 1124–1137 (1964).
[2] G. Kaduchak and C. Loeffler, ''Relationship between material properties and target strength of fluid-filled spherical shells in water: calculations and observations,'' IEEE J. Ocean Eng. **23**, 26–30 (1998).
[3] G. C. Gaunaurd and M. F. Werby, ''Lamb and creeping waves around submerged spherical shells resonantly excited by sound scattering,'' J. Acoust. Soc. Am. **82**, 2021–2033 (1987).
[4] G. C. Gaunaurd and M. F. Werby, ''Sound scattering by resonantly excited, fluid-loaded, elastic spherical shells,'' J. Acoust. Soc. Am. **90**, 2536–2550 (1991).
[5] G. C. Gaunaurd, H. Huang, and H. C. Strifors, ''Acoustic scattering by a pair of spheres,'' J. Acoust. Soc. Am. **98**, 495–507 (1995).
[6] J. Bruning and Y. Lo, ''Multiple scattering of electromagnetic waves by spheres. Pt I. Multipole expansion an ray-optical solutions,'' IEEE Trans. Antennas Propag. **AP-19**, 378–390 (1971).
[7] R. Lim, ''Multiple scattering by many bounded obstacles in a multilayered acoustic medium,'' J. Acoust. Soc. Am. **92**, 1593–1612 (1992).

An evaluation of the accuracy of shallow water matched field inversion results[a)]

Mirjam Snellen[b)] and Dick G. Simons
TNO Physics and Electronics Laboratory, Underwater Acoustics Group, Oude Waalsdorperweg 63, 2509 JG The Hague, The Netherlands

Martin Siderius, Jürgen Sellschopp, and Peter L. Nielsen
SACLANT Undersea Research Centre, Viale S. Bartolomeo 400, 19138 La Spezia, Italy

(Received 24 April 2000; revised 24 October 2000; accepted 15 November 2000)

In this article the accuracy of geo-acoustic and geometric parameter estimates obtained through matched field inversion (MFI) was assessed. Multi-frequency MFI was applied to multi-tone data (200–600 Hz) received at a 2-km source/receiver range. The acoustic source was fixed and the signals were received at a vertical array. Simultaneously with the acoustic transmissions, a CTD (conductivity, temperature and depth)-chain was towed along the acoustic track. A genetic algorithm was used for the global optimization, whereas a normal mode model was applied for the forward acoustic calculations. Acoustic data received at consecutive times were inverted and the stability of the inverted parameters was determined. Also, the parameter estimates were compared with independent measurements, such as multi-channel seismic surveys (for geo-acoustic parameters). The obtained uncertainty in the inversion results was assumed to have two distinct origins. The first origin is the inversion method itself, since each optimization will come up with some solution close to the exact optimum. Parameter coupling and the fact that some parameters hardly influence the acoustic propagation further contribute to this uncertainty. The second is due to oceanographic variability. Both contributions were evaluated through simulation. The contribution of oceanographic variability was evaluated through synthetic inversions that account for the actual sound speed variations as measured by the towed CTD-chain. [DOI: 10.1121/1.1339828]

PACS numbers: 43.30.Pc, 43.60.Pt, 43.30.Ma [DLB]

I. INTRODUCTION

Matched field inversion (MFI) is a technique for obtaining information on unknown parameters that influence the propagation of sound underwater. When employing MFI, a measured acoustic field is compared with acoustic fields that are calculated by a propagation model for many sets of unknown parameters. Since the number of possible parameter combinations is huge and there are many local optima, global optimization methods are needed for guiding the search for the set of unknown parameters that gives the optimum match between measured and calculated acoustic fields. An important application of MFI is geo-acoustic seabed parameter estimation.

An important issue to be addressed is the accuracy of the parameter estimates. In general, performing several independent inversions for the same acoustic field can result in different estimates for the unknown parameters. This uncertainty in parameter estimates is partly due to the optimization method, as global optimization methods often do not determine the exact optimum, but a solution close to it. The resulting variation in parameter estimates is dependent on their influence on the acoustic propagation. For the parameters that hardly influence the acoustic propagation the variation in parameter estimates will comprise a large part of the parameter search space, whereas for the parameters that have a strong influence on the acoustic propagation the variation will be small. Also, the fact that there can be correlations between the unknown parameters contributes to the uncertainty. These two above-mentioned factors can result in optimized parameter values that deviate from the actual parameter values. However, the agreement between optimized and measured acoustic fields can still be high. By adapting the optimization method the uncertainties due to the above-mentioned factors can be reduced. For example, a local search method can be applied after convergence of the global search.[1] In the following we will denote the uncertainties in parameter estimates that result from these two mechanisms as the uncertainties caused by the method itself.

When inverting for unknown parameters and using experimental acoustic fields that were measured at different times, the temporal variability of the oceanographic conditions can result in additional uncertainties in parameter estimates.[2] If the sound speed profile in the water column varies with time and if one does not account for these variations in the inverse modeling, this can result in parameter estimates that also vary with time. In this case, the optimized parameter value is such that it corrects for the difference between the actual sound speed profile and the sound speed profile that is used for the forward model calculations. The effects of the varying oceanographic conditions can directly

[a)]Portions of the work in this paper were presented in ''An evaluation of the accuracy of shallow water broadband inversion results,'' Proceedings of the fifth European Conference on Underwater Acoustics, France, Lyon, 2000.

[b)]Electronic mail: snellen@fel.tno.nl

be seen in the acoustic data, as these variations result in changes in the propagation conditions and consequently varying (in time) received signals. This contribution to the uncertainty can, at least in principle, be eliminated when the sound speed structure between source and receiver is known exactly at the time of each transmission. However, under practical experimental conditions this is rather cumbersome.

Matched field inversion results are presented for experimental data that are obtained during the ADVENT99 sea trial. This sea trial was conducted by SACLANT Centre and TNO-FEL on the Adventure Bank, South of Sicily, in April/May 1999.[3] Several experiments were carried out during ADVENT99. The analysis presented in this article deals with data that were acquired during experiments with both the source and the receiver at a fixed position. The goal of these experiments was to obtain data for a fixed geometry over an extensive period of time. During these experiments many measurements were also carried out to obtain information on the ocean environment. These environmental measurements included conductivity, temperature and depth (CTD) casts and waverider measurements. Further, a CTD-chain was towed along the acoustic track, giving more detailed information on the sound speed structure. Also, an extensive seismic survey was carried out, thereby obtaining independent geophysical information on the seabed. For the survey use was made of a multi-channel streamer, which allows for the estimation of, in addition to the ocean bottom layering, the sound speeds of the layers.

The main goal of this article is to assess the uncertainties in parameter estimates obtained through MFI using experimental data. The uncertainty that is due to the method itself is estimated by simulation, i.e., by performing inversions of synthetic data, and it is demonstrated that for some parameters the uncertainty can thus largely be explained. The CTD-chain data are used to show that the remaining uncertainties originate from oceanographic variations in the water column.

This article is organized as follows: in Sec. II the ADVENT99 sea trial is presented and an overview of the acoustic and environmental data is provided. The inversion problem, comprising the forward acoustic model and the optimization method, is described in Sec. III. In Sec. IV the inversion results are presented and the parameter estimates and their uncertainties assessed.

II. THE ADVENT99 EXPERIMENT

A. Acoustic measurements

A large part of the ADVENT99 sea trial comprised acoustic experiments with both the source and the receiver at a fixed position. These fixed geometry experiments were conducted for source/receiver ranges of 2, 5, and 10 km.

In this article we only consider data of the 2-km experiment. The position of the source during this experiment was 37° 17.966′ N, 12° 15.588′ E. The receiving system was positioned at 37° 17.883′ N, 12° 14.207′ E. This experiment took place on 2 May 1999 from 12:37 to 20:17 UTC time.

The source used for the acoustic transmissions was mounted on a tower that was moored on the sea bottom for keeping it at a fixed position. The Nato Research Vessel (NRV) ALLIANCE was connected to the source by a power supply cable and therefore had to remain close to the source, but at a sufficient distance (a few hundred meters) to reduce risk of damage. Consequently, CTD casts could not be carried out very close to the source. The receiving system consisted of a vertical array (VA), containing 64 elements and spanning 62 m of the water column. The signals received on the VA were sent directly to the data acquisition system on board NRV ALLIANCE by radio link. In Fig. 1 a schematic of the experimental configuration is shown.

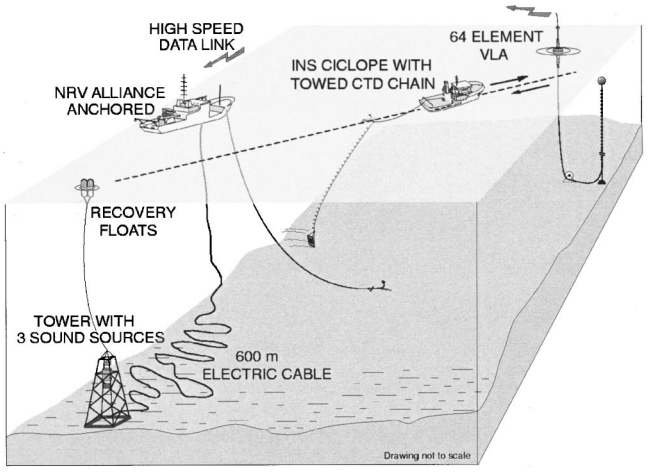


FIG. 1. The configuration of the fixed geometry experiments.

Both low-frequency (200–700 Hz) and high-frequency (800–1600 Hz) multi-tones and linear frequency modulated (LFM) sweeps were transmitted. In this article only the low-frequency multi-tones are considered. Snapshots of 2 s were selected from the received time series and were fast Fourier transformed into the frequency domain. The resulting complex pressures as a function of depth are further referred to as ''pressure fields.'' The magnitudes of 41 pressure fields are displayed in Fig. 2 for the frequencies 200, 300, 400, and 600 Hz. These 41 pressure fields were used in the inversions. The corresponding data were transmitted at 15-min intervals,

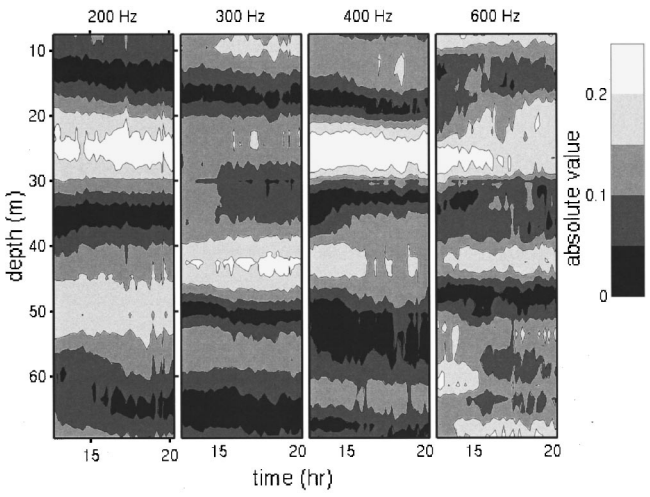


FIG. 2. Absolute values of the measured pressure fields as a function of UTC time for the frequencies 200, 300, 400, and 600 Hz.

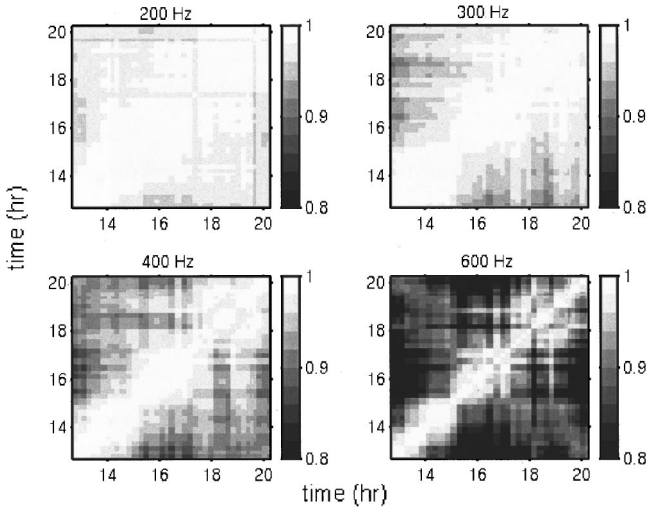


FIG. 3. The quantitative measure M_{ij} [Eq. (1)] for the variability in received signals for the four frequencies.

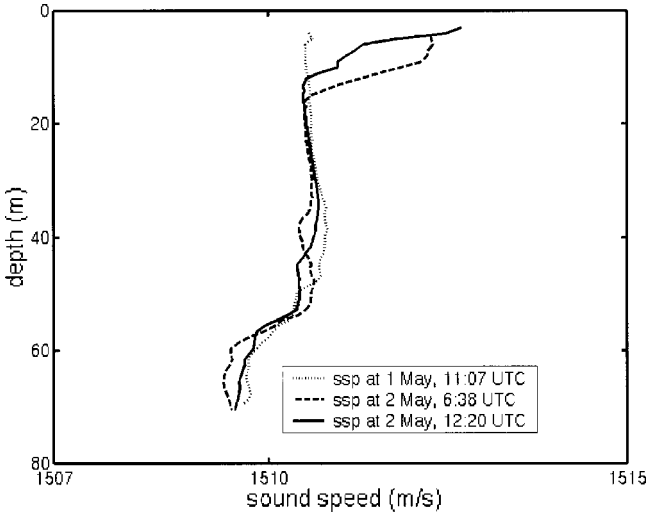


FIG. 4. Sound speed profiles as derived from the CTD casts from NRV ALLIANCE.

spanning the total duration of the 2-km experiment (about 8 h). The figure clearly shows the variability in the received acoustic fields. The set of data considered in this article is thus suitable for the purpose of investigating the contribution of water column variability to uncertainties in the parameter estimates. Note that the variability in the received signals increases with increasing frequency.

Also illustrative is to determine a quantitative measure for the variability in the received signals. For this quantitative measure M_{ij} we have used the correlation between pressure fields $\vec{p}_i$ and $\vec{p}_j$

$$M_{ij}=|\vec{p}_i\cdot\vec{p}_j^{*}| \tag{1}$$

with ''$\cdot$'' indicating the inner product and ''*'' denoting the complex conjugate. The pressure fields $\vec{p}_i$ and $\vec{p}_j$ are normalized such that their norm equals one, i.e., $\|\vec{p}_i\|=\|\vec{p}_j\|=1$. Figure 3 shows, for the four frequencies, the correlation for all possible combinations of received pressure fields. Note that along the diagonal a pressure field is correlated with itself, resulting in a value of one for the correlation. Moving away from the diagonal shows, on the whole, a decrease in correlation, as moving away from the diagonal corresponds to an increase in time span between two transmissions. As with Fig. 2, Fig. 3 also clearly shows increasing variability with increasing frequency. Since the signal-to-noise ratio was high (at least 20 dB), the structure seen in Fig. 3 at 300 Hz and higher must be due to variations in oceanographic conditions and possible variations in the tilt of the vertical array (see Sec. IV A 2).

B. Ocean environmental measurements

1. CTD measurements

From the NRV ALLIANCE a few CTD casts were taken. Figure 4 shows the corresponding sound speed profiles (ssp's). From this figure it can be seen that there is a minor sound speed variation over the water depth (1509–1513 m/s). The ssp plotted as a solid line corresponds to the CTD taken 17 min prior to the 2-km fixed geometry experiment. No further CTD casts were carried out from ALLIANCE during this experiment.

2. CTD-chain measurements

During the acoustic experiments a CTD-chain was towed back and forth over the acoustic track by the Italian Navy Ship (INS) *Ciclope* (see Fig. 1). From the CTD-chain measurements the sound speed as a function of depth in the water column was determined along the acoustic track. The tracks along which the CTD-chain measurements were done have a length of about 8.8 km. Figure 5 presents the water sound speeds, calculated from the CTD-chain measurements that were carried out during the time slot of the 2-km fixed geometry experiment. In order to provide a complete impression of the oceanographic variability encountered during the experiment, we present the CTD-chain data for the full 8.8-km track. The 2-km acoustic track is indicated in the figures by the vertical solid lines. As the tow ship is moving, both time and position are different for succeeding CTD-chain measurements. Hence, these figures do not represent time frozen sound speed structures that can be used directly as input to a (range-dependent) acoustic model. They are indicative of the amount of sound speed variability in the water column. Sound speed realizations will be selected from these CTD-chain data to simulate the effect of this oceanographic variability on inversion results (see Sec. IV B).

3. Seismic measurements

A seismic survey was carried out at the experimental site. In total five tracks were covered in the area around the acoustic track. Figure 6 shows the tracks that were sailed during the survey. As can be seen, the seismic survey covers the full acoustic track (up to 10-km source/receiver distance). The survey was carried out prior to the acoustic measurements and covers an area of 12×2 km^2. Within this area we have selected the acoustic track such that the range dependence in both bathymetry and bottom properties (as derived from the seismic measurements) is minimal. This is because

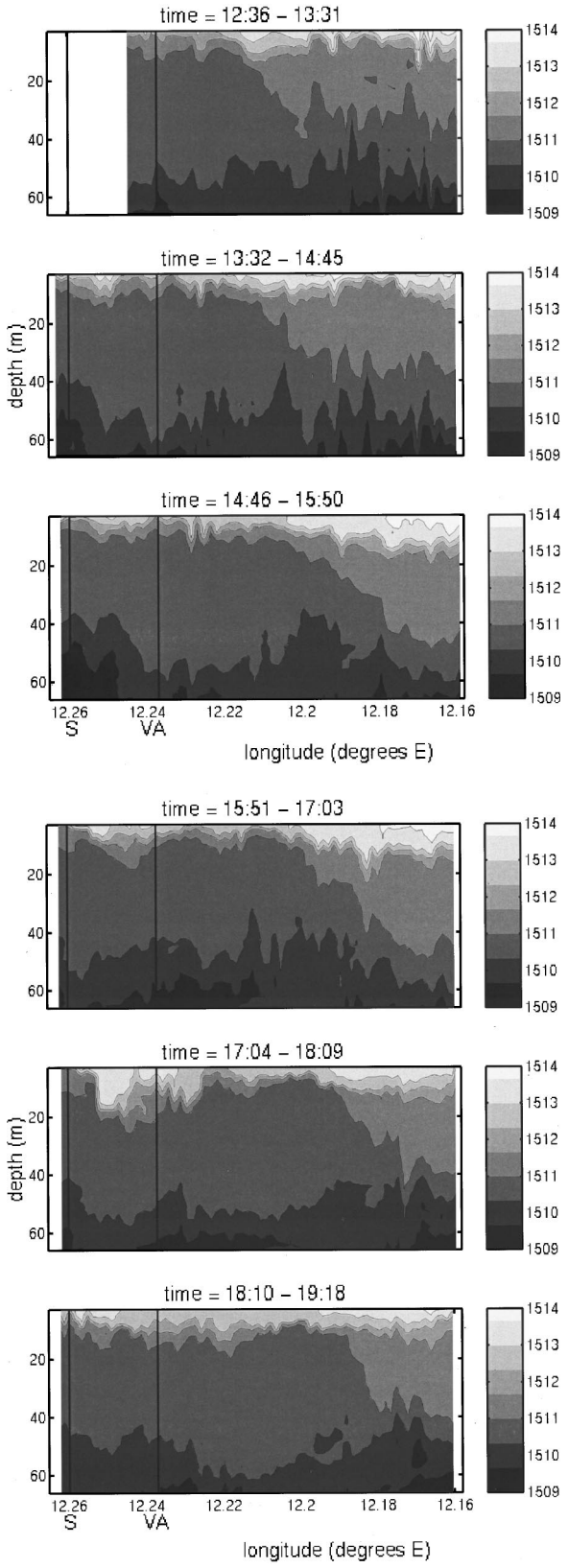


FIG. 5. Sound speed (in m/s) as a function of longitude and time as derived from the CTD-chain measurements made during the 2-km fixed geometry experiment. The vertical lines indicate the position of the source (S) and the vertical array (VA) for this experiment. The total horizontal axis corresponds to about 8.8 km.

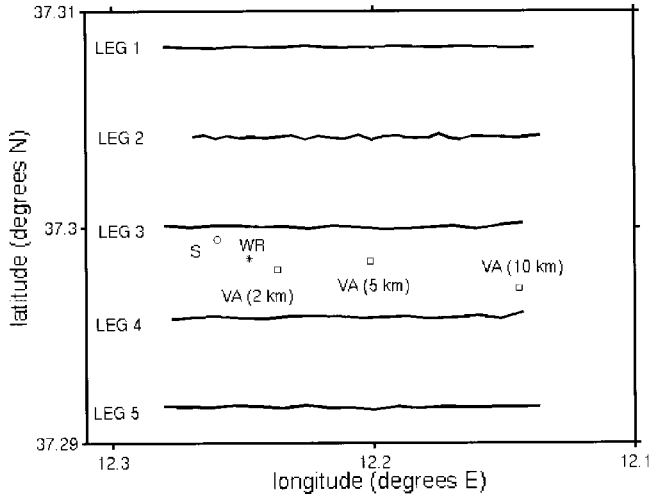


FIG. 6. The tracks sailed during the seismic survey. Note that the vertical distance between the five legs is only 0.5 km. The positions of the vertical array (VA) are denoted by squares, the source tower (S) position by a circle, and a star denotes the waverider (WR) position.

the main scientific issue of this article is to assess the influence of variability in the water column on the geo-acoustic parameter estimates. In the figure the position of the VA for the different experiments, the source position, and the position of the waverider are also indicated.

Use was made of a boomer-type sound source and the signals were received on a multi-channel seismic streamer. The streamer comprised eight groups of hydrophones, i.e., eight channels, and had a total length of 32 m, a group interval of 4 m, and a group length of 3.8 m. Using a multichannel receiving array allows for estimating both the layering of the seabed and the sound speeds of the layers.[4] This is a classical seismic technique, which is based on differences in travel time from reflectors in the bottom towards the various channels in the streamer.[5]

Figure 7 shows the layering on parts of leg 3 and leg 4 as obtained from the seismic measurements. Some internal structures within the sedimentary cover can be seen. This structure consists of a series of subhorizontal layers, i.e., the sedimentary deposits are at a small angle with the horizontal. In these seismic data there is no clear definition of the basement.[6]

When estimating sound speeds from the seismic data use is made of reflections that correspond to a strong reflector. If there are several strong reflectors in the bottom, then the sound speed as a function of depth can be estimated.[5] However, with the seismic data presented here, only one strong reflector could be identified for the majority of the data. Therefore, we only consider layer thicknesses and sound speeds that are estimated for this one reflector. This reflector does not necessarily correspond to the upper layer. Note that the estimated sound speeds represent a value averaged over the layer thickness.

Figures 8 and 9 show the thin layer thicknesses and corresponding thin layer sound speeds as a function of geo-

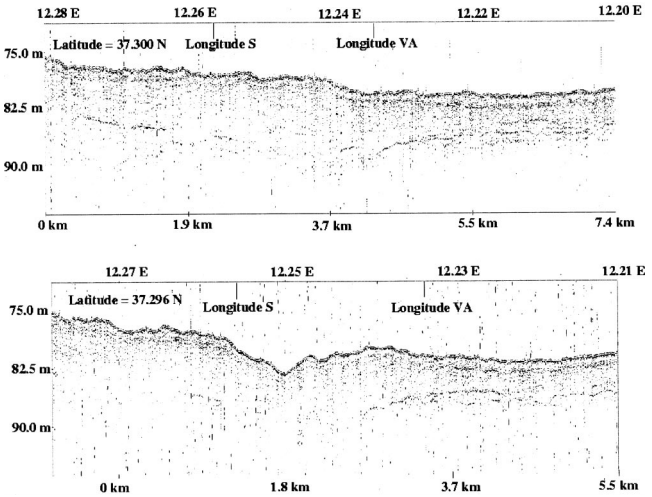


FIG. 7. Bottom layering as derived from seismic measurements on part of leg 3 (upper plot) and part of leg 4 (lower plot). The variable along the *y* axis is depth in meters, whereas position is along the *x* axis. For reference the longitudes of the source (S) and vertical array (VA) positions are also indicated.

graphical position for the full area of the seismic survey. The results for the full area are presented here as then possible systematic changes in bottom properties can be seen. However, no trends in thin layer thickness are visible. In the area surrounding the source and the vertical array, the thin layer thickness ranges from ~2 to ~10 m, i.e., 6 ± 4 m.

As with the thin layer thickness, no trends can be identified in the thin layer sound speeds (see Fig. 9), except for the somewhat higher sound speeds (>1620 m/s) at the right side of the plot. Ignoring these high sound speeds, the values range from ~1500 to ~1600 m/s. Since there is no trend in these sound speeds, we have concluded that this variation is due to errors that are inherent to the seismic method. A histogram of the sound speed estimates is given in Fig. 10, thereby treating the sound speeds as independent observations of the same quantity. The histogram contains all sound speed estimates except those higher than 1620 m/s. A Gaussian fit is applied to these data. The mean and standard deviation of the Gaussian curve is 1552 and 31 m/s, respectively.

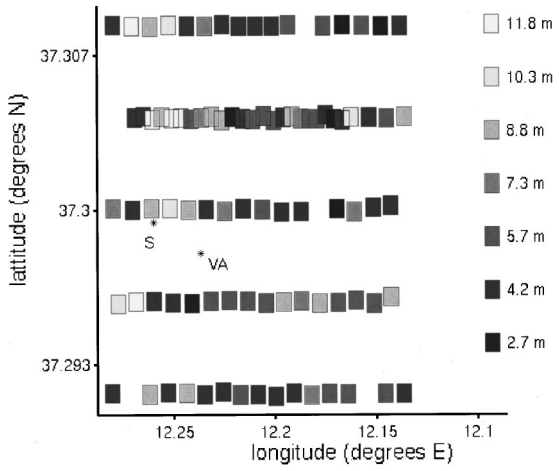


FIG. 8. Thin layer thickness as a function of geographical position. Stars are used for indicating the source (S) and vertical array (VA) position.

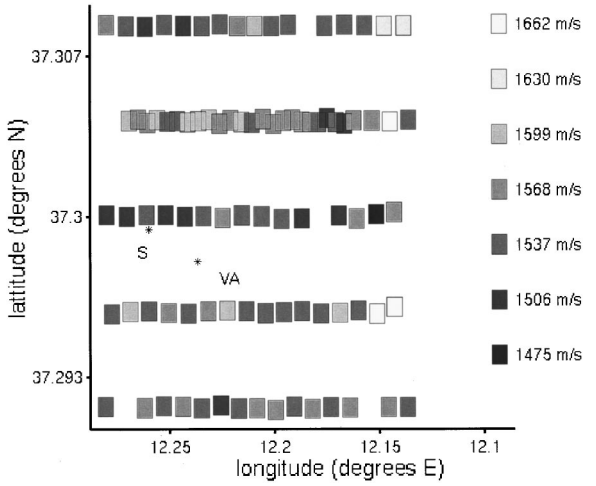


FIG. 9. Thin layer sound speed as a function of geographical position. Stars are used for indicating source (S) and vertical array (VA) positions.

4. Additional ocean environmental measurements

In addition to the environmental measurements described above, bathymetry measurements, waverider measurements, and acoustic Doppler current profiler (ADCP) measurements were also carried out.

The bathymetry of the trial area was measured with the echosounder of NRV ALLIANCE. There was virtually no variation in water depth along the 2-km track with a depth of 77 m at the source position and 79 m at the VA position.

The sea surface wave height spectrum was measured by a waverider. The root-mean-squared wave height during the 2-km experiment was 0.15 m.

The ADCP measurements were collected by two bottom-moored ADCPs and a ship mounted ADCP. The bottom-moored ADCPs were put close to the positions of the VA. Tidal oscillations are found to be the dominant current signals. The ADCP current data are used for validation of the VA tilt estimates (see Sec. IV A 2).

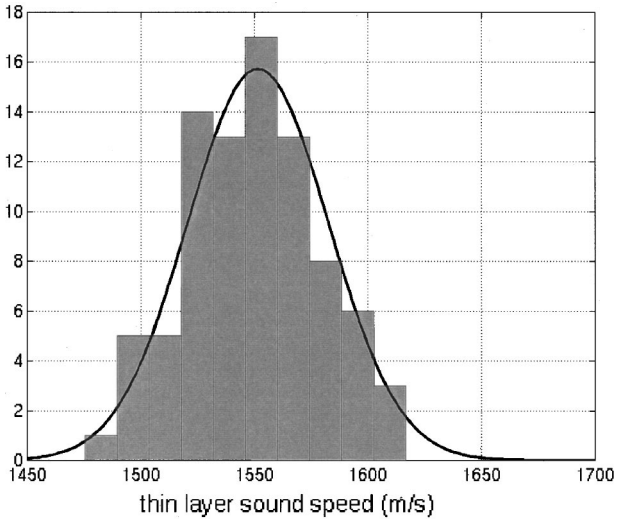


FIG. 10. A histogram of the thin layer sound speeds as estimated from the multi-channel seismic analysis. The black line indicates a Gaussian fit to the histogram.

III. ACOUSTIC INVERSION METHOD

For the forward acoustic model we have applied the standard normal mode technique.[7,8] The sediment layer and the subbottom are treated as fluid layers and the high-loss continuous eigenvalue spectrum is ignored. In Sec. III A the objective function to be minimized is described. In Sec. III B we present details on the applied global optimization method and in Sec. III C we describe the acoustic problem.

A. The objective function

The objective (or energy) function gives a quantitative measure for the agreement between the calculated and measured acoustic fields. We have selected the following objective function E, which is based on the incoherent multi-frequency Bartlett processor[8,9]

$$E(\vec{m}) = 1 - \frac{1}{K}\sum_{k=1}^{K} |\vec{p}(f_k)\cdot\vec{p}_c^{\,*}(f_k,\vec{m})|^2 \tag{2}$$

with $\vec{m}$ the vector containing the unknown parameters, K the number of frequencies, ''$\cdot$'' indicating the inner product of the vectors $\vec{p}(f_k)$, the measured pressure field at frequency f_k, and $\vec{p}_c(f_k,\vec{m})$, the pressure field calculated for parameter set $\vec{m}$ and frequency f_k. Both pressure vectors are normalized such that their norm equals one, i.e., $\|\vec{p}\| = \|\vec{p}_c\| = 1$. From previous experience[8,10] it is known that using multiple frequencies for the inversion results in more accurate and more realistic parameter estimates compared to single-frequency inversion. In order to limit the computation time, which increases quadratic with frequency, we have selected a subset of frequencies (200, 300, 400, and 600 Hz) to be used for the inversions, i.e., K is 4. This covers sufficiently the frequency band transmitted: for a water sound speed of 1500 m/s, the wavelengths corresponding to the frequency subset range from 7.5 to 2.5 m. Including 700 Hz would give a shortest wavelength of 2.1 m, which is only slightly smaller than 2.5 m, but would result in a large increase in computation time.

Minimizing the energy function will lead to the parameter set that corresponds to a simulated pressure field $\vec{p}_c(f_k,\vec{m})$ having maximum similarity with the measured pressure field $\vec{p}(f_k)$.

B. The genetic algorithm

The objective function given in the previous section is a function of many unknown variables, usually on the order of 10, and with many local optima. Finding the global optimum of such a function requires global optimization methods such as simulated annealing and genetic algorithms. We have applied a genetic algorithm (GA) for finding the minimum of the objective function. The application of a GA in underwater acoustics was introduced by Gerstoft.[11] Specifics about the GA applied here can be found in previous work.[8] The basic principle of a GA is summarized below.

First an initial population, consisting of q possible solutions to the problem, is created randomly. This population is the so-called first generation. Out of this initial population elements are selected for establishing a parental population. The selection is such that the most fit members of the initial population, i.e., those with the lowest value for the objective function, have the highest probability of being selected. The elements of the parental population are converted to encoded form such that the numerical values are represented by a string of zeros and ones (bits). This string is called a chromosome and the different parts on the chromosome, which all represent a particular parameter, are called the genes. By applying the operators crossover and mutation to the elements of the parental population, a new population, denoted by children population, is created. The crossover operator results, for each gene, in an exchange between two parents of a (random) fragment of the gene. Crossover occurs with crossover probability p_{cross}. Mutation results in a change of a single bit and occurs with mutation probability p_{mut}.

A next generation is created by replacing the $f_r \cdot q$ least fit members of the first generation by $f_r \cdot q$ members of the children population. The latter are selected at random. f_r is called the reproduction size ($0 < f_r < 1$) and is an important parameter of the GA to be set.[8] We have taken $q = 64$, $f_r = 0.5$, $p_{\text{cross}} = 0.8$, and $p_{\text{mut}} = 0.05$.

By repeating the above-described process, the successive generations become increasingly fit. This process is continued until the optimization process has converged. Convergence is established by taking at least 400 generations, resulting in 12 832 forward acoustic model runs per frequency. For better exploitation of the search space around the global optimum and for diminishing the risk of ending up in a local minimum, five independent GA runs are carried out (i.e., $12\,832 \times 5 \times 4 = 256\,640$ forward runs per pressure field). As we have selected a population size q of 64, each set of five GA runs ends up with 320 parameter sets. From these parameter sets, the one that corresponds to the lowest energy function value is selected. This parameter set is denoted by GA_{best} and is taken to be the solution of the optimization.

C. The acoustic problem

As mentioned in the Introduction, an important application of MFI is geo-acoustic parameter estimation. However, in general, other parameters also need to be optimized as they have an important influence on the propagation, but are not known accurately enough. In this section the parameters included in the inversion are discussed.

Since the bathymetry along the 2-km acoustic track was found to be fairly range-independent (Sec. II B 4) we assume a constant water depth. Due to slight variations in water depth and offsets in the echosounder measurements the water depth H_w is not known exactly and has to be included in the optimization.

The geo-acoustic model selected consists of a single sediment layer with thickness h_{sed}, overlying a homogeneous subbottom. Justification for the single sediment layer assumption is obtained from literature.[12] Here it was found that inversions of synthetic data, calculated for a multi-layer bottom, and using a two-layer model for the forward calculations, resulted in properties of the two-layer bottom that fitted the properties of the actual multi-layer model reasonably well.

The sediment compressional wave speed is assumed to vary linearly with depth from $c_{1,\text{sed}}$ at the top of the sediment

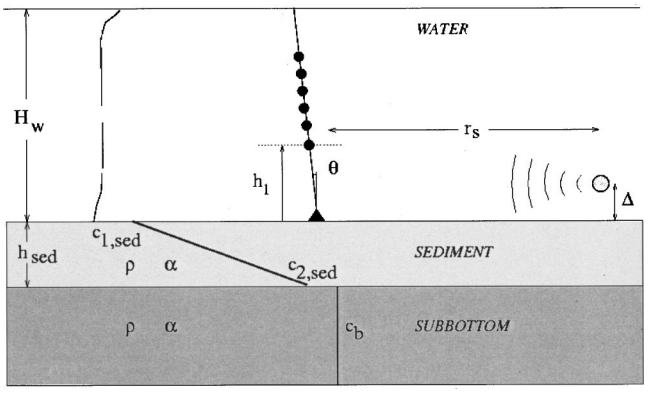


FIG. 11. The parameters to be optimized.

to $c_{2,\mathrm{sed}}$ at the bottom of the sediment, and to have a constant value c_b in the subbottom. The attenuation constant α and density ρ are taken depth independent through both the sediment and the subbottom. Shear is not taken into account. Shear speed effects and the justification for not including it in the inversion are discussed in Sec. IV A 1.

Due to irregularities at the sea bottom the depth of the array is not known precisely. Further, currents can result in a tilt of the array in some direction. Assuming that the effect of a tilt in the azimuth direction can be accounted for by an effective tilt θ in the plane of propagation, the array configuration is completely defined by estimating h_1, which is the distance of the deepest hydrophone to the bottom, and the array tilt θ.

The source range, r_s, and the source depth, here defined by the distance from the source to the bottom Δ, have a large influence on the acoustic propagation and are not known to the required accuracy. The baseline values of r_s, Δ, and h_1 are 2040, 4, and 9.5 m, respectively.

In Fig. 11 the resulting parametrization for the ocean environment and the 11 unknown parameters are shown. The sound speed profile used for the inversions is the sound speed profile that corresponds to the CTD taken from the NRV ALLIANCE on 2 May, 12:20 UTC, i.e., 17 min prior to the execution of the experiment (see Fig. 4).

Table I lists the unknown parameters and their search bounds.

IV. RESULTS AND DISCUSSION

The inversion was carried out for 41 pressure fields. These pressure fields were determined from 2-s snapshots of acoustic data differing in transmission time by approximately 15 min and spanning the entire 2-km experiment (about 8 h). The absolute values of the selected pressure fields are plotted in Fig. 2.

TABLE I. The unknown parameters, their symbols, and search bounds.

Parameter	Symbol	Search bounds
Upper sediment sound speed (m/s)	$c_{1,\mathrm{sed}}$	[1475 1700]
Sediment thickness (m)	h_{sed}	[1 25]
Lower sediment sound speed (m/s)	$c_{2,\mathrm{sed}}$	[1475 1800]
Subbottom sound speed (m/s)	c_b	[1515 1900]
Density (g/cm^3)	ρ	[1 2.3]
Attenuation constant (dB/λ)	α	[0 1]
Source range (m)	r_s	[1700 2500]
Distance source to bottom (m)	Δ	[0 10]
Tilt (degrees)	θ	[−10 10]
Water depth (m)	H_w	[75 85]
Distance lowest hydrophone to bottom (m)	h_1	[7.5 12.5]

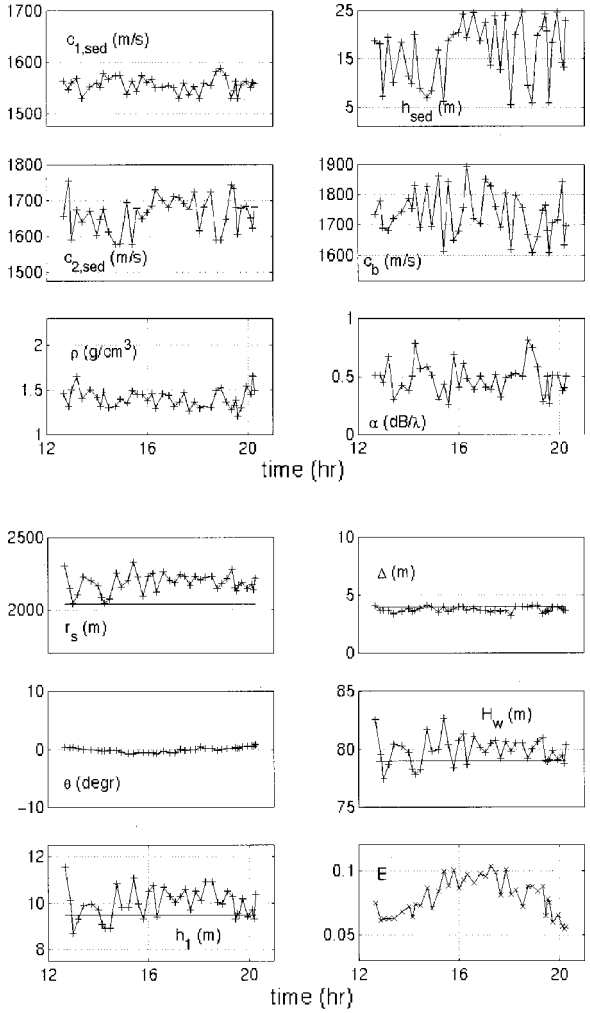


FIG. 12. Parameter estimates ($\mathrm{GA}_{\mathrm{best}}$) and corresponding energy function values as obtained from inversions of the experimental data. The horizontal solid lines in four of the subplots indicate measured values, where for the water depth we have used the depth at the vertical array position (79 m).

Figure 12 presents the results of the inversions. For each parameter a plot of the 41 succeeding $\mathrm{GA}_{\mathrm{best}}$ estimates is given. The lower right subplot shows the corresponding energy function values. The *y*-axes ranges of the plots are equal to the search bounds for the unknown parameters. The variable along the *x* axes is UTC time. Note the behavior of the $\mathrm{GA}_{\mathrm{best}}$ energy function *E* as a function of time. This must be due to temporal oceanographic variations, which are not taken into account in the inversions as we use a single sound speed profile for the forward calculations in all inversions.

Since the experimental configuration is stationary, all unknown parameters (except for the tilt, which might vary due to varying currents) should be constant with time. Therefore, the inversion results can be used for determining the mean and standard deviations for each parameter. Assuming statistically independent observations, also the uncertainty (or error) on the mean (σ_{mean}) and the uncertainty (or error) on the standard deviation (σ_{std}) can be determined by

TABLE II. The means and standard deviations (and their uncertainties) of the inversion results of Fig. 12. The last column lists the true measured values for the geometrical parameters.

Parameter	Mean (σ_{mean})	Std (σ_{std})	True measured values
$c_{1,\text{sed}}$ (m/s)	1556.3 (2.3)	14.8 (1.6)	
h_{sed} (m)	16.49 (0.99)	6.36 (0.70)	
$c_{2,\text{sed}}$ (m/s)	1663.1 (7.6)	48.8 (5.4)	
c_b (m/s)	1734.6 (11.9)	75.9 (8.4)	
ρ (g/cm^3)	1.405 (0.016)	0.101 (0.011)	
α (dB/λ)	0.482 (0.021)	0.134 (0.015)	
r_s (m)	2186 (10)	65.3 (7.2)	2040
Δ (m)	3.772 (0.034)	0.215 (0.024)	4.0
H_w (m)	79.89 (0.18)	1.16 (0.13)	77–79
h_1 (m)	10.00 (0.10)	0.646 (0.071)	9.5

$$\sigma_{\text{mean}}=\frac{\text{std}}{\sqrt{N}}, \quad \sigma_{\text{std}}=\frac{\text{std}}{\sqrt{2N}}, \tag{3}$$

with N the number of observations (here equal to 41), and std the standard deviation of the N observations. The errors on the mean and standard deviation must be accounted for, since the number of observations N is quite small (see Sec. IV A 1).

In Table II the values for the means and standard deviations, and their uncertainties, are listed. Hereafter we further assess the estimates for the different parameters.

A. Assessment of the parameter estimates

From the results presented in Fig. 12 and Table II it can be concluded that some parameters are estimated very accurately. The standard deviation of the upper sediment sound speed is only 15 m/s; the density is determined with a standard deviation of only 0.1 g/cm^3 and the distance of the source to the bottom within 0.2 m. Considering the tilt estimates it can be concluded that the tilt is estimated accurately enough to resolve for a trend. In the following we examine the estimates for the different parameters in detail.

1. Assessment of the geo-acoustic parameter estimates

As mentioned in Sec. II B 3, a seismic survey was carried out for obtaining independent information on bottom layering and bottom sound speeds. According to the seismic analysis the mean sound speed in the upper 6 m of the sediment is 1552 ± 31 m/s. The inverted values for $c_{1,\text{sed}}(1556\,\text{m/s})$, $h_{\text{sed}}(16\,\text{m})$, and $c_{2,\text{sed}}(1663\,\text{m/s})$ give a mean sound speed value of 1576 m/s over the upper 6 m of sediment. It can thus be concluded that this inversion result is in very good agreement with the seismic measurements.

Considering its large search interval (1.0–2.3 g/cm^3), the density is reasonably well resolved within 0.1 g/cm^3 (in agreement with other findings[1,11,12]). In Fig. 13 sediment sound speed and density values are plotted for different types of continental shelf sediment. In this figure also an empirical curve is shown, giving the relation between (compressional) sediment sound speed c_p and density ρ:

$$c_p=2330.4-1257\cdot\rho+487.7\cdot\rho^2\pm33 \ \text{ m/s}. \tag{4}$$

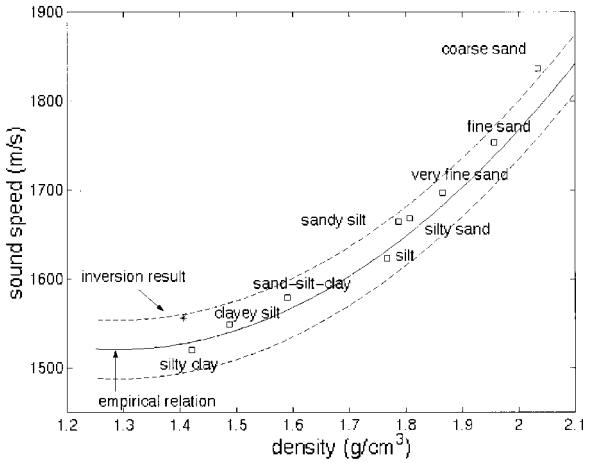


FIG. 13. Sound speeds and densities for continental shelf sediments. Also plotted is an empirical curve that gives the relation between sediment sound speed and density (Ref. 13). The star indicates the inversion result.

This expression has been derived as a fit through a large number of measurements.[13] The expression is valid for $1520<c_p<1840$ m/s and $1.25<\rho<2.10$ g/cm^3.

The inverted sediment density (1.41 ± 0.10 g/cm^3) and upper sediment sound speed (1556.3 m/s) values satisfy the above expression, but it is at the very low end of the brackets (see star in Fig. 13). We feel confident about our estimation of sediment sound speed since it is in agreement with independent seismic measurements and therefore we think the error is on the density. From the expression given here, and taking into account the error of 33 m/s, the inverted sediment sound speed should correspond to a density of 1.56 ± 0.12 g/cm^3, instead of the optimized density of 1.41 g/cm^3.

This deviation can be an indication for the presence of shear waves, as neglecting the effect of shear is accounted for by an effective density.[14,15] This is expressed through the following relation between actual density ρ, shear speed c_s, and an effective density ρ_{eff}:

$$\rho_{\text{eff}}=\rho\left[1-2\left(\frac{c_s}{c_w}\right)^2\right]^2 \tag{5}$$

with c_w the sound speed in the water. Employing this expression, a sediment density of 1.56 ± 0.12 g/cm^3 and an effective sediment density of 1.405 ± 0.101 g/cm^3 are found to correspond to a shear speed of 241 ± 121 m/s. This basically means that shear cannot be inverted since the presence of shear results in an effective density. This is illustrated in Figs. 14 and 15 where the complex reflection coefficient[7] is plotted as a function of incident grazing angle for homogeneous bottoms with a compressional wave speed of 1556.3 m/s and attenuation constant of 0.5 dB/λ and with values for the density and the shear speed as listed below:

(i) BOTTOM1: $\rho=1.56$ g/cm^3, $c_s=0$ m/s;
(ii) BOTTOM2: $\rho=1.56$ g/cm^3, $c_s=241$ m/s; and
(iii) BOTTOM3: $\rho=1.41$ g/cm^3, $c_s=0$ m/s.

BOTTOM1 is included for illustration: comparing the results for BOTTOM1 and BOTTOM2 shows the influence of shear speeds. The reflection coefficient for BOTTOM2 and BOTTOM3 (both the absolute value and the phase shift)

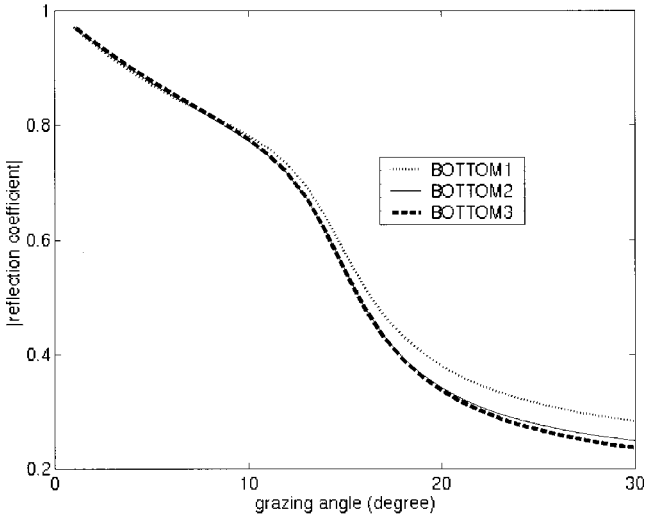


FIG. 14. Magnitude of the reflection coefficient versus angle for BOTTOM1, BOTTOM2, and BOTTOM3. Note that the curves for BOTTOM2 and BOTTOM3 virtually coincide.

virtually coincide, which demonstrates that the effect of shear is compensated by a decrease in density as given by Eq. (5).

Another method for estimating shear speeds is to apply the following regression equation,[16] relating shear speed c_s (in m/s) to compressional wave speed c_p (also in m/s):

$$c_s = 1.137 \cdot c_p - 1485. \tag{6}$$

This equation is valid for $1555 < c_p < 1650$ m/s.

From this equation a shear speed value of 285 ± 17 m/s is calculated for a compressional wave speed of 1556.3 ± 14.8 m/s. The two derived values for the shear speed, viz., 241 ± 121 m/s and 285 ± 17 m/s, are in agreement and are therefore considered to be reliable indicative estimates for the sediment shear speed. Note that due to their similar effects on the reflection coefficient, it will not be possible to unambiguously estimate both the density and the shear speed simultaneously through inversion.

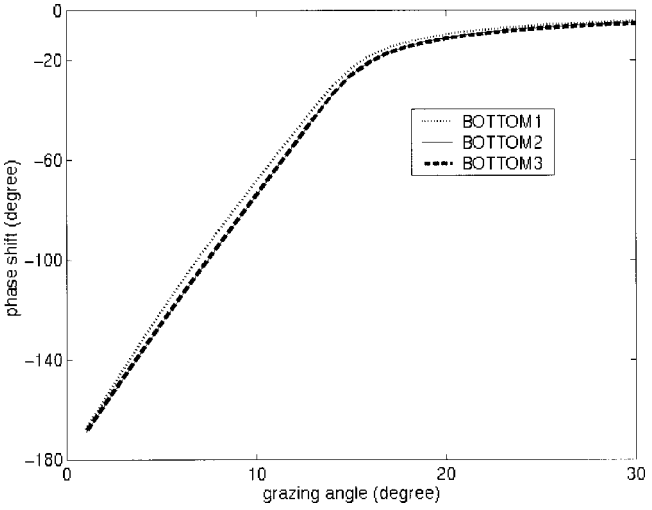


FIG. 15. Phase shift of the reflection coefficient versus angle for BOTTOM1, BOTTOM2, and BOTTOM3. Note that the curves for BOTTOM2 and BOTTOM3 virtually coincide.

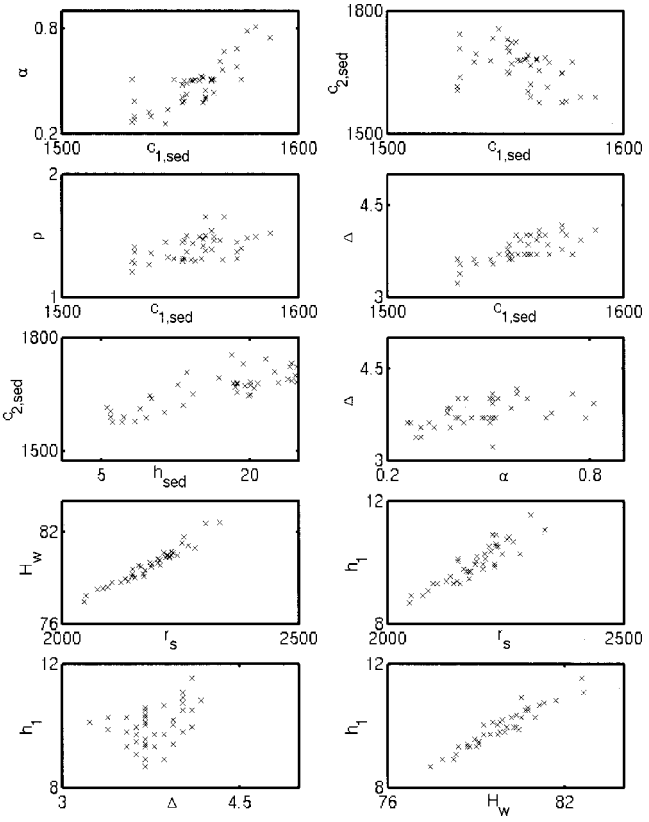


FIG. 16. Final parameter estimates plotted against each other for ten parameter combinations with strong coupling.

The sediment attenuation constant is not well resolved, but remains within the bounds of 0.25 and 0.81 dB/λ, corresponding roughly to the values for clay and sand.[7]

Parameter h_{sed} should be interpreted carefully. The multi-layer structure of the bottom is not included in the geo-acoustic model. However, from the seismic data analysis we know that, in reality, the sediment has some internal layering and is not made up of a single layer (see Fig. 7). Also, the layer's thicknesses are not independent of range (as assumed in the inversions). h_{sed} represents a breakpoint, setting a depth in the bottom below which the bottom is considered homogeneous and above which it fits the sound speed gradient determined by $c_{1,\text{sed}}$ and $c_{2,\text{sed}}$. No independent measurements are available for the sediment thickness h_{sed} since from the seismic data no clear reflector that marks the boundary between sediment and subbottom could be identified. The parameter is highly undetermined, as the optimized values for h_{sed} comprise nearly the entire search space. From the inversion results we found the two parameters h_{sed} and $c_{2,\text{sed}}$ to be strongly correlated. This indicates that these parameters do have an influence on the propagation, but that they cannot be estimated separately. Parameter correlation is illustrated in Fig. 16 in which the 41 final parameter estimates are plotted against each other. Ten parameter combinations with strong coupling are shown here. It is clearly observed from this figure that some parameters, such as h_{sed} and $c_{2,\text{sed}}$, are highly correlated. Avoiding parameter coupling possibly requires a completely different parametrization of the environment[11,17–19] (e.g., inverting for the sediment sound

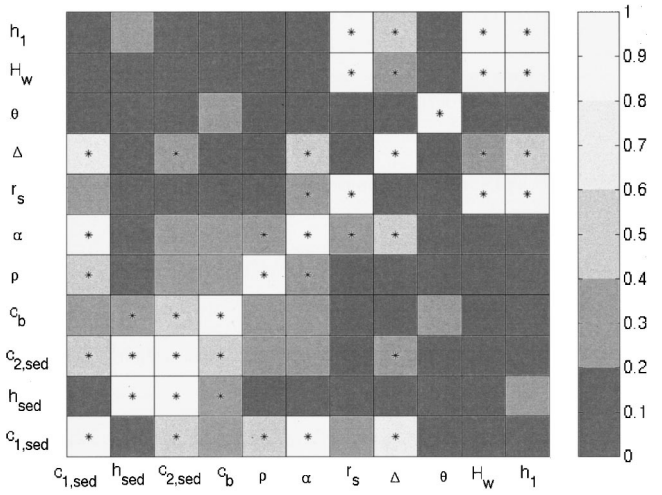


FIG. 17. The correlation coefficient for all parameter combinations.

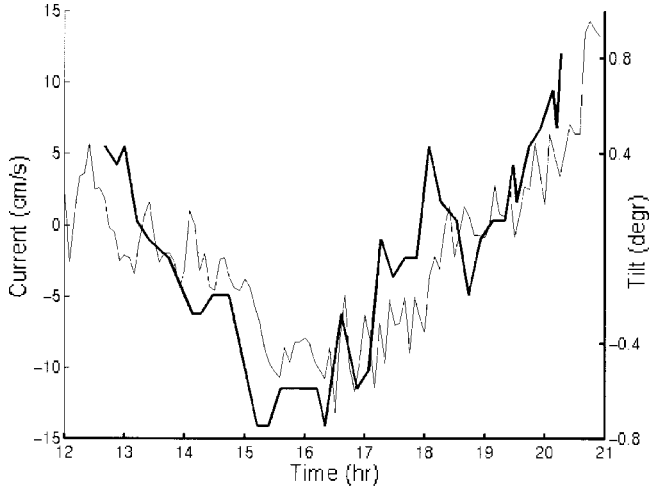


FIG. 18. Eastward current (thin solid line) and tilt estimates (thick solid line) as a function of UTC time.

speed gradient and h_{sed}, instead of h_{sed} and $c_{2,\text{sed}}$). The full parameter coupling of the inversion problem is illustrated in Fig. 17, which shows the magnitude of the linear correlation coefficient calculated for all combinations of parameter estimates. It is emphasized that these calculated values for the linear correlation coefficient are based on a fairly limited number of data pairs (41), i.e., also the statistical significance of each of the correlation coefficient values was determined. The stars in Fig. 17 indicate for which parameter combination the confidence of the calculated correlation coefficient exceeds 95%.[20] For instance, the correlation coefficient between Δ and H_w is 0.34 with a confidence of 96.9%, i.e., there is a probability of 3.1% that the observed correlation can occur between two random (uncorrelated) data sequences of length 41. Correlation coefficients greater than 0.5 have a confidence of at least 99.98%. The advantage of reparametrization of the environment in this specific inversion problem is subject to further research.

2. Assessment of the array tilt estimates

From the parameter estimates as a function of time as shown in Fig. 12 it can be seen that there is a trend in the estimates for the VA tilt. This trend should correspond to some similar trend in the currents close to the VA. The VA and the source are at nearly equal latitudes and we consider array tilt along the acoustic track. Therefore we consider only the east/westwards directed currents. In Fig. 18 we have plotted both the estimates for the tilt and the eastward component of the currents as measured by the bottom-moored ADCP at a depth of 27 m. The tilt is positive in the eastward direction (towards the tower source) and negative westwards (away from the tower source). This also holds for the currents. It is clear that the current and the tilt are in excellent agreement as they show a very similar trend. This indicates that the inverted tilt estimates are reliable.

3. Assessment of the geometrical parameter estimates

From Fig. 12 and Table II it can be seen that for the geometrical parameters, i.e., r_s, Δ, H_w, and h_1, the inverted values are in good agreement with the true values, especially when considering the uncertainty (standard deviation) in these parameter estimates. The true range, being 2040 m as calculated from the navigation data, slightly deviates from the mean inversion result. This can be partly ascribed to a

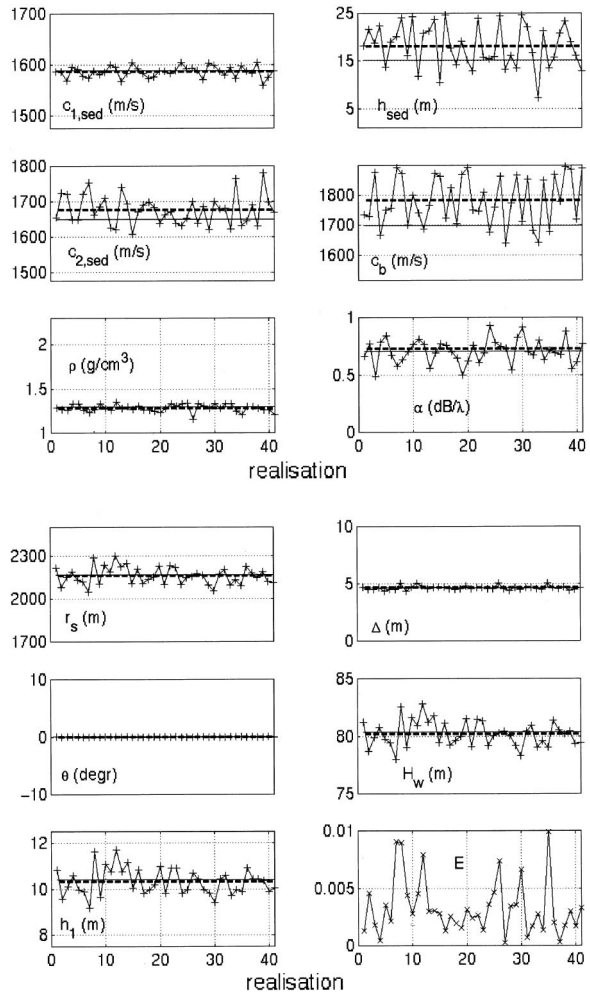


FIG. 19. Parameter estimates and corresponding values for the energy function (lowest, most right subplot) as obtained from inversions of synthetic data (SIM1). The dashed horizontal lines are the mean values, whereas the solid horizontal lines indicate the true values.

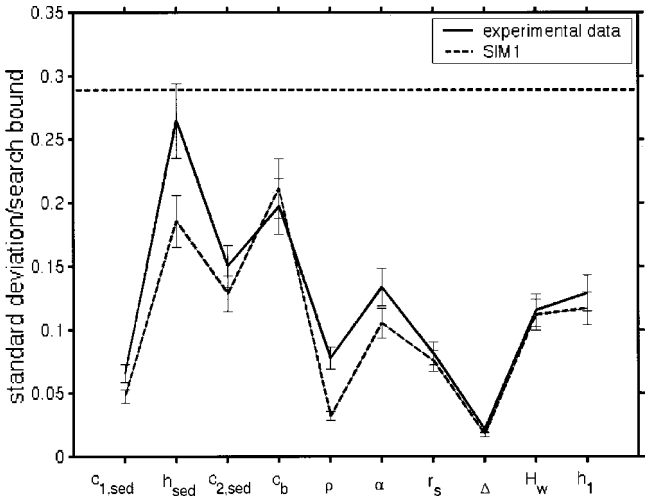


FIG. 20. The standard deviations and their statistical errors divided by the appropriate search bounds, for all parameters except for the tilt, for both the experimental data and simulated data. The horizontal dashed line indicates the relative standard deviation for a uniform distribution ($1/\sqrt{12}$).

difference between the DGPS positions of the ship at the time of source and VA deployment and the final source and VA position.

The water depth, the source range, and the distance of the deepest hydrophone to the bottom are strongly correlated (see Fig. 17). This can explain the minor differences between true and inverted values. Neglecting the variations in water depth over the acoustic track in the inversions will result in an effective water depth. It is known that there is a small variation in water depth over the 2-km acoustic track (from 77 at the source to 79 m at the VA). Therefore, the estimates for the geometrical parameters that correspond to the resulting effective water depth might deviate from their actual values. In the next section it is shown that also oceanographic variability can result in a shift of geometrical parameter values.

B. Assessment of the parameter uncertainties

As mentioned before, there are two main origins for the variations in the parameter estimates. The first origin is the method itself. The second is oceanographic variability. The contribution of both is determined through simulation and is presented and discussed in this section. Other contributions to the uncertainty can be due to a low signal-to-noise ratio, e.g., due to noise of passing ships, and wrong parametrization. As the signal-to-noise ratio was very high, its contribution to the uncertainty is negligible. We will demonstrate that the above-mentioned two main origins can fully explain the obtained parameter uncertainties. Wrong parametrization will not be considered here.

The contribution of the method to the parameter uncertainty was determined through the repetitive inversion (41 times) of a simulated pressure field, i.e., 41 optimization runs were performed on the same synthetic pressure field. As a genetic algorithm is a Monte Carlo search method, this approach requires the use of a different random seed for each of these 41 optimization runs. The parameters used for creating this synthetic field are values close to the mean parameter values obtained from inversion of the experimental data. This simulation is further denoted as SIM1. The results of these inversions can be found in Fig. 19. Also shown are the true values (solid horizontal lines) and the means of the inversion results (dashed horizontal lines). For all parameters, except for h_{sed}, $c_{2,\text{sed}}$, and c_b, the mean values virtually coincide with the true values. In Table III the deviation δ_{mean}

TABLE III. The deviation of the mean values from the true values (δ_{mean}) and the standard deviation (Std) obtained for the two simulations. The last column lists the standard deviation for the experimental data inversions. The statistical errors on δ_{mean} and std are given in parentheses.

		SIM1		SIM2		EXP
Parameter	True	δ_{mean} (σ_{mean})	std (σ_{std})	δ_{mean} (σ_{mean})	std (σ_{std})	std (σ_{std})
$c_{1,\text{sed}}$ (m/s)	1588.9	−2.8	10.7	−12.9	16.0	14.8
		(1.7)	(1.2)	(2.5)	(1.8)	(1.6)
h_{sed} (m)	15.2	2.88	4.46	2.49	5.09	6.36
		(0.70)	(0.49)	(0.80)	(0.56)	(0.70)
$c_{2,\text{sed}}$ (m/s)	1650.4	24.9	41.8	36.9	46.4	48.8
		(6.5)	(4.6)	(7.2)	(5.1)	(5.4)
c_b (m/s)	1698	84	81.4	75	81.2	75.9
		(13)	(9.0)	(13)	(9.0)	(8.4)
ρ	1.292	−0.0154	0.0418	0.004	0.0777	0.101
		(0.0065)	(0.0046)	(0.012)	(0.0086)	(0.011)
α (dB/λ)	0.724	−0.018	0.105	−0.123	0.128	0.134
		(0.016)	(0.012)	(0.020)	(0.014)	(0.015)
r_s (m)	2168	−7.4	60.7	−11.3	59.3	65.3
		(9.5)	(6.7)	(9.3)	(6.6)	(7.2)
Δ (m)	4.63	−0.004	0.177	−0.190	0.260	0.215
		(0.028)	(0.020)	(0.041)	(0.029)	(0.024)
θ (degrees)	0.024	−0.0250	0.0397	−0.0192	0.0498	0.419
		(0.0062)	(0.0044)	(0.0078)	(0.0055)	(0.046)
H_w (m)	80.35	−0.16	1.12	−0.36	1.15	1.16
		(0.17)	(0.12)	(0.18)	(0.13)	(0.13)
h_1 (m)	10.42	−0.104	0.585	−0.39	0.693	0.646
		(0.091)	(0.065)	(0.11)	(0.077)	(0.071)

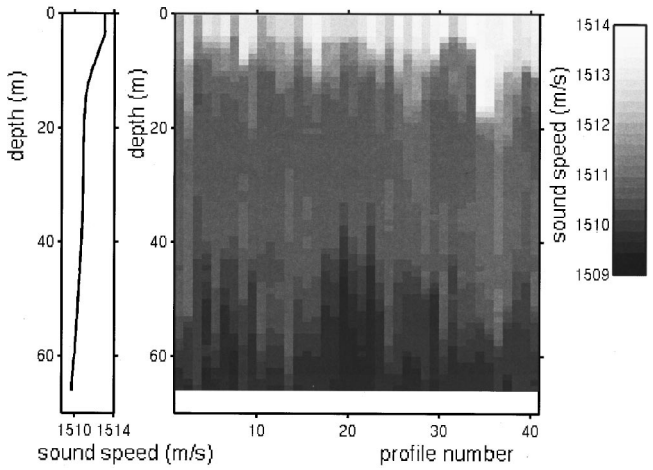


FIG. 21. The sound speed profiles for which synthetic pressure fields were calculated. The data are obtained from the CTD-chain measurements on the 2-km acoustic track. The left-most plot shows the mean sound speed profile.

of the mean values from the true values is listed together with the standard deviations.

In Fig. 20 the standard deviations of the parameter estimates (std) are plotted, both for the experimental data inversion results and the simulation. The uncertainties in the standard deviation (σ_{std}), representing the statistical error, are also plotted in the figure as error bars. By dividing both by the appropriate search bounds a direct comparison between parameter uncertainties can be made. The parameter tilt is not considered as in the experimental configuration the tilt varies, whereas it is constant in the simulation. In the figure also the relative standard deviation for uniformly distributed parameter values (being equal to $1/\sqrt{12}$) is plotted (horizontal dashed line). Parameters whose standard deviation divided by the search bound approach this value are badly determined: they have a uniform distribution over the entire search area. From Fig. 20 it can be seen that this is the case for the sediment thickness as determined from the experimental data. It can be concluded that for $c_{2,\text{sed}}$, c_b, r_s and Δ, H_w, and h_1, the uncertainties in the inversion results for the experimental and the synthetic data coincide within the statistical error. Hence, for these parameters the uncertainties can be completely ascribed to the method itself. For the remaining parameters the additional uncertainty should be caused by oceanographic variability. Note, however, that also for these parameters a significant part of the uncertainty stems from the method itself.

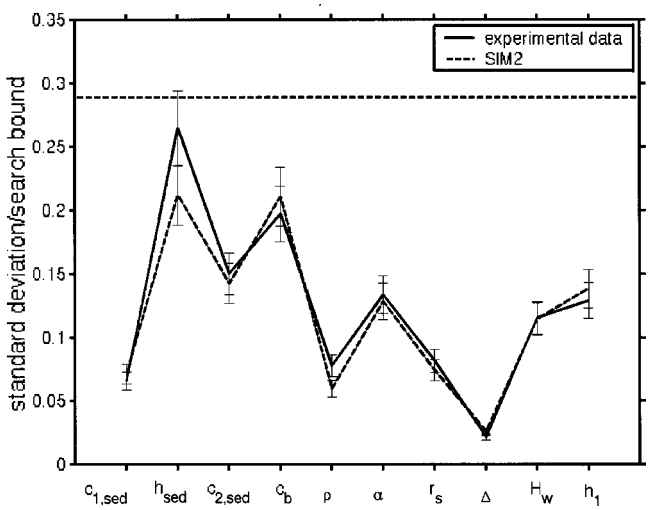


FIG. 23. The standard deviations and their statistical errors divided by the appropriate search bounds, for all parameters except for the tilt, for both the experimental data and for simulation SIM2. The horizontal dashed line indicates the relative standard deviation for a uniform distribution ($1/\sqrt{12}$).

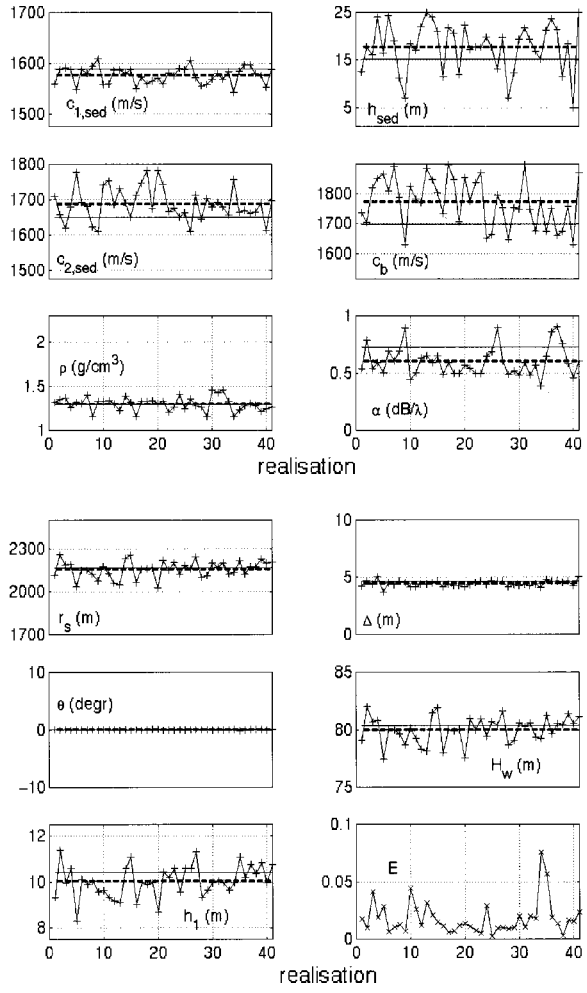


FIG. 22. SIM2 results: Parameter estimates and corresponding values for the energy function (lowest, right-most subplot). The means of the parameter values are plotted as dashed lines. The solid lines indicate the true values.

From the CTD-chain measurements it is known that the sound speed profile varies with time (and range). For investigating the effect of these variations on the inversion results we have proceeded as follows: 41 sound speed profiles measured by the CTD-chain were selected on the 2-km acoustic track, i.e., in between 12° 15.588′E (source position) and 12° 14.207′E (VA position). The selected profiles, plotted in Fig. 21, represent realistic oceanographic variations on the 2-km acoustic track. For these 41 sound speed profiles, the corresponding pressure fields were calculated (for the four frequencies). For the unknown parameters we have taken the same values as those for SIM1. The spatial structure of the sound velocity field between source and receiver is not exactly accounted for in these calculations, as each of the 41 calculations is still range independent.

The resulting pressure fields were subsequently inverted. The sound speed profile used for the forward calculations is the mean of all profiles from the CTD-chain data as given in Fig. 5. This mean profile, also shown in Fig. 21, is obtained by simply taking the mean of all sound speeds at each depth. This simulation is further denoted as SIM2.

In Fig. 22 the resulting parameter estimates can be found. Also plotted are the true values (solid lines) and the means of all 41 estimates (dashed lines). The bottom, rightmost subplot shows the corresponding lowest energy function values. As for SIM1, Table III lists the deviation δ_{mean} of the mean parameter estimates from the true values. Also given are the standard deviations.

From the results presented in Fig. 22 and Table III several conclusions can be drawn. Introducing oceanographic variability can result in a shift of parameter estimates. A more important effect of the oceanographic variability is an increase in the standard deviation for some parameters. This is illustrated in Fig. 23. Variability in the water column has resulted in a statistically significant increase of the standard deviation for the parameters $c_{1,\text{sed}}$, ρ, and Δ (compare with Fig. 20). From this figure it can also be seen that now we have almost completely explained the uncertainty for all parameters. For h_{sed} and ρ the experimental data still show a statistically significant larger standard deviation compared to the simulations. This might be due to mismatch: in SIM2 the geo-acoustic model assumed is per definition correct. This is not true for the inversions of experimental data as, for example, we have assumed a single layer sediment. Further we have assumed the density and attenuation to be constant throughout the bottom, which is probably not true in reality.

V. SUMMARY AND CONCLUSIONS

In this article we reported matched field inversion (MFI) results of multi-tone shallow water acoustic data that were obtained during the ADVENT99 sea trial. The multi-tone data were collected during an experiment with both the source and the receiver at a fixed position. The range between source and receiver was 2 km. In order to be able to assess the influence of oceanographic variability, the experiment was continued for as long as 8 h. A genetic algorithm was applied as the global optimization method, whereas a normal mode model was used for the forward propagation calculations. The parameters to be inverted, i.e., the unknown parameters, comprise both geo-acoustic and geometric parameters.

Many measurements were carried out for determining parameters of the ocean environment. These parameters are needed as input for the forward acoustic model and to obtain independent information about the unknown parameters (e.g., multi-channel seismic measurements for obtaining estimates of sediment thickness and sediment speed). The third objective of these environmental measurements is to monitor oceanographic variability. For this, a CTD-chain was towed back and forth over the acoustic track, thereby measuring sound speeds as a function of time and position. It was found that many of the unknown parameters could be estimated very accurately through inversion and that they are in good agreement with the independent measurements.

An important item addressed in this article is the uncertainty of the parameter estimates. It is shown through simulation that the main part of this uncertainty stems from the method itself, as the optimization does not locate the exact global optimum and many of the parameters are correlated. Simulations including the water column variability, as determined from the CTD-chain measurements, show that for nearly all parameters the remaining uncertainties can be ascribed to this oceanographic variability.

ACKNOWLEDGMENTS

The authors would like to thank all participants of the ADVENT99 sea trial, E. Michelozzi and P. Boni from SACLANT Centre in particular. Thanks are also due to the crews of NRV ALLIANCE and INS *Ciclope*. The fruitful collaboration with Dr. M. Ainslie from CORDA is especially acknowledged. The authors further thank C. Mesdag and P. Frantsen from TNO-NIAG for performing the seismic experiments and subsequent seismic data analysis.

[1] P. Gerstoft, ''Inversions of acoustic data using a combination of genetic algorithms and the Gauss–Newton approach,'' J. Acoust. Soc. Am. **97**, 2181–2190 (1995).
[2] F. Gingras and P. Gerstoft, ''Inversion for geometric and geoacoustic parameters in shallow water: Experimental results,'' SR-229, SACLANT Undersea Research Centre, La Spezia, Italy (1994).
[3] J. Sellschopp, M. Siderius, and P. Nielsen, ''ADVENT99, pre-processed acoustic and environmental cruise data,'' CD-35, SACLANT Undersea Research Centre, La Spezia, Italy (2000).
[4] T. M. Siderius, M. Snellen, D. G. Simons, and R. Onken, ''An environmental assessment in the Strait of Sicily: Measurement and analysis techniques for determining bottom and oceanographic properties,'' IEEE J. Ocean Eng. **25**(3), 364–387 (2000).
[5] L. Hatton, M. H. Worthington, and J. Makin, *Seismic Data Processing, Theory and Practice* (Blackwell Science, London, 1986).
[6] C. S. Mesdag and T. A. M. de Groot, ''Seismic survey in the Strait of Sicily, part of the ADVENT99 cruise,'' TNO-NIAG report, NITG 99-233-C (1998).
[7] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994).
[8] D. G. Simons and M. Snellen, ''Multi-frequency matched-field inversion of benchmark data using a genetic algorithm,'' J. Comput. Acoust. **6**(1&2), 135–150 (1998).
[9] A. Tolstoy, *Matched Field Processing for Underwater Acoustics* (World Scientific, River Edge, NJ, 1993).
[10] G. Haralabus and P. Gerstoft, ''Variability of multi-frequency parameter estimates in shallow water,'' in *Third European Conference on Underwater Acoustics*, edited by J. S. Papadakis (Heraklion, Crete, 1996), Vol. I, pp. 355–360.
[11] P. Gerstoft, ''Inversion of seismoacoustic data using genetic algorithms and a *posteriori* probability distributions,'' J. Acoust. Soc. Am. **95**, 770–782 (1994).
[12] M. Siderius, P. Gerstoft, and P. Nielsen, ''Broadband geoacoustic inversion from sparse data using genetic algorithms,'' J. Comput. Acoust. **6**(1&2), 117–134 (1998).
[13] E. L. Hamilton and R. T. Bachman, ''Sound velocity and related properties of marine sediments,'' J. Acoust. Soc. Am. **72**, 1891–1904 (1982).
[14] D. M. F. Chapman, P. D. Ward, and D. D. Ellis, ''The effective depth of a Pekeris ocean waveguide, including shear effects,'' J. Acoust. Soc. Am. **85**, 648–653 (1989).
[15] Z. Y. Zhang and C. T. Tindle, ''Improved equivalent fluid approximations for a low shear speed ocean bottom,'' J. Acoust. Soc. Am. **98**, 3391–3396 (1995).
[16] E. L. Hamilton, ''V_p/V_s and Poisson's ratios in marine sediments and rocks,'' J. Acoust. Soc. Am. **66**, 1093–1101 (1979).

[17] P. Gerstoft, ''Global inversion by genetic algorithms for both source position and environmental parameters,'' J. Comput. Acoust. **2**, 251–266 (1994).
[18] M. D. Collins and L. Fishman, ''Efficient navigation of parameter landscapes,'' J. Acoust. Soc. Am. **98**, 1637–1644 (1995).
[19] C. A. Zala and J. M. Ozard, ''Estimation of geoacoustic parameters from narrowband data using a search-optimization technique,'' J. Comput. Acoust. **6**(1&2), 223–243 (1998).
[20] J. S. Bendat and A. G. Pierson, *Random Data: Analysis and Measurement Procedures* (Wiley, New York, 1971), pp. 125–129.

The intensity coherence function of time for partially saturated acoustic propagation through ocean internal waves

James S. Gerber and Stanley M. Flatté
Physics Department and Institute of Tectonics, University of California at Santa Cruz, Santa Cruz, California 95064

(Received 5 August 1999; revised 17 May 2000; accepted 6 September 2000)

The intensity coherence function of time for partially saturated acoustic propagation through internal waves is calculated with a method that is improved over previous treatments. Two specific improvements are introduced: the usual expansion in $(1/\Lambda\Phi^2)$ is carried out to a higher order, and then the terms of that expansion are calculated with a new perturbative method. The method is applied to propagation without a sound channel, for both phase-screen and continuous-medium cases. The validity of the new perturbative method is estimated by calculating the next order error terms. Accuracies at the few-percent level are found. The new analytic formulas are also corroborated with numerical integration. Finally, the method is applied to a specific ocean-acoustic experiment [Azores Fixed Acoustic Range (AFAR)]. In order to achieve good agreement with experiment it will be necessary to add an accurate treatment of the sound channel to the present perturbation method. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1322020]

PACS numbers: 43.30.Re, 43.60.Cg, 43.60.Gk, 43.30.Ft

I. INTRODUCTION

The acoustic field received after propagation through many kilometers of ocean internal waves contains important statistical information about the internal-wave field. Typical observables used as internal-wave probes have been phase and amplitude fluctuations, pulse travel time, and field coherences, both temporal and spatial. In order to calculate such internal-wave effects, a path-integral approach was introduced in the early 1980s.[1,2] That work introduced the Λ–Φ diagram, on which the various regimes of propagation behavior can be located. The parameter Φ is a measure of the strength of the internal-wave-induced sound-speed fluctuations, and the parameter Λ is a measure of the importance of diffraction in the propagation. The regimes for which the quantity $\Lambda\Phi^2$ is large are called the saturated regimes. If $\Lambda\Phi<1$ the regime is called partially saturated.

The use of intensity as an observable was discussed in detail at that time, but it has proven difficult for two reasons: first, the theory predicting the effects of internal waves is more complex; and second, the acoustic intensity is more sensitive to the deterministic environment (the sound channel) than are the other observables mentioned.[1,2]

In 1987, Flatté *et al.*[3,4] used the above-mentioned path-integral approach to calculate the intensity fluctuations for a 35-km acoustic experiment, done with acoustic frequencies between 400 and 5000 Hz, at a geographical location off the Azores (AFAR).[3,4] Though the comparison was reasonably successful, the evaluation of the path integrals had many *ad hoc* aspects. It is the purpose of this paper to present improvements in the evaluations of the path integrals necessary for the calculation of the temporal coherence function of intensity, in cases for which $1/\Lambda\Phi^2$ is small, and at the same time, $\Lambda\Phi$ is also small. (Qualitatively, this condition can be achieved by a number of combinations of long range, high internal-wave strength, or high frequency.) We have chosen a set of internal-wave parameters that correspond to the expansion parameter being reasonably small in order to test our formulas. But we could not take too small a value, or we would not be able to do comparable simulations. For the cases calculated, we have found that the new analytic formulas are accurate to a few percent or less, as determined by explicit calculation of the next term in the expansion. We have also compared with numerical simulations using the parabolic approximation, and have found excellent agreement.

Finally, we have used internal-wave parameters that mimic the AFAR experiment, and have then used the analytic formulas to calculate the intensity coherence function and compare with the experimental results. The comparison is off by a factor of order 2. However, the analytic improvements have not included a sound channel, and we associate the discrepancy with that fact; thus comparisons with ocean experimental results require that the sound channel be included in the theory.

The organization of the paper is as follows: Section II describes the internal-wave spectral model we have used, in order first to derive and then to test our formulas for the intensity coherence function; Sec. III gives the path-integral expression from which we start our calculations; and Sec. IV describes the derivations and results. Section V describes our method of numerical simulation that is used to calculate results to compare with our derived formulas. Sections VI and VII give the results of our calculations and the comparisons with simulations for the single-screen and continuous-medium cases, respectively. Section VIII shows our comparisons with the AFAR experimental results. Section IX gives a summary.

II. THE INTERNAL-WAVE SPECTRUM

Internal waves vary in both space and time. A statistical description of the internal-wave field in the oceans of the world, based on empirical measurements, has been proposed

by Garrett and Munk in 1972.[5] The so-called Garrett–Munk spectrum has been modified repeatedly over the years to conform to current oceanographic measurements.[6,7]

The salient features of the Garrett–Munk spectrum include separability of the wave number and frequency dependence of the spectrum, and power-law behavior of the wavenumber power spectrum. Since the power spectrum factorizes, so do the medium correlation functions, which will simplify a number of our expressions.

In order to treat a sound-propagation problem, we need the spectrum of the integral of the sound-speed fluctuations along a ray, rather than the spectrum of the sound-speed fluctuations themselves. We introduce the temporal and spatial power spectra ($A(\omega)$ and $P(\kappa)$, respectively), and express the power spectrum of the phase integral along the ray as $S(\omega,\kappa)$:

$$S(\omega,\kappa)=A(\omega)P(\kappa), \tag{1}$$

where ω is the internal-wave frequency and κ is the wave number of the vertical fluctuations.

A. Temporal behavior

We define the temporal internal-wave correlation function $R_t(t)$ as the Fourier transform of $A(\omega)$:

$$R_t(t)=\frac{1}{2\pi}\int e^{i\omega t}A(\omega)d\omega. \tag{2}$$

In order that $R_t(0)=1$, we require that $\int A(\omega)d\omega = 2\pi$.

1. Taylor expansion

Typically, one is interested in studying acoustic coherence over times which are very short compared to the average internal-wave correlation time. Consequently, whenever a functional form for $R_t(\Delta t)$ is needed, it will suffice to Taylor expand $R_t(\Delta t)$

$$R_t(\Delta t)\approx 1-\tfrac{1}{2}\{\omega^2\}(\Delta t)^2+\ldots, \tag{3}$$

where $\{\omega^2\}$ is the average internal-wave frequency, defined following Esswein and Flatté as a weighted spectral average:[8]

$$\{\omega^2\}\equiv\frac{\int\omega^2A(\omega)d\omega}{\int A(\omega)d\omega}. \tag{4}$$

2. Time scales

Appropriate definitions of the time scales of the internal-wave field, the acoustic field, and the acoustic intensity are the following:

The internal-wave field itself:

$$t_{IW}\equiv\{\omega^2\}^{-1/2}, \tag{5}$$

whose typical value is 1 h.[9]

The coherence time of the acoustic field:

$$t_c\equiv\frac{t_{IW}}{\Phi}. \tag{6}$$

The coherence time of the acoustic intensity:

$$t_I\equiv\frac{t_{IW}}{\Lambda\Phi^2}. \tag{7}$$

These definitions are consistent with the relation $t_I\approx t_c/\Lambda\Phi$, which holds in the regime of partial saturation.[10]

B. Spatial behavior

We have used the following for P:

$$P(\kappa)=\begin{cases}0, & |\kappa|<\kappa_m\\ 2\pi\Phi^2\kappa_m^2\kappa^{-3}, & |\kappa|>\kappa_m.\end{cases} \tag{8}$$

The quantities Φ and κ_m can be given in terms of oceanographic and acoustic parameters in the following way[1]:

$$\Phi^2=\frac{1}{\pi^2}\frac{j_*}{B}\langle\mu_0^2\rangle\frac{k_0^2}{\kappa_m^2}R, \tag{9}$$

where μ_0 is the rms fractional sound-speed fluctuation at a depth at which the buoyancy frequency is 3 cph; k_0 is the acoustic wave number, and R is the range between source and receiver.

For analytic convenience, the spectrum has an abrupt cutoff at the outer scale, represented by a cutoff wave number κ_m. We specify the precise value of κ_m so that the total variance of the fluctuations remains unchanged from the value it has in a modal representation, where the exact effects of the surface and bottom boundary conditions are met. The condition on κ_m in terms of common oceanographic parameters is

$$\kappa_m^{-2}=2\left(\frac{B}{\pi}\right)^2\sum_{j=1}^{\infty}\frac{1}{(j^2+j_*^2)j}. \tag{10}$$

The quantity j_* determines the shape of the spectrum in vertical mode number j. The scale length B characterizes the roughly exponential drop in the density gradient, or buoyancy frequency, as a function of depth in the main thermocline (that is, below a few hundred meters). Typical values for B and j_* are respectively 1 km and 3.[1]

Values of $2(\kappa_mB/\pi)^2=[\Sigma_{j=1}^{\infty}1/[(j^2+j_*^2)j]]^{-1}$, for $j_*=1$, 2, 3, 4, and 5, are respectively 1.49, 3.10, 5.34, 8.13, and 11.42.

In this way we have introduced Φ, the strength parameter, one of two internal-wave parameters that determine the acoustic propagation regime. The other important parameter is the diffraction parameter Λ, defined here as

$$\Lambda\equiv\overline{R_f^2}\kappa_m^2, \tag{11}$$

where $\overline{R_f^2}$ is the square of Fresnel length, averaged over range. The Fresnel length R_f at a distance R from the receiver is approximately $R_f=[R/k_0]^{1/2}$. This definition of Λ has been chosen for analytic convenience and differs slightly from that of Flatté *et al.*[1] For the phase-screen case, we find $\overline{R_f^2}=R/k_0$, and for the continuous case, $\overline{R_f^2}=R/2k_0$.

A spatial correlation function of depth interval may be derived from the vertical spatial spectrum $P(\kappa)$, but it turns out that the more important spatial function is the one-dimensional spatial phase-structure-function density. This

function represents the difference between the integrals of the sound-speed fluctuation μ along two parallel lines separated vertically by Δz:[8] Because the sound-speed fluctuations due to internal waves are uncorrelated for separations beyond a few kilometers, the integral may be extended to infinity without loss in accuracy. Methods for dealing with this integral, including cases in which the rays curve significantly, are discussed and treated by Flatté and Rovner.[11]

$$d(\Delta z)=\frac{1}{\pi}\int_{-\infty}^{\infty}(1-e^{i\kappa\Delta z})P(\kappa)d\kappa. \tag{12}$$

We can also express the phase structure-function density $d(\Delta z)$ as a function of transverse separation u rather than in terms of a spectrum of transverse wave number. However, even in the absence of a waveguide and heterogeneous random medium, the result is a rather more complicated form than when expressed in terms of the transverse spectrum.[10]

$$d(u)=\frac{\Phi^2\kappa_m^2u^2}{R}\left[\ln\left(\frac{e^{3-2\gamma}}{\kappa_m^2u^2}\right)+\sum_{m=2}^{\infty}\frac{(u^2\kappa_m^2)^{m-1}(-1)^m}{(2m)!(m-1)}\right]. \tag{13}$$

The quantity $\gamma\approx0.5772$ is the Euler–Masheroni constant. In the regime of partial saturation, terms of quartic and higher order in u (that is, terms in the summation) can be ignored.[10]

C. The combination of temporal and spatial behavior

It will be necessary later to have an expression for the phase structure-function density as a function of u and Δt simultaneously. That expression is

$$d(u,\Delta t)=R_t(t)d(u). \tag{14}$$

III. THE PATH-INTEGRAL APPROACH

A. Path space

Consider first a sound source at the origin of an (x,y,z) coordinate system, and a sound receiver at $(R,0,0)$. A general path from the source to the receiver may be defined as $\mathbf{u}(x)$ where $\mathbf{u}$ is a two-dimensional transverse vector having y and z components. The complex sound field at the receiver can be expressed as a sum over all possible paths from the source to the receiver. Because the internal-wave field in the ocean is very anisotropic, with much more rapid variations in the depth coordinate (z) than in either horizontal coordinate, the y component of the paths under consideration can be integrated out, thus leaving only the depth as a function of range, delineated by a single-component $z(x)$. Furthermore, since the range from the source to the receiver is more easily remembered if it is represented by r, we change the path to $z(r)$ where r runs from 0 to R.

In general we are interested in more than just the complex field itself. For example we may be interested in the coherence of the field, represented by the product of two complex fields at two different points. Or we may be interested in the intensity coherence at two different points, which require four complex fields for its expression. If each of these fields were expressed as a path integral, we would require four paths, say, $z_1(r)$, $z_2(r)$, $z_3(r)$, and $z_4(r)$.

However, it turns out that, for intensity coherence, one can integrate out two of the four paths. The first integration comes because one is not dealing with a general fourth moment, and the second comes because of an assumption of quasi-homogeneity. The two paths that are left are linear combinations of the four z-paths mentioned above, so that they may be represented by any letters one likes. We shall use $u(r)$ and $v(r)$ to avoid conflicts with other usage in this paper. Previous papers have made a variety of choices: e.g., $[\beta,\gamma]$,[3] and $[u,w]$.[2]

The general description of paths that are used in the path-integral expressions of interest, and the reasons why the horizontal transverse coordinate can be suppressed, are discussed at length in Chapter 12 of the book by Flatté *et al.*[1]

B. The path-integral

Flatté, Reynolds, and Dashen[3] treated the special case of the intensity coherence in time for acoustic propagation through internal waves. Their resulting two-fold path integral can be expressed in terms of the functions defined so far:

$$\langle I(0)I(\Delta t)\rangle=N\int\int D[u(r)]D[v(r)]\times e^{[ik_0\int_0^R\dot{u}(r)\dot{v}(r)dr]}e^{-M(u(r),v(r),\Delta t)}. \tag{15}$$

Here, k_0 is the acoustic wave number, R is the propagation range, $D[\cdots]$ is the measure of integration over paths $u(r)$ and $v(r)$, and the dot over the variables u and v denotes a derivative with respect to r.

The potential function M contains the effects of internal waves.

The normalizing constant N is defined so that $\langle I(0)I(\Delta t)\rangle$ is unity in the absence of fluctuations, that is, when $M=0$:

$$N^{-1}\equiv\int\int D[u(r)]D[v(r)]e^{ik_0\int_0^R\dot{u}\dot{v}dr}. \tag{16}$$

Like any path integral, Eq. (15) represents a weighted summation over all possible paths connecting source and receiver. The weight is given by a phase factor, which is affected by the curvature of the paths (through the terms $\dot{u}$ and $\dot{v}$) and the presence of a potential $(M(u,v,\Delta t))$. A strength of this formalism is that all possible paths are considered, but only those which are physically important tend to interfere constructively.

The potential function M is defined as

$$M(u,v,\Delta t)=\int_0^R\left[d(u,0)+\left[d(v,\Delta t)-\frac{1}{2}(d(u+v,\Delta t)+d(u-v,\Delta t))\right]\right]dr. \tag{17}$$

The definition of the phase structure-function density as a function of two variables is given in Eq. (14). Note that the internal-wave field is probed at $\Delta t=0$ by the u path, and at Δt by the paths v, $u+v$, and $u-v$. Note also that M is a monotonically increasing function of both u and v, so that for large path separations, e^{-M} grows small and cuts off the integral.

For the calculations in this paper, we ignore the sound channel and we consider only rays that travel horizontally, so that we may take the internal-wave statistical strength to be constant along the entire ray.

IV. CALCULATION OF INTENSITY COHERENCE

A. Strip approximation

In the saturated regions, it can be easily shown[3] that the main contributions to the integral in Eq. (5) come from two regions, which correspond to u or v small. The ''strip approximation'' entails confining the region of integration to two strips, along the u and v axes. We write

$$\langle I(0)I(\Delta t)\rangle = I_u + I_v + \mathbf{O}\left(\frac{1}{\Lambda\Phi^2}\right),$$

$$I_u = N\int\int_{u\ \text{small}} D[u]D[v]e^{ik_0\int_0^R \dot{u}\dot{v}\,dr}e^{-M(u,v,\Delta t)}, \tag{18}$$

$$I_v = N\int\int_{v\ \text{small}} D[u]D[v]e^{ik_0\int_0^R \dot{u}\dot{v}\,dr}e^{-M(u,v,\Delta t)}.$$

We now expand M in powers of $v/u(u/v)$, in order to calculate the v small (u small) integral. After expanding M, the v small (u small) integral can be carried out over the entire u v plane because M will increase quadratically as v (u) increases, leading to an exponential cutoff of the integral away from the strip axis.

$$I_u = I_u^0 + I_u^1 + \ldots, \tag{19}$$

$$I_u^0 = N\int\int D[u]D[v]e^{ik_0\int_0^R \dot{u}\dot{v}\,dr}e^{-M_u^0}, \tag{20}$$

$$I_u^1 = N\int\int D[u]D[v]e^{ik_0\int_0^R \dot{u}\dot{v}\,dr}e^{-M_u^0}[e^{-M_u^1}-1], \tag{21}$$

$$\begin{aligned} M_u^0 &= \int_0^R d(u)\,dr, \\ M_u^1 &= R_t\int_0^R u^2 d''(v)\,dr, \end{aligned} \tag{22}$$

$$I_v = I_v^0 + I_v^1 + \ldots, \tag{23}$$

$$I_v^0 = N\int\int D[u]D[v]e^{ik_0\int_0^R \dot{u}\dot{v}\,dr}e^{-M_v^0}, \tag{24}$$

$$I_v^1 = N\int\int D[u]D[v]e^{ik_0\int_0^R \dot{u}\dot{v}\,dr}e^{-M_v^0}[e^{-M_v^1}-1], \tag{25}$$

$$\begin{aligned} M_v^0 &= \int_0^R\left[\frac{\{\omega^2\}(\Delta t)^2}{2}d(u)+d(v)\right]dr, \\ M_v^1 &= \int_0^R v^2 d''(u)\,dr. \end{aligned} \tag{26}$$

We note at this point that $I_u^0=1$ for all times and $I_v^0=1$ at $t=0$. We therefore replace I_u^0 by unity, and we introduce the normalized time-coherence function $\overline{K}(\Delta t)$ by equating it with I_v^0: $\overline{K}(\Delta t)\equiv I_v^0$. The function $\overline{K}(\Delta t)$ is similar to the function $K(\Delta t)$ defined in Flatté *et al.*,[1] the main difference being that $\overline{K}(\Delta t)$ is normalized to unity at the origin. In the zeroth approximation, then, the intensity coherence function is given by

$$\langle I(0)I(\Delta t)\rangle = 1 + \overline{K}(\Delta t). \tag{27}$$

We see that at $\Delta t=0$, $\langle I^2\rangle = 2$, which is the far asymptotic limit of irradiance variance equal to unity.

Keeping the two lowest-order terms for each strip, we write

$$\langle I(0)I(\Delta t)\rangle = 1 + I_u^1 + \overline{K}(\Delta t) + I_v^1. \tag{28}$$

The treatment in previous work[3] consisted of the zeroth-order approximation to $\langle I(0)I(\Delta t)\rangle$ given in Eq. (27) with the right-hand side multiplied by an *ad hoc* constant, and $\overline{K}(\Delta t)$ approximated in a manner to be described in the next section. In this paper, we present a more accurate calculation of $\overline{K}(\Delta t)$ as well as methods for calculating the two first-order terms I_u^1 and I_v^1 appearing in Eq. (28). These latter methods allow us to deal with the difference between $\langle I(0)I(\Delta t)\rangle$ and 2 in a manner that is more rigorous.

B. Analysis of $\overline{K}(\Delta t)$

Here we describe the various perturbative methods of calculating $\overline{K}(\Delta t)$. For simplicity, we will restrict attention to the case of propagation through a homogeneous internal-wave field in the absence of a sound channel,

$$\begin{aligned} \overline{K}(\Delta t) = N\int\int D[u(r)]D[v(r)]\exp\left[ik_0\int_0^R \dot{u}\dot{v}\,dr\right] \\ \times\exp\left[-\int_0^R\left(d[v(r)]+\frac{\{\omega^2\}(\Delta t)^2}{2}d[u(r)]\right)dr\right], \end{aligned} \tag{29}$$

$$d(u) = R^{-1}\Phi^2\kappa_m^2 u^2\left[\ln\frac{e^{3-2\gamma}}{\kappa_m^2 u^2}\right]. \tag{30}$$

Computation of Eq. (29) would be straightforward if the phase-structure-function density $d(u)$ were quadratic.[12] Each of our methods, including the old, is essentially a perturbative expansion in powers of the difference of $d(u)$ and a suitably chosen quadratic approximation.

We introduce a constant u_0 and expand d as follows:

$$d(u) = R^{-1}\Phi^2\kappa_m^2 u^2\ln\frac{e^{3-2\gamma}}{\kappa_m^2 u_0^2} + p_u(u), \tag{31}$$

$$p_u(u) \equiv R^{-1}\Phi^2\kappa_m^2 u^2\ln\frac{u_0^2}{u^2}. \tag{32}$$

The term $p_u(u)$, and an analogous term $p_v(v)$ (with an analogous constant v_0), will take the role of the small quantity, allowing us to expand $\overline{K}(\Delta t)$ in a perturbation series based on expansion of the exponentials of p_u and p_v in a Taylor series,

$$\overline{K}(\Delta t) = \overline{K^0}(\Delta t) + \overline{K_u} + \overline{K_v} + \ldots\ . \tag{33}$$

To simplify notation, we introduce a weighted path integral

$$\langle\langle f[u(\cdot),v(\cdot),r',\Delta t]\rangle\rangle$$
$$\equiv N\int\int D[u]D[v]f[u,v,r',\Delta t]$$
$$\times\exp\left[ik_0\int_0^R \dot{u}\dot{v}\,dr - R^{-1}\Lambda\Phi^2\right.$$
$$\left.\times\int_0^R\left[v^2\ln\frac{e^{3-2\gamma}}{\kappa_m^2 v_0^2}+\frac{\{\omega^2\}(\Delta t)^2}{2}u^2\ln\frac{e^{3-2\gamma}}{\kappa_m^2 u_0^2}\right]dr\right]. \tag{34}$$

It is helpful to consider $f[u(\cdot),v(\cdot),r',\Delta t]$ both as a function *and* a functional of u and v. In other words, f depends in the usual manner on the functional integration variables $u(r)$ and $v(r)$, and in addition we will treat f as a function of u and v evaluated at a particular range, r'.

Using Eqs. (29), (31), and (34), we write

$$\overline{K^0}(\Delta t)=\langle\langle 1\rangle\rangle,$$
$$\overline{K_u}(\Delta t)=-\frac{\{\omega^2\}(\Delta t)^2}{2}\int_0^R dr'\langle\langle p_u[u(r')]\rangle\rangle, \tag{35}$$
$$\overline{K_v}(\Delta t)=-\int_0^R dr'\langle\langle p_v[v(r')]\rangle\rangle.$$

1. Zeroth-order perturbation terms

Carrying out the quadratic path integrals,[10] we find

$$\overline{K^0}(\Delta t)=\frac{1}{|\cos(\phi)|}, \tag{36}$$

where

$$\phi\equiv 2e^{i3\pi/4}\Phi\sqrt{\Lambda}\left[\frac{\{\omega^2\}(\Delta t)^2}{2}\ln\left(\frac{e^{3-2\gamma}}{\kappa_m^2 u_0^2}\right)\ln\left(\frac{e^{3-2\gamma}}{\kappa_m^2 v_0^2}\right)\right]^{1/4}. \tag{37}$$

2. First-order perturbation terms

$$\overline{K_u}(\Delta t)=-\{\omega^2\}(\Delta t)^2\frac{\Phi^2}{R}\frac{1}{|\cos(\phi)|}$$
$$\times\int_0^R dr'\left[\ln\left[\frac{\kappa_m^2 e^{\gamma-2}}{\epsilon_u(r')}\right]\epsilon_u(r')\right], \tag{38}$$

$$\overline{K_v}(\Delta t)=-2\frac{\Phi^2}{R}\frac{1}{|\cos(\phi)|}\int_0^R dr'\left[\ln\left[\frac{\kappa_m^2 e^{\gamma-2}}{\epsilon_v(r')}\right]\epsilon_v(r')\right], \tag{39}$$

$$\epsilon_u(r')\equiv\frac{\Lambda^2\Phi^2}{\sqrt{2}}\left(\frac{t}{t_I}\right)^{-3/2}\left[\frac{\ln\left(\frac{e^{3-2\gamma}}{\kappa_m^2 v_0^2}\right)}{\ln\left(\frac{e^{3-2\gamma}}{\kappa_m^2 u_0^2}\right)^3}\right]^{1/4}$$
$$\times\Re\left[\frac{\sin(\phi)+\sin\left(\phi\left(1-2\frac{r'}{R}\right)\right)}{e^{i\pi/4}\cos(\phi)}\right], \tag{40}$$

$$\epsilon_v(r')\equiv\frac{1}{4\sqrt{2}\Phi^2}\left(\frac{t}{t_I}\right)^{1/2}\left[\frac{\ln\left(\frac{e^{3-2\gamma}}{\kappa_m^2 u_0^2}\right)}{\ln\left(\frac{e^{3-2\gamma}}{\kappa_m^2 v_0^2}\right)^3}\right]^{1/4}$$
$$\times\Re\left[\frac{\sin(\phi)+\sin\left(\phi\left(1-2\frac{r'}{R}\right)\right)}{e^{i\pi/4}\cos(\phi)}\right]. \tag{41}$$

3. Determination of u_0,v_0

We present three methods for determining the values of u_0 and v_0: the old, the mean-field, and the new approximations. The old method was introduced by Flatté, Reynolds, and Dashen.[3]

Old approximation. The ''old'' approximation involves order-of-magnitude estimates of the parameters u_0 and v_0 based on the region of the dominant contribution to the integrand of Eq. (29). Once chosen, these parameters are put into the expression for $\overline{K^0}(\Delta t)$ of Eq. (36), leading to the desired approximation to $K(\Delta t)$. The values of u_0 and v_0 are found to be[3]

$$v_0^2=\frac{\overline{R_f}^2}{\Lambda\Phi^2},\quad u_0^2=4\overline{R_f}^2\Lambda\Phi^2\ln(\Phi^2 e^{3-2\gamma}). \tag{42}$$

Mean-field approximation. The mean-field approximation is conceptually identical to the old approximation. However, the emphasis is on finding an approximation valid for larger time separations, and on allowing one of these parameters to vary with the time Δt,

$$v_0^2=\frac{\overline{R_f}^2}{\Lambda\Phi^2\ln(e^{3-2\gamma}\Phi^2)},\quad u_0^2=\frac{t_I^2}{(\Delta t)^2}\frac{4\overline{R_f}^2\Lambda\Phi^2}{\ln\left(\frac{e^{3-2\gamma}}{\Lambda^2\Phi^2}\right)}. \tag{43}$$

New approximation. In the new approximation, u_0 and v_0 are chosen by constraining the first-order perturbation terms to vanish: $\overline{K_u}=\overline{K_v}=0$. These constraints turn out to be integral equations.

$$\int_0^R dr'\ln\left[\frac{\kappa_m^2 e^{\gamma-2}}{\epsilon_u(r')}\right]\epsilon_u(r')=0,$$
$$\int_0^R dr'\ln\left[\frac{\kappa_m^2 e^{\gamma-2}}{\epsilon_v(r')}\right]\epsilon_v(r')=0. \tag{44}$$

C. Analysis of first-order terms I_u^1 and I_v^1

1. Calculation of I_u^1

It is noted that the sum $I_u^0+I_u^1$ is an integral exactly analogous to that given in Eqs. (14) and (15) of Dashen and Wang,[13] except that β (or g) now contains a factor of R_t. As a consequence, we are able to estimate[10] the value of I_u^1 as $\Delta/2$ where Δ is the experimentally observed difference between $\langle I^2\rangle$ and 2.

2. Calculation of I_v^1

(a) $\Delta t\ll t_I$. In this region we use an *ad hoc* interpolation between the point at $\Delta t=0$, which we take from the ex-

perimentally observed $\langle I^2\rangle$, and the calculations of the next section, which give the result at $\Delta t = t_I$ and larger.

(b) $\Delta t \gtrsim t_I$. For $\Delta t \gtrsim t_I$, the integrand of Eq. (25) is cut off when the M^0 exponent grows to order unity. The M^1 term remains small, so it is appropriate to expand out the term in brackets. We introduce the notation $h(\Delta t)$ for this large Δt approximation for I_v^1.

$$h(\Delta t) \equiv \int_0^R dr' \int\int D[u]D[v] e^{ik_0 \int_0^R \dot{u}\dot{v}\, dr} e^{-M^0(u,v,\Delta t)} \times [v(r')^2 d''[u(r')]]. \tag{45}$$

We analyze Eq. (45) with the techniques of Sec. IV B. The values of u_0 and v_0 used to approximate $K(\Delta t)$ can be used for approximating $h(\Delta t)$ because the support of the integral in Eq. (45) is substantially similar to that of Eq. (29).

To simplify the expression for $h(t)$, we introduce the substitutions,

$$d''[u] = \frac{2\Lambda\Phi^2}{R} \int_{-\infty}^{\infty} dq \frac{\Theta(|q|-1)}{q} e^{-iq\kappa_m u},$$
$$v^2 = -\frac{1}{\kappa_m^2} \int_{-\infty}^{\infty} dq\, \delta''(q) e^{-iq\kappa_m v}. \tag{46}$$

The first equality above is obtained by differentiating both sides of Eq. (12) with respect to Δz and substituting in the expression for $P(\kappa)$ of Eq. (8). The second equality follows from the properties of the Dirac delta function. Equation (45) can be expressed in terms of the bracket notation of Sec. IV B,[10]

$$h(\Delta t) = -R_t \frac{\Phi^2}{R} \int_0^R dr' \int_{-\infty}^{\infty} dq_1 \frac{\Theta(|q_1|-1)}{q_1} \int_{-\infty}^{\infty} dq_2 \delta''(q_2) \times \langle\langle \exp[-iq_1\kappa_m u(r') - iq_2\kappa_m v(r')]\rangle\rangle. \tag{47}$$

Carrying out this quadratic path integral, we find[10]

$$\langle\langle \exp[-iq_1\kappa_m u(r') - iq_2\kappa_m v(r')]\rangle\rangle = \frac{1}{|\cos(\phi)|} e^{-\epsilon_u q_1^2 - \epsilon_v q_2^2 - 2i\eta_{uv} q_1 q_2}, \tag{48}$$

where the quantities ϕ, ϵ_u, and ϵ_u have been defined in Eqs. (37), (40), and (41), respectively. The quantity η_{uv} is defined by

$$\eta_{uv}(r') \equiv \frac{\sqrt{\Lambda}}{4\Phi} \frac{\mathcal{I}\left[\dfrac{2\sin\left[\phi\left(1-\dfrac{r'}{R}\right)\right]\cos\left(\phi\dfrac{r'}{R}\right)}{e^{i\pi/4}\cos(\phi)}\right]}{\left[\ln\left(\dfrac{e^{3-2\gamma}}{\kappa_m^2 v_0^2}\right)\ln\left(\dfrac{e^{3-2\gamma}}{\kappa_m^2 u_0^2}\right)\dfrac{\{\omega^2\}(\Delta t)^2}{2}\right]^{1/4}}. \tag{49}$$

The symbol $\mathcal{I}[X]$ denotes the imaginary part of X.

Keeping terms of first order in the small parameters, we expand Eq. (48) in a Taylor series, and retain the lowest order expression for h,

$$h(\Delta t) \approx \frac{2R_t\Phi^2}{|\cos(\phi)|R} \times \int_0^R dr' \left\{ \epsilon_v(r')[-\ln[\epsilon_u(r')] - \gamma] + 2\frac{\eta_{uv}^2(r')}{\epsilon_u(r')} \right\}. \tag{50}$$

V. SIMULATION OF ACOUSTIC PROPAGATION

We have implemented a numerical simulation of acoustic propagation through internal waves to provide a basis for comparison with the theory. The simulation uses a parabolic-equation-based algorithm to propagate a plane wave through phase screens which mimic internal-wave behavior.[14,15]

A. Propagation algorithm

It has been remarked by Flatté *et al.* that the parabolic-equation simulation method uses exactly the same expressions that Feynman used to introduce the path-integral formalism.[1] Feynman wrote the path integral in terms of phase screens, and then took the limit in which the number of screens goes to infinity. We use a parabolic-equation-based approach with a finite number of screens to simulate the model described in Sec. I. See Jensen *et al.*[16] for a review of the literature on parabolic-equation methods.

B. Simulation parameters

The length scales which must be specified are the Fresnel radius R_f and the length scales associated with the power spectral density of the propagating medium. In other words, a given simulation corresponds to an infinite set of range-frequency combinations (i.e., those which determine the specified Fresnel radius).

The functional form of the power spectrum of the phase screens must be specified. For all simulations presented in this work, the power spectral density of the phase screen is assumed separable in time and space. The spatial power spectrum is given by Eq. (8), and the temporal power spectrum is chosen to be consistent with Eq. (3).

C. Sampling requirements

Martin and Flatté have carefully discussed spatial sampling constraints with particular regard to adequate representation of the entire dynamical range of the intensity fluctuations.[17]

Minimal sampling requirements were first estimated by the methods of Martin and Flatté[17] and then confirmed by convergence testing. For the simulation results presented in this paper with $\Lambda\Phi^2 = 40.0$ and $\Lambda\Phi = 0.2$, the Fresnel radius was taken to be 120 in grid units. To obtain $\Lambda = 1 \cdot 10^{-1}$, the outer scale was approximately 26 000 grid units, and the spatial dimension of the simulation was $2^{17} = 131\,072$. Since the inner scale is given by the spacing of the grid, there are approximately 26 000 modes representing the internal-wave field. These parameters represent an oversampling, but this was found to lead to fast statistical convergence of the simulation.

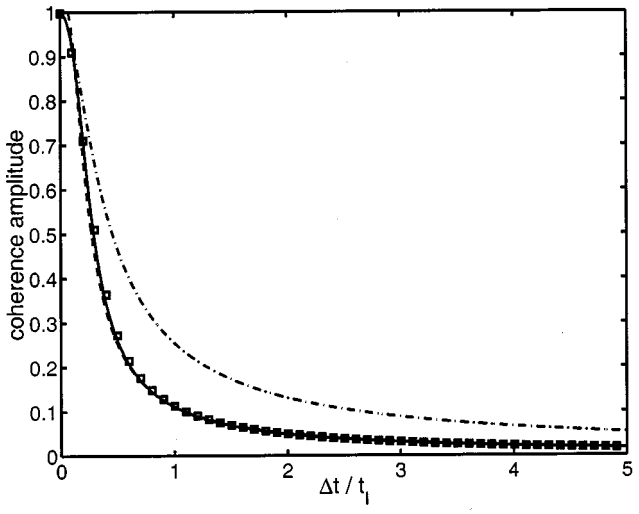


FIG. 1. Approximations to $\bar{K}(\Delta t)$, plotted with the numerical-integration result in the single phase-screen case with $\Lambda\Phi^2=40$ and $\Lambda\Phi=0.2$. The squares represent the results of a numerical integration of Eq. (29). The solid curve is the approximation to $\bar{K}(\Delta t)$ obtained in the ''new'' approach, the dashed curve is the ''mean-field'' approximation, and the dash-dot curve is the ''old'' result.

VI. SINGLE-SCREEN COMPARISONS

A. Perturbative calculation of $\bar{K}(\Delta t)$

Results which indicate the validity of the perturbative calculation of $\bar{K}(\Delta t)$ outlined in Sec. IV B will be summarized here.

Figure 1 shows the normalized time-coherence function $\bar{K}(\Delta t)$ for various approximations, for the case of a single-phase screen. In this situation, where we calculate $\bar{K}$ rather than $\langle I(0)I(\Delta t)\rangle$, it is possible to obtain the exact result by numerical integration of Eq. (29). It is seen that the approximations that incorporate something of the new approach are excellent, while the old approximation technique misses by as much as a factor of two at particular times.

Figure 2 provides an indication of the size of the perturbation terms. The sum of the first- and second-order perturbation terms, divided out by the zeroth-order term, is plotted for the various approximations. It is seen that the new approximation provides a very accurate estimate of the normalized time-coherence function $\bar{K}(\Delta t)$.

Figure 3 shows the behavior of u_0 and v_0 as a function of time, for the various approximations. The ability of the newer approximations to allow u_0 and v_0 to differ for different time intervals in the coherence function is important to their success.

B. Intensity coherence in time

We present two theoretical results for $\langle I(0)I(\Delta t)\rangle$ and compare them with simulation data. Both of these theoretical results make use of the new approximation calculation of $\bar{K}(\Delta t)$. The first theory is obtained from the first-order expression given in Eq. (28), with the results of Secs. IV B and IV C used to replace $\bar{K}(\Delta t)$, I_u^1, and I_v^1. However, these equations are valid only for $\Delta t > t_I$. In order to provide a result for small Δt, we use a linear combination of $\bar{K}(\Delta t)$

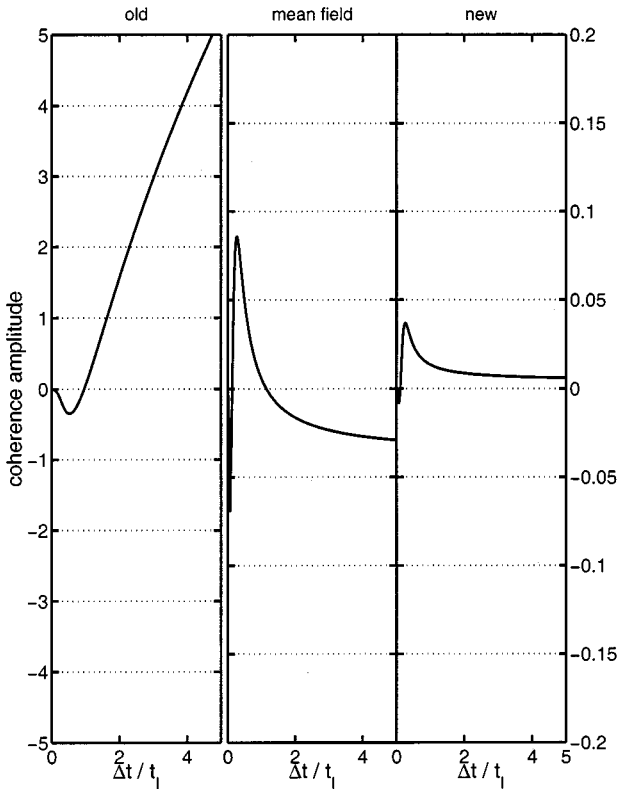


FIG. 2. Sum of first- and second-order perturbation terms divided by the zeroth-order result for the single-screen case with $\Lambda\Phi^2=40$ and $\Lambda\Phi=0.2$. Note that the ordinate axis for the old approximation is scaled by a factor of 25.

and a constant. The coefficient of $\bar{K}(\Delta t)$ and the constant are determined by $\langle I(0)I(\Delta t)\rangle$ at $\Delta t=t_I$ (calculated from our equations) and at $\Delta t=0$ (determined empirically from experimental data or from simulations). The second theory is the zeroth-order result of Eq. (27) multiplied by $\langle I^2\rangle/2$ (again determined from experimental data or from simulations) so that there is agreement at the origin. Figure 4 shows the simulation results compared with these two theories. The agreement is excellent.

VII. CONTINUOUS-CASE COMPARISONS

A. Perturbative calculation of $\bar{K}(\Delta t)$

Results which indicate the validity of the perturbative calculation of the normalized time-coherence function outlined in Sec. IV B will be presented here.

Figure 5 shows $\bar{K}$ for various approximations, for the continuous case. In this situation it is not possible to obtain the exact result by numerical integration. However, we have confidence in the analytic calculations because of the agreement seen in the single-screen case, and because of the agreement between simulation and analytics that we will show later for $\langle I(0)I(\Delta t)\rangle$. It is seen that all the approximations that incorporate something of the new approach give very closely the same result, while again the old approximation technique misses this presumably correct answer by as much as a factor of two at particular times.

Figure 6 shows the normalized sum of perturbation terms through second order for the various approximations.

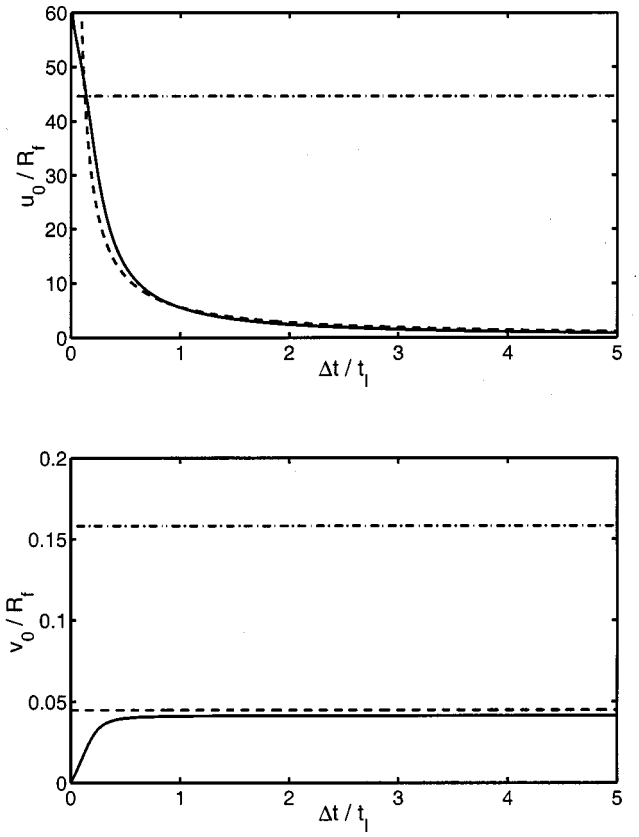


FIG. 3. Behavior of u_0 and v_0, the logarithmic constants, for the single phase-screen case with $\Lambda\Phi^2=40$ and $\Lambda\Phi=0.2$. The values are represented for the old, mean-field, and new approximations by dash-dot, dashed, and solid lines, respectively.

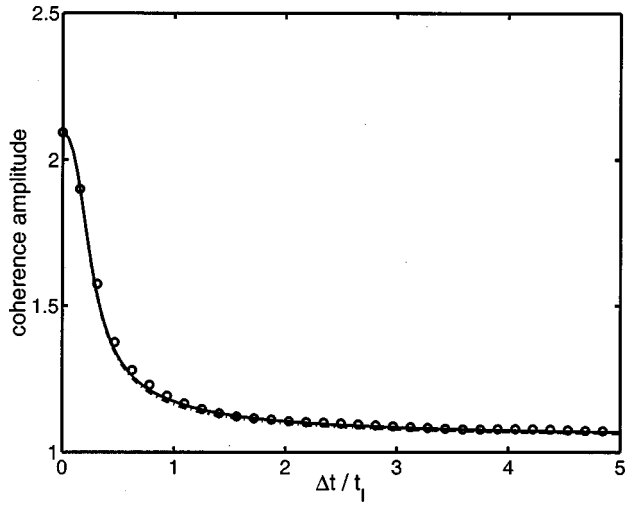


FIG. 4. Simulated intensity coherence function in time $\langle I(0)I(\Delta t)\rangle$ for $\Lambda\Phi^2=40$ and $\Lambda\Phi=0.2$ plotted against theoretical results presented in the text for the single phase-screen case. The simulation data are marked with the circles. The dash-dot curve is the zeroth-order result for $\langle I(0)I(\Delta t)\rangle$ [Eq. (27)] multiplied by a constant to force agreement at the origin. The solid curve is the first-order result for $\langle I(0)I(\Delta t)\rangle$ [Eq. (28)]. This result is calculated explicitly for $\Delta t > t_I$ and approximated with a linear function of the zeroth-order result for $\Delta t < t_I$. This approximation is constrained to agree with simulation data at the origin and with the calculated value of $\langle I(0)I(t_I)\rangle$.

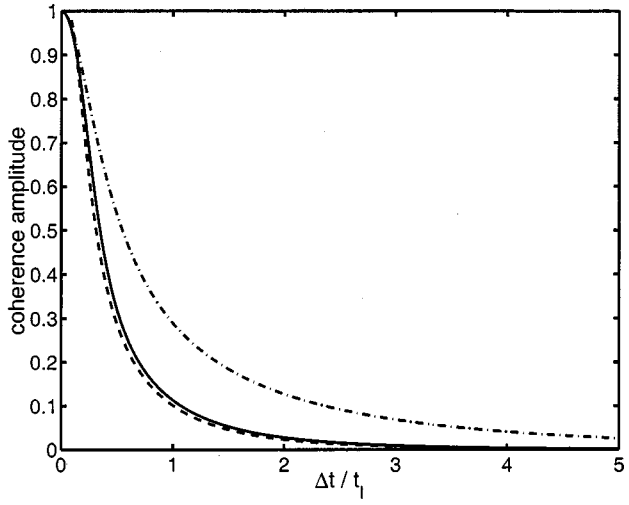


FIG. 5. Approximations to $\bar{K}(\Delta t)$ in the continuous case with $\Lambda\Phi^2=40$ and $\Lambda\Phi=0.2$. The solid curve is the approximation to $\bar{K}(\Delta t)$ obtained in the ''new'' approach, the dashed curve is the ''mean-field'' approximation, and the dash-dot curve is the ''old'' result.

It is seen that the new approximation is quite accurate for all relevant times.

Figure 7 shows the behavior of u_0 and v_0 as a function of time, for the various approximations. Again, the ability of the newer approximations to allow u_0 and v_0 to differ for different time intervals in the coherence function is important to their success.

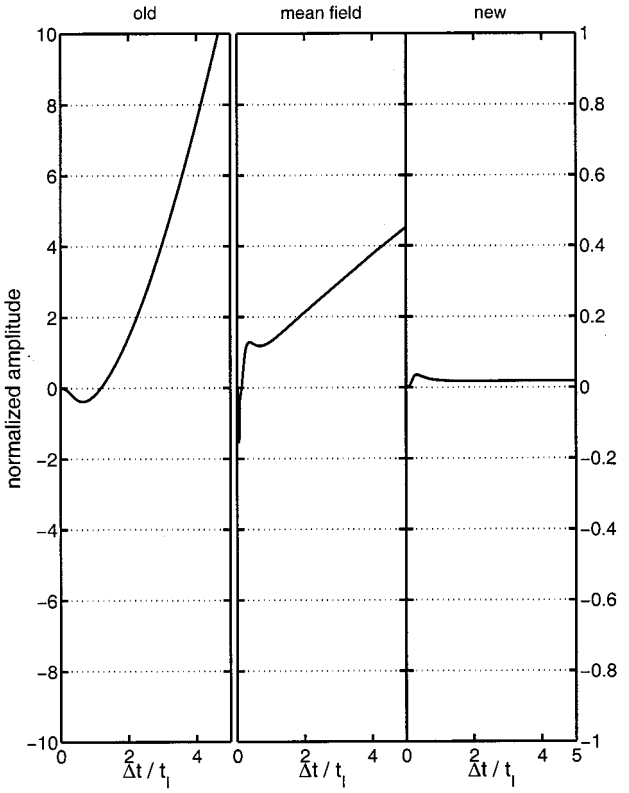


FIG. 6. Sum of first- and second-order perturbation terms divided by the zeroth-order result for the continuous case with $\Lambda\Phi^2=40$ and $\Lambda\Phi=0.2$. Note that the ordinate axis for the old approximation is scaled by a factor of 10.

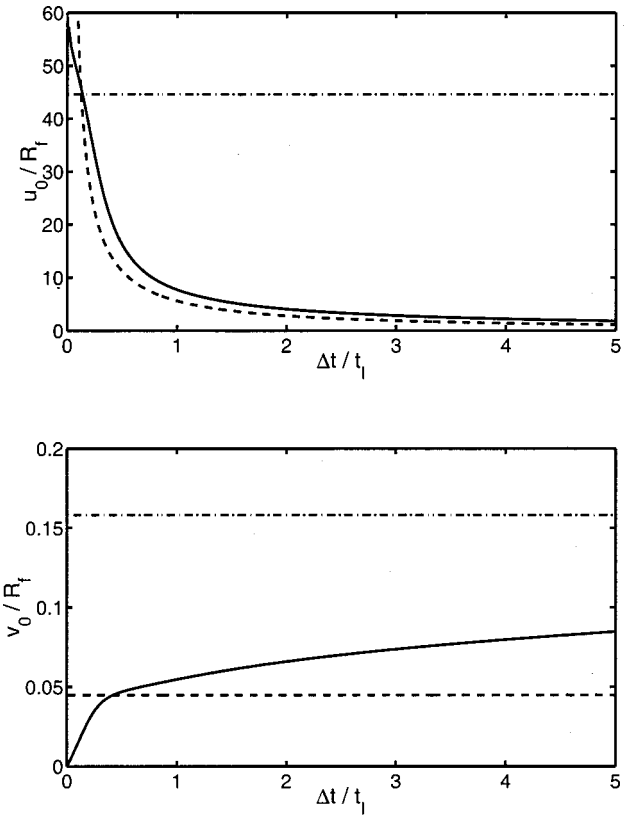


FIG. 7. Behavior of u_0 and v_0, the logarithmic constants, for the continuous case with $\Lambda\Phi^2=40$ and $\Lambda\Phi=0.2$. The values are represented for the old, mean-field, and new approximations by dash-dot, dashed, and solid lines, respectively.

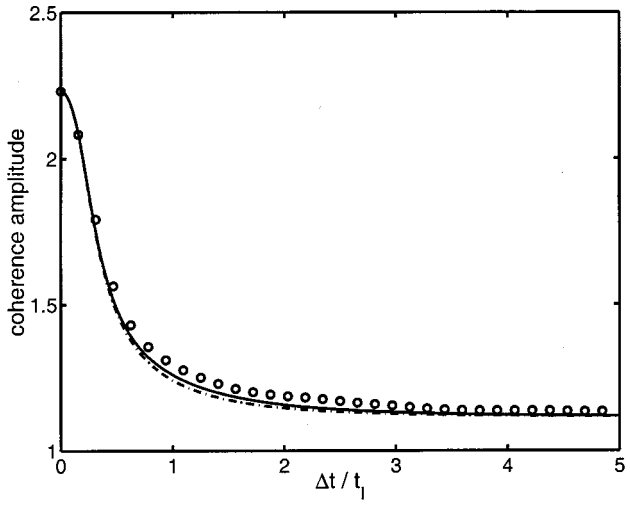


FIG. 8. Simulated intensity coherence function in time $\langle I(0)I(\Delta t)\rangle$ for $\Lambda\Phi^2=40$ and $\Lambda\Phi=0.2$ plotted against theoretical results presented in the text for the continuous case. The simulation data are marked with the circles. The dot-dashed curve is the zeroth-order result for $\langle I(0)I(\Delta t)\rangle$ [Eq. (27)] multiplied by a constant to force agreement at the origin. The solid curve is the first-order result for $\langle I(0)I(\Delta t)\rangle$ [Eq. (28)]. This result is calculated explicitly for $\Delta t > t_I$ and approximated with a linear function of the zeroth-order result for $\Delta t < t_I$. This approximation is constrained to agree with simulation data at the origin and with the calculated value of $\langle I(0)I(t_I)\rangle$.

B. Intensity coherence in time

This section will present a comparison of the results of the simulation with the theoretical calculation of the intensity coherence in time $\langle I(0)I(\Delta t)\rangle$ [which includes the calculation of the normalized time-coherence function $K(\Delta t)$].

We present two theoretical results for $\langle I(0)I(\Delta t)\rangle$. These are the same two theories that are explained in the single-screen case (Sec. VI B). Figure 8 shows the simulation results compared with these two theories. The agreement is quite good.

VIII. COMPARISON WITH THE AFAR EXPERIMENT

We use the results summarized above to predict the observed intensity coherence function in time for a transmission from the Azores Fixed Acoustic Range (AFAR) experiment.[18,19] We focus specifically on the 35-km transmission centered at 1010 Hz, which is found to be in the partially saturated regime. Reynolds *et al.* also found the statistics of the sound field to be nominally fit by the Garrett–Munk internal-wave formulation with the following parameter values: $n_0 B = 3.0$ cph-km, $\zeta_0 = 13.1$ m, and $j_* = 3$, and determined the fluctuation parameters, using the definitions given in Flatté *et al.*,[1] as $\Lambda\Phi^2 = 3.0$ and $\Lambda\Phi = 0.17$.[20]

In order to compare the experimental results with the theoretical predictions made above, we proceed from the assumption that the environment is homogeneous along the ray. The average Fresnel radius is found to be $\sqrt{R/6k_0} = 37$ m. Using Eq. (12) we find our spectral cutoff to be $\kappa_m = 2.7 \cdot 10^{-3}\,\mathrm{m}^{-1}$. We determine the average internal-wave frequency by equating ν^2 given in Reynolds *et al.*[4] with $\Phi^2\{\omega^2\}(\Delta t)^2$, and we find $\{\omega^2\} = 1.28\,\mathrm{h}^{-1}$.

The predictions for the time-coherence function made using these numbers are plotted in Fig. 9 against the AFAR experimental data.[4] It can be seen that the decoherence time predicted by the mean-field approximation is a factor of 2 larger than that indicated by the data, whereas the old approximation predicts a decoherence time a factor of 4 larger.

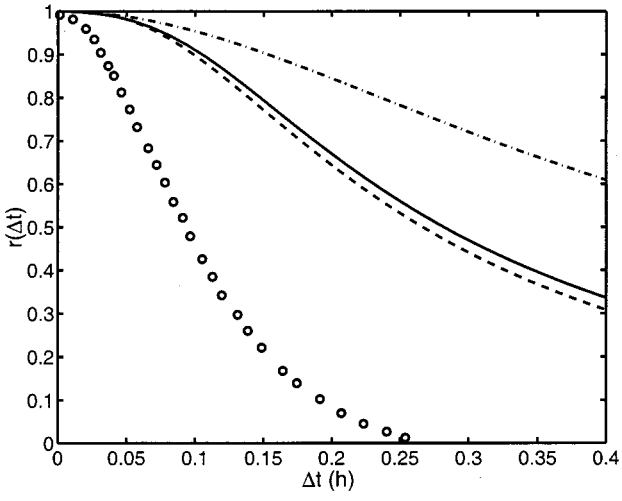


FIG. 9. Intensity coherence function for 35-km AFAR transmission at 1010 Hz. AFAR data from Flatté *et al.* (Ref. 4) are shown as open circles. Theoretical curves represent calculations of the time-coherence function $K(\Delta t)$ in the old (dash-dot line), new (solid line), and mean-field (dashed line) approximations with the assumption of no waveguide and homogeneous internal waves.

This discrepancy can be understood in light of the importance of the sound channel to the experimental analysis. Flatté *et al.* showed that changing just the algorithm used to smooth the experimental sound channel can change the predicted decoherence time by a substantial factor.[4]

The substantial difference between the predictions of the old and newer approximations suggests that it will be worthwhile to predict the time-coherence function using the methods presented here in conjunction with the actual sound channel of the AFAR experiment. This is proposed as an avenue of future research.

IX. SUMMARY

We have developed an analytic method for evaluating the temporal coherence function of acoustic intensity for propagation through ocean internal waves. We have shown by comparison with simulations using the parabolic equation that our analytic method is accurate within a few percent for cases in partial saturation. Our demonstrations and first calculations have been done without a sound channel, and this has prevented us from getting good agreement with (the AFAR) experiment at this time. However, even without a sound channel, the method is useful for predicting rough values of intensity coherence times for a wide variety of ocean cases. We expect to be able to generalize to the sound-channel case in the near future.

ACKNOWLEDGMENTS

This work has been partially supported by the Ocean Acoustics office (Code 321 OA) of the Office of Naval Research. We are grateful for a grant from the W. M. Keck Foundation. The late Roger Dashen contributed to early work on this project. This is Contribution Number 405 of the Institute of Tectonics, and was submitted in partial satisfaction of the Ph.D. requirement of J. Gerber.

[1] S. Flatté, R. Dashen, W. Munk, K. Watson, and F. Zachariasen, *Sound Transmission through a Fluctuating Ocean* (Cambridge University Press, Cambridge, 1979).

[2] S. M. Flatté, ''Wave propagation through random media: Contributions from ocean acoustics,'' Proc. IEEE **71**, 1267–1294 (1983).

[3] S. Flatté, S. Reynolds, and R. Dashen, ''Path-integral treatment of intensity behavior for rays in a sound channel,'' J. Acoust. Soc. Am. **82**, 967–972 (1987).

[4] S. Flatté, S. Reynolds, R. Dashen, B. Buehler, and P. Maciejewski, ''AFAR measurements of intensity and intensity moments,'' J. Acoust. Soc. Am. **82**, 973–980 (1987).

[5] C. Garrett and W. H. Munk, ''Space-time scales of internal waves,'' Geophys. Fluid Dyn. **3**, 225–264 (1972).

[6] C. Garrett and W. H. Munk, ''Space-time scales of internal waves: A progress report,'' J. Geophys. Res. **80**, 291–297 (1975).

[7] W. Munk and C. Garrett, ''Internal waves in the ocean,'' Annu. Rev. Fluid Mech. **11**, 339–369 (1979).

[8] R. Esswein and S. Flatté, ''Calculation of the phase-structure function density from oceanic internal waves,'' J. Acoust. Soc. Am. **70**, 1387–1396 (1981).

[9] S. M. Flatté and R. B. Stoughton, ''Theory of acoustic measurement of internal wave strength as a function of depth, horizontal position, and time,'' J. Geophys. Res. **91**, 7709–7720 (1986).

[10] J. S. Gerber, ''Calculation by perturbative methods of the time coherence function of acoustic intensity from the effects of internal waves,'' Ph.D. thesis, University of California at Santa Cruz, 1999.

[11] S. Flatté and G. Rovner, ''Calculations of internal-wave-induced fluctuations in ocean-acoustic propagation,'' J. Acoust. Soc. Am. **108**, 526–534 (2000).

[12] R. Feynman and A. Hibbs, *Quantum Mechanics and Path Ingetrals* (McGraw-Hill, New York, 1965).

[13] R. Dashen and G.-Y. Wang, ''Intensity fluctuation for waves behind a phase screen: A new asymptotic scheme,'' J. Opt. Soc. Am. A **10**, 1219–1225 (1993).

[14] R. H. Hardin and F. Tappert, ''Application of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations,'' SIAM Rev. **15**, 423 (1973).

[15] S. Flatté and F. Tappert, ''Calculation of the effect of internal waves on oceanic sound transmission,'' J. Acoust. Soc. Am. **58**, 1151–1159 (1975).

[16] F. Jensen, W. Kuperman, M. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 2000).

[17] J. Martin and S. Flatté, ''Intensity images and statistics from numerical simulation of wave propagation in 3-D random media,'' Appl. Opt. **27**, 2111–2126 (1988).

[18] A. W. Ellinthorpe, ''The Azores Range,'' Technical Report No. 4551, NUSC, New London, CT, 1973.

[19] B. G. Buehler, ''Volume propagation experiments at the Azores fixed acoustic range,'' Technical Report No. 5785, NUSC, New London, CT, 1979.

[20] S. Reynolds, S. Flatté, R. Dashen, B. Buehler, and P. Maciejewski, ''AFAR measurements of acoustic mutual coherence functions of time and frequency,'' J. Acoust. Soc. Am. **77**, 1723–1731 (1985).

Time-reversing array retrofocusing in noisy environments

Sunny R. Khosla and David R. Dowling[a)]
Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109-2121

(Received 9 August 1999; revised 11 August 2000; accepted 13 November 2000)

Acoustic time reversal is a robust means of retrofocusing acoustic energy, in both time and space, to the original sound-source location. However, noise may limit the performance of a time-reversing array (TRA) at long source–array ranges, or when the original-source or TRA-element power levels are low. The operation of a TRA requires two steps (reception and transmission) so both TRA-broadcast noise and ambient noise must be taken into account. In this paper, predictions are made for how a simple omnidirectional noise field influences the probability that the signal amplitude from a narrow-band TRA will exceed the noise at the TRA's retrofocus. A general formulation for the probability of TRA retrofocusing, which can be used for TRA design, is developed that includes: the variance of the noise field, the original source strength, the TRA's element output power, the number of TRA elements (N), and the propagation characteristics of the environment. This formulation predicts that a TRA's array gain (in dB) at the retrofocus may be as high as $+10\log_{10}(N)$ to $+20\log_{10}(N)$ depending on the relative strengths of the original source and the TRA's elements. Monte Carlo simulations in both a free-space environment and a shallow-ocean sound-channel environment compare well to this probability formulation even when simple approximate parametric relationships for the appropriate Green's functions are used. The dominant deviation between theory and simulation in the sound channel is caused by acoustic absorption. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1338560]

PACS numbers: 43.30.Vh, 43.30.Yj, 43.60.Cg, 43.30.Nb [DLB]

I. INTRODUCTION

In the past decade, acoustic time reversal has been studied as a solution to a wide range of focusing problems involving complex and unknown media. Recent work in underwater acoustics (Kuperman *et al.*, 1998; Song *et al.*, 1998; Khosla and Dowling, 1998; Hodgkiss *et al.*, 1999; Song *et al.*, 1999; Roux and Fink, 2000; Dungan and Dowling, 2000), and ultrasound (Fink, 1997; Tanter *et al.*, 1998; Prada and Fink, 1998; Draeger and Fink, 1999a, b; Mordant *et al.*, 1999) illustrate this alluring technique's capabilities and recent extensions.

For example, two-way underwater communication systems conform to the operational requirements of TRAs and might benefit from their special properties. Consider communication between a surface ship and an untethered submerged vehicle or instrumentation package which may be several or several tens of kilometers away. A complete round-trip acoustic transmission involves an initial signal broadcast from the submerged device, signal reception and decoding at the surface ship, a response signal transmitted from the ship, and final signal reception back at the submerged device. This type of underwater acoustic data transmission is limited primarily by bandwidth constraints, ambient noise, transducer broadcast power, and by the complexities of underwater sound propagation including multipath propagation and random variability arising from time-varying ocean surface and water column properties (Catipovic, 1997).

The data rate and/or signal-to-noise ratio in this communication scenario can be enhanced if a time-reversing array (TRA) is used by the surface ship (or the submerged device) because a TRA can compensate for multipath propagation. This compensation is based on a TRA's inherent ability to spatially and temporally retrofocus sound in unknown complex acoustic environments by exploiting all available propagation paths between the source and the array (Jackson and Dowling, 1991; Dowling, 1994; Kuperman *et al.*, 1998; Roux and Fink, 2000).

Such compensation may be lost when the effective TRA aperture is too small and is typically degraded in dynamic media at a rate that depends on the source–array range, the acoustic frequency, and the characteristics of the environment (Dowling and Jackson, 1992; Dowling, 1993, 1994; Khosla and Dowling, 1998; Dungan and Dowling, 2000). However, little or no information is available on the limits of TRA performance in noisy environments. The present paper discusses the limitations imposed on the narrow-band performance of a TRA by a simple ambient noise field. Complementary results are provided by Yoo and Yang (1998) on signal-processing performance improvements for receive arrays in the presence of ambient noise, and by Baggeroer *et al.* (1988), who studied the influence of noise on the Bartlett matched-field processor—which is computationally similar to acoustic time reversal—and evaluated its Cramer–Rao bound.

Aspects of narrow-band TRA performance include retrofocus amplitude, retrofocus size, and the presence or absence of spatial sidelobes (spurious retrofoci). Here, the main indication of TRA performance will be the relative amplitudes of signal and noise at the retrofocus location. A field-amplitude-based performance description was chosen because it is the most important for the envisioned underwater

[a)]Author to whom correspondence should be addressed. Electronic mail: drd@engin.umich.edu

communications applications of this technology with stationary or slowly moving sources and arrays. Here, the intended receiver lies at or near the retrofocus location (the intended focus) so it does not listen to or become confused by spatial sidelobes (unintended foci). Thus, spatial sidelobes need only be considered for this communication application if the transmitted signals must have a low probability of interception. In addition, our investigations showed that the TRA retrofocus size lacked any trend with increasing noise level and was instead primarily governed by the acoustic frequency and the characteristics of the environment. Computational results for TRA retrofocus size (and peak-to-sidelobe ratios) are provided in Dungan and Dowling (2000).

The main goal of this paper is to provide a theoretical framework along with some specific performance predictions for TRAs in noisy free-space and shallow-ocean sound-channel environments. Time-varying environments are not considered. The framework leads to a general analytical scaling law for the probability (φ) that the intended signal amplitude exceeds the noise amplitude at the TRA's retrofocus and includes the noise field variance (σ_n^2), the original source strength (M_s), the broadcast power of the TRA's elements (Π_o), the number of TRA elements (N), the source–array range (R), and other propagation characteristics of the environment. The utility of the formal results lies in their capacity to separate the regions in parameter space where a TRA will, and will not, reliably work. For example, the 50%-probability-of-retrofocus contour may adequately define such a boundary for preliminary TRA design or application-selection purposes. However, the framework developed here is sufficiently flexible to allow lower or higher probabilities.

An omnidirectional noise field was used in both of the acoustic environments considered in this investigation, even though it may not accurately represent the noise field in a shallow-ocean sound channel. The hope here is that results obtained with such a simple noise field will be broadly indicative of TRA performance for any noise field. There were several reasons for this selection, the main one being simplicity; the entire noise field is characterized by one parameter, σ_n^2. Additionally, the effects of nonuniform or directional noise would best be addressed after the omnidirectional noise results are established as a basis for comparison. Moreover, under actual operational conditions, known directional noise sources could be removed from the TRA output by adaptive-beamformer nulling techniques (Steinberg, 1976; Yoo and Yang, 1998). And finally, omnidirectional noise may, in fact, present the most serious challenge to the array because a portion of the noise source always appears to be collocated with the original sound source.

The remainder of this paper is organized into four sections. A description of the mathematical framework for acoustic time reversal with ambient noise is discussed in the next section. Here, the necessary development and definitions for the TRA's amplification factor, the received signal-to-noise ratio at the array, and the signal-to-noise ratio at the retrofocus are provided. These definitions are then used to generate a general scaling law for the probability of retrofocusing. Section III presents the results of a study of time reversal in a noisy free-space environment. The results from a similar investigation for a shallow-ocean sound channel are described in Sec. IV. The final section summarizes the findings and states the conclusions drawn from this research.

II. MATHEMATICAL FORMULATION

This section presents the formal development of how noise affects a time-reversing array. The operating cycle for a TRA involves two essential steps. First, the acoustic waves emanating from a sound source propagate to and are recorded by the TRA along with the ambient noise at the array. Second, this recording of signal plus noise is time reversed, scaled by an amplification factor, and retransmitted into the acoustic environment. As previously mentioned, some extra signal processing or analysis may take place as part of the second step. Ambient noise creates at least three challenges for a TRA. First, the array always rebroadcasts some (or all) of the noise it receives and this introduces additional noise into the environment which can obscure the retrofocus. Second, the maximum possible signal amplification is lower in noisy environments because the array wastes power rebroadcasting noise. And third, at long ranges, ambient noise may overwhelm an otherwise acceptable retrofocus field. All these factors are accounted for in the following development.

For consistency with previous work (Jackson and Dowling, 1991), a discrete version of the time-invariant, narrow-band weighted-monopole formulation for the time-reversal process was chosen as the starting point

$$P_{\mathrm{TRA}}(\mathbf{r}_f,\mathbf{r}_s,\omega)=\sum_{i=1}^{N}\{M_s G(\mathbf{r}_i,\mathbf{r}_s,\omega)+n(\mathbf{r}_i)\}^* \times A G(\mathbf{r}_f,\mathbf{r}_i,\omega)+n(\mathbf{r}_f). \quad (1)$$

Here, $P_{\mathrm{TRA}}(\mathbf{r}_f,\mathbf{r}_s,\omega)$ is the TRA's pressure-field response at the field point $\mathbf{r}_f$ due to a point source located at $\mathbf{r}_s$, $\mathbf{r}_i$ denotes the array-element locations, ω is the acoustic frequency in radians per second, $G(\mathbf{r}_2,\mathbf{r}_1,\omega)$ is the Helmholtz-equation Green's function at frequency ω between locations $\mathbf{r}_1$ to $\mathbf{r}_2$, $n(\mathbf{r}_i)$ and $n(\mathbf{r}_f)$ are the complex random noise field at the array-element and field-point locations, respectively, A is the array's amplification factor, and M_s is the source strength. It is related to the source level: $\mathrm{SL}=20\log_{10}\{M_s/\sqrt{2}P_{\mathrm{ref}}(1m)\}$, where P_{ref} is 1 μPa. The total field received by an array element appears in $\{\ \}$ braces in (1) and consists of the signal from the source, $M_s G(\mathbf{r}_i,\mathbf{r}_s,\omega)$, plus the local ambient noise, $n(\mathbf{r}_i)$. This total field is then complex conjugated (complex conjugation is equivalent to time reversal for narrow-band signals; Jackson and Dowling, 1991) and multiplied by the array amplification factor prior to retransmission from the array and propagation to the field point, $\mathbf{r}_f$.

For any actual TRA, there will be a limit on how much power each transducer element can deliver. In addition, the relationship that determines the broadcast amplitude of the array must only contain quantities that are measured by the TRA or can be easily estimated. These restrictions are built into the amplification factor A, used in this study

$$A^2=\frac{\rho_o c\Pi_o/2\pi}{(1/N)\,\Sigma_{i=1}^{N}\{|M_s G(\mathbf{r}_i,\mathbf{r}_s,\omega)|^2+\langle|n(\mathbf{r}_i)|^2\rangle\}}, \qquad (2)$$

which limits the average power output per monopole array element to Π_o. In (2), the $\langle\ \rangle$ brackets denote an ensemble average, ρ_o is the local fluid density, and c is the local speed of sound. In practice, A would probably be set so that Π_o is comfortably below the maximum power output per element to prevent clipping during the TRA broadcast, with ρ_o and c set to appropriate reference (or average) values. The denominator of (2) is the average mean-square pressure received by an array element assuming that the signal, $M_s G(\mathbf{r}_i,\mathbf{r}_s,\omega)$, and the noise, $n(\mathbf{r}_i)$, are uncorrelated. For a water-column-spanning TRA, averaging over the array makes the value of A much less sensitive to the spatial distribution of the signal and noise fields.

Because the TRA operation cycle requires two steps, there are two signal-to-noise ratios that govern TRA performance. For ease of manipulation, both are defined without decibels. The first is the average received signal-to-noise ratio for an individual element at the array, SNR_r. It is one of the main independent parameters of this investigation and is defined here as

$$\mathrm{SNR}_r=\frac{\Sigma_{i=1}^{N}|M_s G(\mathbf{r}_i,\mathbf{r}_s,\omega)|^2}{\Sigma_{i=1}^{N}\langle|n(\mathbf{r}_i)|^2\rangle}. \qquad (3)$$

The second is the signal-to-noise ratio at the retrofocus, SNR_f. This is the main performance parameter for TRA and is defined by

$$\mathrm{SNR}_f(\mathbf{r}_f)=\frac{|\hat{s}|^2}{\langle|\hat{n}|^2\rangle}=\frac{A^2|\Sigma_{i=1}^{N}M_s G^*(\mathbf{r}_i,\mathbf{r}_s,\omega)G(\mathbf{r}_f,\mathbf{r}_i,\omega)|^2}{\langle|n(\mathbf{r}_f)|^2\rangle+A^2\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}\langle n(\mathbf{r}_i)n^*(\mathbf{r}_j)\rangle G(\mathbf{r}_f,\mathbf{r}_i,\omega)G^*(\mathbf{r}_f,\mathbf{r}_j,\omega)}, \qquad (4)$$

because it must include the two noise components. The numerator of (4) is the mean-square signal pressure $|\hat{s}|^2$ produced by the TRA at $\mathbf{r}_f$. The denominator of (4) is the expected mean-square noise pressure $\langle|\hat{n}|^2\rangle$ at $\mathbf{r}_f$; the first term represents the ambient noise at the field point, and the second terms represents the noise unintentionally broadcast by the TRA. Here, it has been assumed that $\mathbf{r}_f$ is sufficiently distant from the array (located at $\mathbf{r}_i$) so that these two noise components are uncorrelated. The main conclusion to be drawn from (4) is that the desired realm of high SNR_f will only occur when the contributions of ambient noise and TRA-broadcast noise are both small.

Up to this point no restrictions have been placed on the noise field, and (1)–(4) are fully general for narrow-band signals in time-invariant environments. However, (3) and (4) include the noise field's statistics so further simplification of these formulas necessitates the selection of a noise model. In this paper, a uniform omnidirectional noise field is used

$$\langle n(\mathbf{r}_i)n^*(\mathbf{r}_j)\rangle=\sigma_n^2\frac{\sin(k|\mathbf{r}_i-\mathbf{r}_j|)}{k|\mathbf{r}_i-\mathbf{r}_j|}, \qquad (5)$$

where $k=\omega/c$ is the acoustic wave number, and σ_n^2 is the variance of the noise field. For this simple noise field, if the array elements are assumed to be spaced at integer multiples of half an acoustic wavelength, or in a more general noise field when the element spacing is sufficiently large so that $\langle n(\mathbf{r}_i)n^*(\mathbf{r}_j)\rangle\approx 0\ \ (i\neq j)$, then at $\mathbf{r}_f=\mathbf{r}_s$ (4) reduces to

$$\mathrm{SNR}_f(\mathbf{r}_s)=\frac{M_s^2A^2\{\Sigma_{i=1}^{N}|G(\mathbf{r}_i,\mathbf{r}_s,\omega)|^2\}^2}{\sigma_n^2[1+A^2\Sigma_{i=1}^{N}|G(\mathbf{r}_i,\mathbf{r}_s,\omega)|^2]}. \qquad (6)$$

Furthermore, if the array elements are relatively loud or the source–array range is not too long [i.e., $AG(\mathbf{r}_i,\mathbf{r}_s,\omega)\gg 1$], then (3) and (4) in decibel form yield

$$\mathrm{SNR}_f(\mathbf{r}_s)=\mathrm{SNR}_r+10\log_{10}N. \qquad (7)$$

Thus, in this idealized case, the ideal array gain (see Jensen *et al.*, 1994) should be recovered at the original source location independently of the acoustic environment.

As will be seen (Fig. 2), a clear retrofocus forms reliably once SNR_f is somewhat above $+10$ dB while the retrofocus is consistently overwhelmed by noise when SNR_f is less than -10 dB or so. Hence, the last step in this development is to relate SNR_f to the probability φ that the mean-square signal amplitude $|\hat{s}|^2$ exceeds the mean-square noise amplitude $|\hat{n}|^2$ at $\mathbf{r}_f=\mathbf{r}_s$. This probability criterion can be stated as

$$\int_0^{|\hat{s}|^2}\mathrm{PDF}(|\hat{n}|^2)d(|\hat{n}|^2)=\varphi, \qquad (8)$$

where $\mathrm{PDF}(|\hat{n}|^2)$ is the probability density function for the total retrofocus noise field. If $|\hat{n}|$ is Rayleigh distributed (a typical assumption), then $\mathrm{PDF}(|\hat{n}|^2)$ is merely a decaying exponential so (8) produces

$$1-\exp\left\{-\frac{|\hat{s}|^2}{\langle|\hat{n}|^2\rangle}\right\}=1-\exp\{-\mathrm{SNR}_f(\mathbf{r}_s)\}=\varphi, \qquad (9)$$

a form of the desired relationship. It must be noted here that (9) is an approximation because $|\hat{n}|^2$ may not be exponentially distributed since it is a weighted sum of ambient and TRA-broadcast noise components. However, the exponential assumption for the distribution of $|\hat{n}|^2$ should be good at long ranges [i.e., when $AG(\mathbf{r}_i,\mathbf{r}_s,\omega)\ll 1$] where $|\hat{n}|^2$ is dominated by the lone noise term representing ambient noise at the retrofocus location.

These formal results can be cast into a convenient analytical form that includes all the various parameters and facilitates comparisons with the Monte Carlo simulations described in the next two sections. Combining (2), (3), (6), and (9) produces

$$\mathrm{SNR}_f(\mathbf{r}_s)=\ln\left[\frac{1}{1-\varphi}\right]=\frac{N^2XY}{1+NX+Y}, \tag{10a}$$

or

$$Y=\frac{\mathrm{SNR}_f(\mathbf{r}_s)\left(N+\dfrac{1}{X}\right)}{N^2-\dfrac{\mathrm{SNR}_f(\mathbf{r}_s)}{X}}, \tag{10b}$$

where X and Y are dimensionless ratios that represent the loudness of the TRA's elements and the original source with respect to the noise. These ratios are defined by

$$Y=\frac{M_s^2}{N\sigma_n^2}\sum_{i=1}^{N}|G(\mathbf{r}_i,\mathbf{r}_s,\omega)|^2=\mathrm{SNR}_r, \tag{11}$$

and

$$X=\frac{\rho_o c\Pi_o}{2\pi N\sigma_n^2}\sum_{i=1}^{N}|G(\mathbf{r}_i,\mathbf{r}_s,\omega)|^2. \tag{12}$$

The dimensionless ratio Y is proportional to M_s^2/σ_n^2 and is merely the received signal-to-noise ratio of (3) simplified for uncorrelated noise between array elements. The dimensionless ratio X is proportional to Π_o/σ_n^2 and is the average signal-to-ratio at $\mathbf{r}_s$ for the sounds transmitted by a single element of the TRA. When $X\gg 1$, the array elements are capable of producing loud sounds at the original source location compared to the ambient noise. For a fixed transducer power, this situation will occur at short source–array ranges or when the noise level is low. When $X\ll 1$, the TRA's transmission is quiet compared to the ambient noise at the original source location. For a fixed transducer power, this situation will occur at long source–array ranges or when the ambient noise level is high.

Equations (10)–(12) include all the parameters of interest to a system designer $(M_s,\Pi_o,\sigma_n^2,N,\varphi)$ and can be evaluated for particular source–array ranges (R) and acoustic frequencies if $|G(\mathbf{r}_i,\mathbf{r}_s,\omega)|$ can be calculated or estimated. Fortunately, only the magnitude of $G(\mathbf{r}_i,\mathbf{r}_s,\omega)$ is needed to evaluate (11) and (12). This facilitates the use of (10) for TRA system design because approximate scaling laws for $|G(\mathbf{r}_i,\mathbf{r}_s,\omega)|^2$ can be developed while there is little hope of correctly determining the phase of $G(\mathbf{r}_i,\mathbf{r}_s,\omega)$ in complex environments without measurements or full-wave computations.

The utility of (10) can be illustrated as follows. If it is necessary to achieve a certain value of φ, (10) specifies the combinations of N, X, and Y necessary to reach this objective. For example if a $\varphi=0.95$, or equivalently $\mathrm{SNR}_f=3.00$, is required from a TRA with a dozen array elements, then any combination of X and Y satisfying $Y(144X-3)=3(12X+1)$ should be effective. Thus, a system designer can determine how to partition resources between the source (represented by Y) and the array (represented by X) with only minimal information.

A TRA should work well when both Y and X are large, i.e., when both the original source and the TRA's broadcast are loud compared to the noise. However, TRA performance is limited when one or the other is relatively quiet. Given that X and Y are always positive, the form of (10b) leads to two limiting regimes: *strong source* and *strong array*. In both regimes, N plays a differing but important role in determining whether or not a TRA will retrofocus.

The strong source regime occurs when $Y\to\infty$. In this limit, (10a) requires $\mathrm{SNR}_f(\mathbf{r}_s)\to N^2X$ and (9) reduces to $\varphi\cong 1-\exp\{-N^2X\}$. Thus, when Y is large, the array produces a dB broadcast array gain of $+20\log_{10}N$ with respect to X, and the retrofocus probability is determined by the array's broadcast capabilities. The lower limit on X in this regime, $\mathrm{SNR}_f(\mathbf{r}_s)/N^2$, corresponds to the minimum possible TRA broadcast (i.e., no rebroadcast noise). At this limit, further improvement of the source broadcast (i.e., an increase in Y) does not produce any performance improvement because $\mathrm{SNR}_f(\mathbf{r}_s)$ is determined entirely by how loud the array's broadcast is compared to the focal ambient noise. Naturally, values of X below its lower limit do not produce the requisite $\mathrm{SNR}_f(\mathbf{r}_s)$.

The strong array regime occurs when $X\to\infty$. In this limit, (10b) requires $Y\to\mathrm{SNR}_f(\mathbf{r}_s)/N$, and (9) reduces to $\varphi\cong 1-\exp\{-NY\}$. Thus, when X is large, the array produces a dB broadcast array gain $+10\log_{10}N$ with respect to Y [see (7)], and the retrofocus probability is determined by the quality of the received source broadcast. Here, the broadcast array gain is smaller than that for the strong source regime because the array is rebroadcasting noise. In fact, the lower limit on Y in this regime, $\mathrm{SNR}_f(\mathbf{r}_s)/N$, corresponds to the maximum possible rebroadcast noise that the TRA can manage while still achieving the requisite $\mathrm{SNR}_f(\mathbf{r}_s)$.

The final results embodied in (10)–(12) capture all the essentials. This formulation is useful because contours of constant φ in the $X-Y$ plane can be used to predict the parametric regions where a TRA will or will not work.

III. NOISY FREE-SPACE ENVIRONMENT

As can be clearly seen by the presence of $G(\mathbf{r}_2,\mathbf{r}_1,\omega)$ in (1)–(4), (6), (11), and (12), the acoustic environment influences how well a TRA works. This section presents both theoretical and simulation results for a free-space environment, i.e., $G(\mathbf{r}_2,\mathbf{r}_1,\omega)=(1/|\mathbf{r}_2-\mathbf{r}_1|)\exp(ik|\mathbf{r}_2-\mathbf{r}_1|)$ where $k=\omega/c=2\pi/\lambda$, and λ is the acoustic wavelength. The simulation results are based on a numerical implementation of (1) where the real and imaginary parts of $n(\mathbf{r})$ were obtained from a Gaussian random number generator. For each trial, a successful retrofocus was declared if the TRA-broadcast signal amplitude exceeded the total noise amplitude at the intended retrofocus location, $\mathbf{r}_s$. The retrofocus probability for each set of parameters was determined by tabulating the results from 1000 trials. For this effort, the TRA's retrofocus location is defined as the original source location whether or not an absolute local maximum in the TRA-produced acoustic field exists there. Thus, this definition includes the array's far field where an acoustic beam is formed and there is no field maximum in the propagation direction, and the array's extreme near field where a complicated interference pattern may form. Considerations for near-field and far-field focusing in free-space are provided by Ziomek (1993).

The basic broadside performance of a linear TRA in a noisy free-space environment is shown in Figs. 1 and 2 for

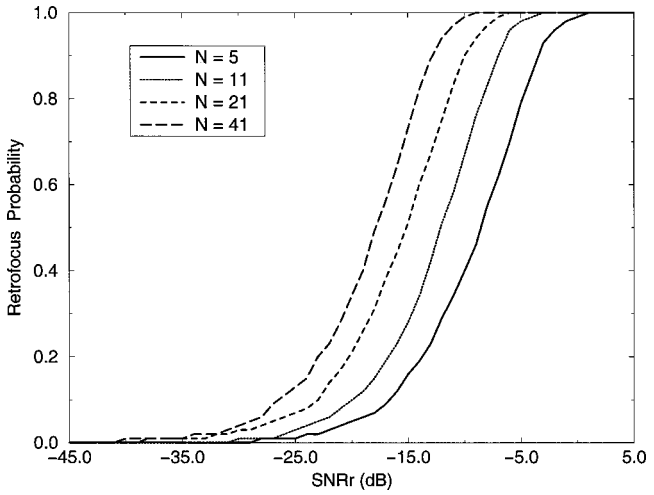


FIG. 1. Monte Carlo retrofocus probability versus received signal-to-noise ratio, SNR_r, for a variable number of array elements, N, when the ambient noise at the retrofocus location is small in a free-space environment. The array length is $L=20\lambda$, and the source–array range is $R=2000\lambda$. The noise field is omnidirectional and the array elements are integer half-wavelength spaced. As expected the retrofocus probability increases monotonically with increasing SNR_r, and larger N allows a TRA to operate at lower SNR_r.

the strong array limit. Figure 1 presents results for the far field of the array where $R>L^2/4\lambda$ with R=source–array range and L=length of the array (Kinsler *et al.*, 1982). The retrofocus probability is plotted versus SNR_r in decibels, for array-element numbers $N=5$, 11, 21, and 41 for $L=20\lambda$ and $R=2000\lambda$. As expected, for any fixed N, the retrofocus probability increases monotonically with increasing SNR_r. In addition, at any SNR_r, the curves for larger N have higher focal probabilities. Figure 2 contains both near-field and far-field results, but employs SNR_f in decibels on the horizontal axis. Here, the retrofocus probability curves collapse as suggested by (7) and (9) even though the two shortest ranges, $R=2\lambda$ and $R=20\lambda$, lie in the array's near field ($R<L^2/4\lambda$) and N varies by approximately a factor of 8. Although it is not plotted, the analytical curve (9) would lie amongst the tight bundle of curves in Fig. 2.

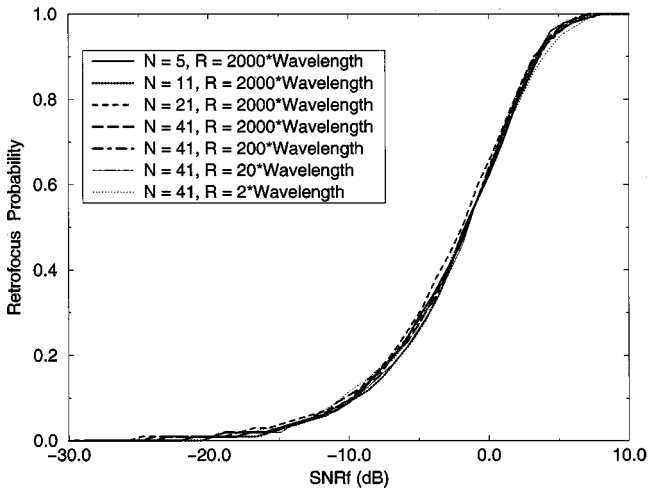


FIG. 2. Monte Carlo retrofocus probability versus the signal-to-noise ratio at the retrofocus, SNR_f, for a variable N and R. All conditions are the same as in Fig. 1 except where noted. The probability curves collapse independent of N, R, and whether or not the retrofocus lies in the near field of the array.

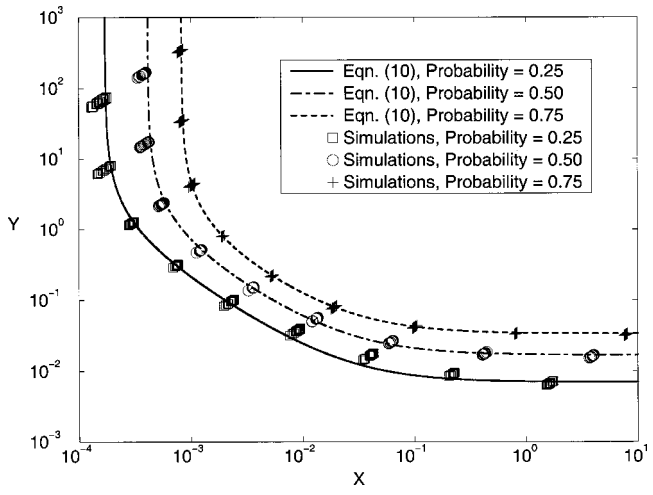


FIG. 3. Performance diagram for a TRA retrofocusing in a noisy free-space environment for $N=41$ and $L=20\lambda$. The horizontal axis is $X=\rho_0 c\Pi_0/2\pi R^2\sigma_n^2$, a signal-to-noise ratio for the TRA broadcast. The vertical axis is the received signal-to-noise ratio $Y=M_s^2/R^2\sigma_n^2=\mathrm{SNR}_r$. The smooth curves are the locus of retrofocus probabilities of 1/4, 1/2, and 3/4 calculated from Eq. (10). The plotted symbols are the results of the Monte Carlo simulations for $166.7\lambda\leqslant R\leqslant 4000\lambda$. A TRA will reliably focus in a noisy environment when both X and Y are large. Agreement between the theory and the simulations is good.

The elementary scaling shown in Fig. 2 is successful but it is incomplete because it does not include the strong source regime. A full comparison between (10) and simulations that cover both limiting regimes is given in Fig. 3 for $166.7\lambda \leqslant R\leqslant 4000\lambda$, $N=41$, $L=20\lambda$, and $\rho_o c=1.5\times 10^6$ Pa s/m. The horizontal and vertical axes are X and Y from (12) and (11), respectively. For this figure and the next, the sums of Green's functions in (11) and (12) are approximated by: $\Sigma|G(\mathbf{r}_i,\mathbf{r}_s,\omega)|^2\approx N/R^2$. The three curves at increasing distance from the origin of Fig. 3 are the locus of retrofocus probabilities of 1/4, 1/2, and 3/4, respectively. The discrete symbols corresponding to each probability are plotted at the appropriate parametric locations of the simulations. The theoretical curves match the simulations well over five orders of magnitude in both X and Y. The observed small deviations are probably caused by the use of an assumed probability distribution in (9) and (10).

Figure 3 diagrams the parametric regions where TRAs are likely and unlikely to retrofocus. Large values of X and Y lying to the right or above the probability curves will lead to reliable TRA retrofocusing. Combinations of X and Y lying to the left or below of the curves are regions where TRAs will be inoperable because of noise. This figure also displays three parametric ranges: the strong source regime where the probability curves turn vertical ($X<10^{-3}$ in Fig. 3), a transitional regime where the retrofocus probability depends on both X and Y ($10^{-3}<X<10^{-1}$ in Fig. 3), and the strong array regime where the probability curves are horizontal ($X>10^{-1}$ in Fig. 3). In the strong source regime, the performance of the TRA is limited by the array's inability to broadcast loudly enough to overcome the ambient noise at the retrofocus so the retrofocus probability becomes independent of Y and depends only on X and N. This regime sets the TRA operating range for loud sources. In the strong array regime, the retrofocus noise is dominated by noise broadcast

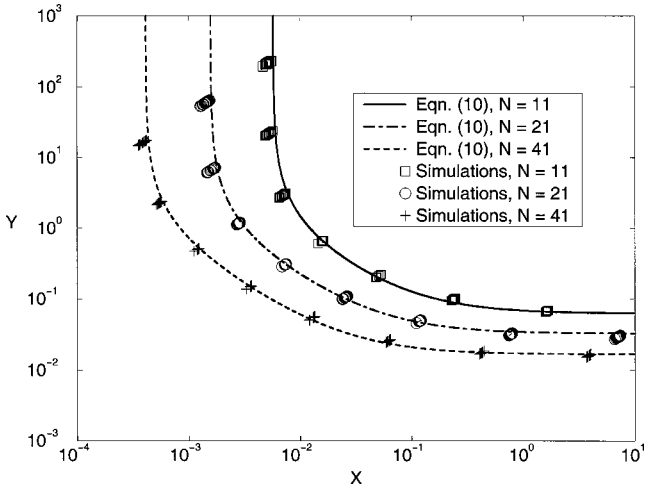


FIG. 4. Same as Fig. 3 except here N is varied and only the 50% retrofocus probability curves from Eq. (10) are plotted. As expected, TRAs with more elements can operate at smaller values of X and Y, i.e., with quieter sources and less powerful elements. As with Fig. 3, agreement between the theory and the simulations is good.

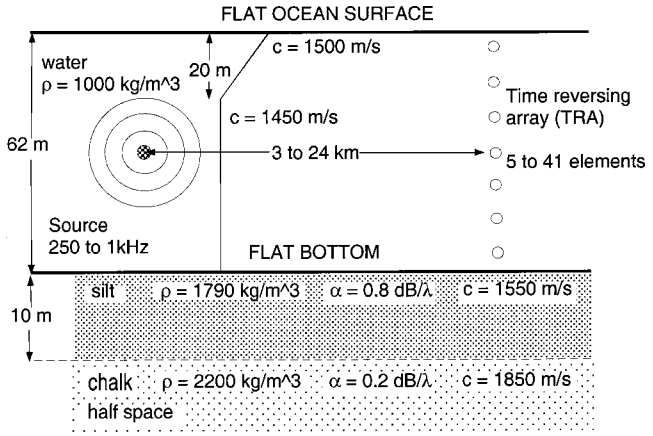


FIG. 5. Shallow-water sound-channel acoustic environment. A discrete, linear, vertical, integer multiple half-wavelength spaced array spans a simple, time-invariant sound channel with an ambient noise field. A harmonic point source located at ranges of 3 to 24 km from the TRA launches the initial acoustic wave. The TRA records both the signal from the source and the ambient noise field. The bottom has two layers with properties chosen from Jensen *et al.* (1994).

by that array so Y and N determine the retrofocus probability, independent of the broadcast strength of the array. This regime sets the TRA operating range for quiet sources and produces the elementary collapse of retrofocus probability curves seen in Fig. 2.

The effect of different numbers of array elements is shown in Fig. 4 which displays the locus of 50% retrofocus probability in $X-Y$ coordinates for $N=11$, 21, and 41. Note the greater sensitivity to changes in N in the strong source regime. As in Fig. 3, the discrete symbols mark the parametric locations of the simulations. Again, the theory–simulation agreement is good, which suggests that (10) correctly predicts the necessary parametric dependence on N.

The comparisons shown in this section make a compelling case for the accuracy of the noise formulation, but were all constructed using the simple, free-space Green's function. While such results may be directly applicable to ultrasonic TRAs, the effects of a more realistic underwater environment are considered in the next section. Of course, the *a priori* expectation is that the essential features seen in Figs. 3 and 4 will remain unchanged since the Green's function of the environment exerts equal influence on both the X and Y axes.

IV. NOISY SHALLOW-WATER SOUND CHANNEL

In this section, the theory described in Sec. II is compared to Monte Carlo simulations of TRAs in a noisy shallow-water sound channel. The effort is conceptually identical to that in Sec. III except for the replacement of the free-space Green's function with one computed using the wide-angle parabolic-equation code RAM (Collins, 1993, 1994, 1998). The main differences between the sound-channel and free-space environments are: the sound channel supports multipath (multimode) propagation, and it incorporates acoustic absorption which strongly attenuates high-order modes and hinders TRA retrofocusing. Modal propagation explicitly brings the acoustic wave number (or frequency) into the formulation, and the absorption losses make simple range scaling of the Green's function (like $|G|^2 \approx 1/R^2$ for free space) less accurate because each mode is attenuated at a different rate with increasing source–array range. The main operational difference between the two environments is that both the signal and the rebroadcast noise undergo modal propagation in the sound channel, while in free-space the noise unintentionally broadcast by the TRA effectively radiates away omnidirectionally.

The computational shallow-water sound channel chosen for these simulations was range- and time independent. Its geometry and properties are shown in Fig. 5. The linear TRA spanned the water column and the source was placed at mid-depth in the sound channel (i.e., broadside to the array). This source–array configuration was chosen to facilitate comparisons with the free-space results. Investigations at several acoustic frequencies were conducted for $R=3$, 6, 12, and 24 km with $N=5$, 11, 21, and 41. The source–array geometry at 500 Hz and $R=6$ km is precisely the same as the free-space investigations at $R=2000\lambda$, and $L=20\lambda$. Based on convergence studies, the depth grid spacing was chosen to be $\lambda/30$ or smaller, the range grid spacing was $4\lambda/3$ or smaller, eight Padé terms were used, and the depth of the computational half-space was 500 m. Within the last 4λ of the computational bottom the attenuation was ramped up to 10 dB/λ to prevent artificial reflections. One hundred Monte Carlo simulations were performed for each set of parameters.

Figure 6 shows simulation results for the retrofocus probability versus SNR_f in the strong array regime at 500 Hz for $N=5$, 11, 21, and 41 at $R=6$ km; and for $N=41$ with $R=3$, 6, 12, and 24 km. As expected, the retrofocus probability increases with increasing SNR_f. Embedded in this figure is the improvement in the retrofocus probability with increasing N suggested by (7). These results are essentially identical to those shown in Fig. 2. The main difference between the two figures is the greater statistical convergence seen in Fig. 2 resulting from the larger number of free-space Monte Carlo trials. Additionally, the theoretical result (10) falls nicely among the simulation curves in Fig. 6, thereby

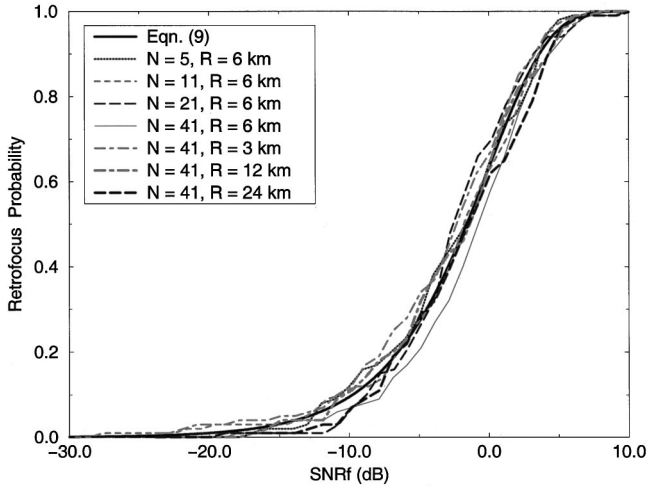


FIG. 6. Retrofocus probability versus the signal-to-noise ratio at the retrofocus, SNR_f, for a variable N and R, when the ambient noise at the retrofocus location is small in the shallow-ocean sound channel of Fig. 5. The jagged curves are from the Monte Carlo simulations at 500 Hz with $N=41$ and $L=60$ m. The smooth curve is given by (9). The simulation results collapse well around the theoretical curve. The remaining scatter is likely caused by incomplete statistical convergence. The equivalent of Fig. 1 for these results can be recovered through (7).

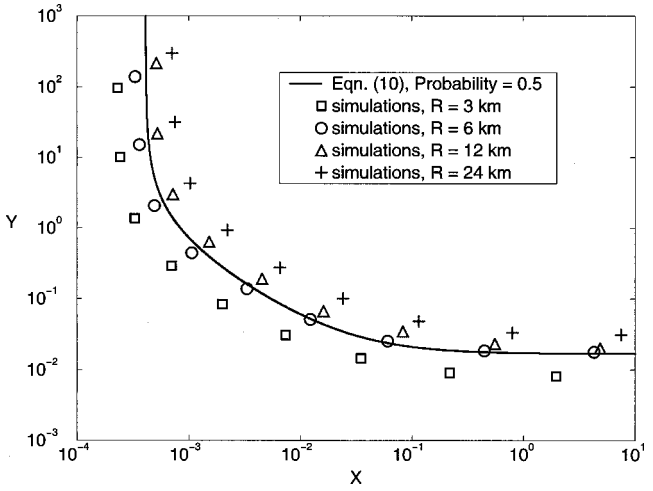


FIG. 7. Performance diagram for a TRA retrofocusing in a noisy shallow-water sound channel for $N=41$ and $L=60$ m for several source–array ranges, R, at a frequency of 500 Hz ($M=23$). Here, X and Y are computed from (11) and (12) using the approximation (11). The individual symbols represent the results of Monte Carlo simulations. The solid line is the locus of 50% retrofocus probability. Theory–simulation agreement is acceptable, although acoustic absorption, which is not included in (13), leads to better TRA performance at smaller R.

suggesting that the chosen probability density function for the noise is acceptable for parameters leading to the results in this figure.

A performance diagram similar to the free-space results in Fig. 3 can be constructed for the sound-channel environment once a scaling law for $|G(\mathbf{r}_i,\mathbf{r}_s,\omega)|^2$ is developed. An approximate form based on the classical Pekeris-waveguide modal sum was used in these investigations

$$|G(\mathbf{r}_i,\mathbf{r}_s,\omega)|\approx\beta\frac{M}{D\sqrt{kR}}, \tag{13}$$

where β is a dimensionless empirical constant, D is the depth of the sound channel, and M is the number of modes with propagation angles less than the critical angle for penetration into the bottom (Frisk, 1994)

$$M=\frac{2D}{\lambda}\sqrt{1-\frac{c^2}{c_1^2}}+\frac{1}{2}, \tag{14}$$

where c is the average or reference speed of sound in the water column, c_1 is a representative sound speed in the bottom. Here, M was calculated from (14) with $D=62$ m, $c=1500$ m/s, and $c_1=1800$ m/s, and then rounded down to the nearest integer before use in (13) with $\beta=0.2$. For the present purposes, the main liability of (13) is that is does not account for absorption.

The sound-channel results for the locus of the 50% retrofocus probability curve are given in Figs. 7–9 in $X-Y$ coordinates. These figures show the parametric effects of changing source–array range, acoustic frequency, and the amount of bottom absorption. In all these figures, (10)–(14) were used to produce the theoretical 50% retrofocus probability curves (solid lines) while the statistical simulation results appear as discrete points. Figure 7 shows 50% probability contour results at 500 Hz ($M=23$) for $N=41$ and $R=3$, 6, 12, and 24 km. In this figure, the simulation results move further away from the origin with increasing R. Thus, TRA operations require a louder sources and more powerful elements at longer ranges. Figure 8 shows the combined influence of changes in source-array range and frequency for $N=21$. It portrays the same four source–array ranges as Fig. 7, but includes the results from three acoustic frequencies: 250 Hz ($M=12$), 500 Hz ($M=23$), and 1 kHz ($M=46$). For clarity each frequency is plotted as the same symbol with the longer range simulations producing 50% retrofocus probability points further from the origin. Comparing different

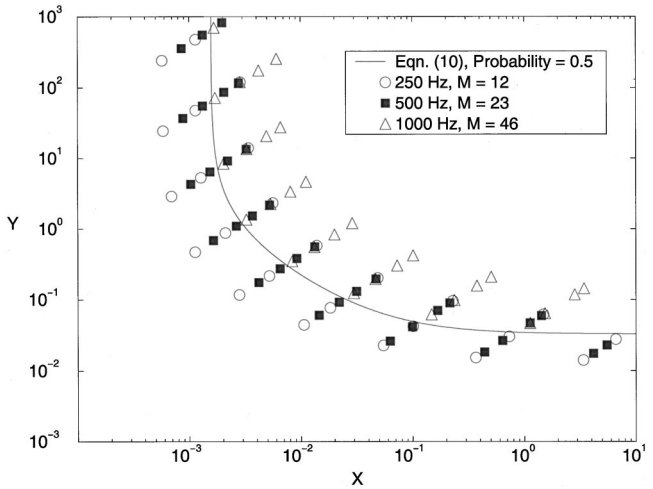


FIG. 8. Performance summary for TRA retrofocusing in a noisy shallow-water sound channel for several source–array ranges, R, at three frequencies with $N=21$. The axes are the same as in Fig. 7. The smooth curve is Eq. (10). The individual symbols represent the results of Monte Carlo simulations at each frequency for 50% retrofocus probability. In any string of points, longer range results lie further from the origin. The spread within each string of points and between the various frequencies is caused by the lack of absorption in the approximate Green's function magnitude Eq. (13) used to generate X and Y.

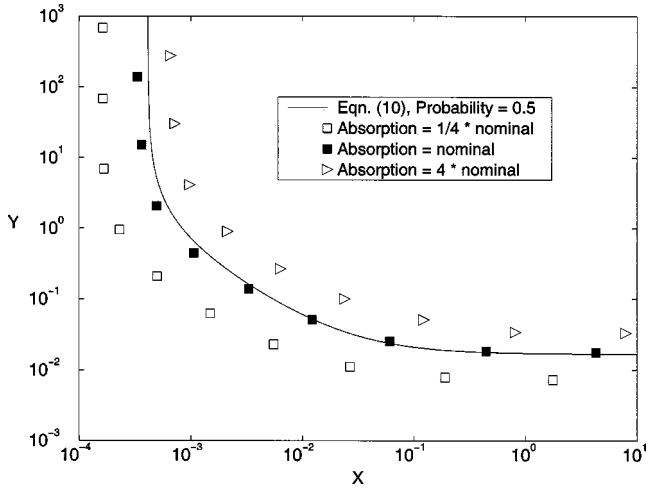


FIG. 9. Performance summary for TRA retrofocusing in a noisy shallow-water sound channel at 500 Hz for $N=41$, $L=60$ m, and $R=6$ km for variable bottom attenuation levels. The axes and the theoretical curve are the same as in Fig. 7. The individual symbols represent the results of Monte Carlo simulations for 50% retrofocus probability. Higher absorption pushes the simulation points further from the origin. Clearly, greater source strength and/or array element power is needed to maintain TRA performance in noisy environments with higher absorption.

frequencies in Fig. 8 shows that increasing frequency causes the 50% retrofocus probability points to move farther from the origin, essentially the same effect as increasing range. The results for 24 km at 250 Hz were excluded from the plot because a proper retrofocus does not form in this case.

The spread in the simulation results shown in Figs. 7 and 8 is primarily caused by the effects of acoustic absorption at the different ranges and frequencies. In general, acoustic absorption makes it increasingly difficult for the TRA to retrofocus at longer ranges or higher frequencies. Figure 9 illustrates this point by showing the 50% retrofocus probability results for 500 Hz at $R=6$ km with differing absorption levels in the first layer of the bottom. Clearly, higher absorption requires larger X and Y values, i.e., louder sources or more powerful array elements, to reach the region of reliable retrofocusing.

V. SUMMARY AND CONCLUSIONS

The impact of an omnidirectional noise field on the retrofocusing of a time-reversing array (TRA) operating in free-space and in a shallow-water sound channel has been presented in this paper. A general theory for predicting the probability that the intended-signal amplitude will exceed the noise at the retrofocus has been constructed. This theory includes the noise field variance, the original source strength, the broadcast power of the TRA's elements, the number of TRA elements, the source–array range, and other propagation characteristics of the environment. It compares well to Monte Carlo simulations of TRA retrofocusing in noisy environments even when approximate forms for the Green's function of environment are used.

There are three main conclusions that can be drawn from this study. First of all, TRAs can operate in noisy environments. For example, the results in Fig. 6 suggest that a TRA with 41 array elements operating at 500 Hz may produce a retrofocus probability of 75% at a range of 6 km in a shallow-ocean sound channel even when the received signal-to-noise ratio is as low as -15 dB. Moreover, a TRA with relatively loud elements should benefit from significant array gain, $+10\log_{10}N$ to as much as $+20\log_{10}N$, *on broadcast* in an omnidirectional noise field as the number of array elements (N) increases. Second, TRA retrofocusing performance is limited in noisy environments by low received signal-to-noise ratios, or by insufficient element power. When the received signal-to-noise ratio is low, a TRA wastes broadcast power retransmitting received noise, and this broadcast noise interferes with the intended signal at the retrofocus, even when the array elements are arbitrarily powerful. Likewise, when the TRA element power is insufficient to pull the intended signal above the ambient noise field at the retrofocus, even a perfect signal reception from the original source is no guarantee that a proper retrofocus will be produced. However, the additional array elements help in both cases.

And finally, the theoretical scaling of the omnidirectional noise field is acceptable in both free-space and a shallow-water sound channel even when approximate forms for the Green's function are used. Thus, it can safely be concluded that the theory applies to omnidirectional noise fields in any environment. One clear advantage of the probability formulation developed in Sec. II lies in the fact that the performance of a TRA in a noisy environment may be predicted or estimated with a simple analytic rule or scaling law. Such laws can form the basis for TRA system design because they can be evaluated easily with scant information. For example, consider a half-wavelength-spaced TRA having 1-μW transducers, operating in an aqueous free-space environment with ambient noise variance of 1 Pa^2. A system designer might wish to know how many array elements are required to achieve a probability of retrofocus of 90% at a range of 1 m with a source that produces a signal amplitude of 0.3 Pa at 1 m. Under these circumstances, (9), (11), and (12) can be evaluated to find $X=0.12$, $Y=0.09$, and $\mathrm{SNR}_f(\mathbf{r}_s)=2.3$. Plugging these values into (10) and solving for the number of array elements produces $N\approx 30$. Design calculations for a sound channel proceed in similar manner with the Green's function magnitude estimated from (13). However, uncertainty in the extent of acoustic absorption in a shallow-water sound channel would always require a system designer to conservatively choose system parameters.

ACKNOWLEDGMENTS

The research was sponsored by the Ocean Acoustics Program of the Office of Naval Research under Grant No. N00014-96-1-0040.

Baggeroer, A. B., Kuperman, W. A., and Schmidt, H. (**1988**). ''Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem,'' J. Acoust. Soc. Am. **83**, 571–587.

Catipovic, J. A. (**1997**). ''Acoustic Telemetry,'' in *Encyclopedia of Acoustics* (Wiley, New York), Vol. 1, Pt. 4, Ch. 51, pp. 591–596.

Collins, M. D. (**1993**). ''A split-step Padé solution for parabolic equation method,'' J. Acoust. Soc. Am. **93**, 1736–1742.

Collins, M. D. (**1994**). ''Generalization of the split-step Padé solution,'' J. Acoust. Soc. Am. **96**, 382–385.

Collins, M. D. (**1998**). ''New and improved parabolic equation models,'' J. Acoust. Soc. Am. **104**, 1808(A).
Dowling, D. R. (**1993**). ''Phase-conjugate array focusing in a moving medium,'' J. Acoust. Soc. Am. **94**, 1716–1718.
Dowling, D. R. (**1994**). ''Acoustic pulse compression using passive phase-conjugate processing,'' J. Acoust. Soc. Am. **95**, 1450–1458.
Dowling, D. R., and Jackson, D. R. (**1992**). ''Narrow-band performance of acoustic phase-conjugate arrays in dynamic random media,'' J. Acoust. Soc. Am. **91**, 3257–3277.
Draeger, C., and Fink, M. (**1999a**). ''One channel time reversal in chaotic cavities; Theoretical limits,'' J. Acoust. Soc. Am. **105**, 611–617.
Draeger, C., and Fink, M. (**1999b**). ''One channel time reversal in chaotic cavities: Experimental results,'' J. Acoust. Soc. Am. **105**, 618–625.
Dungan, M. R., and Dowling, D. R. (**2000**). ''Computed time-reversing array retrofocusing in a dynamic shallow ocean,'' J. Acoust. Soc. Am. **107**, 3101–3112.
Fink, M. (**1997**). ''Time reversed acoustics,'' Phys. Today **50**, 34–40.
Frisk, G. V. (**1994**). *Ocean and Seabed Acoustics, A Theory of Wave Propagation* (Prentice-Hall, Englewood Cliffs, NJ), pp. 145–152.
Hodgkiss, W. S., Song, H. C., Kuperman, W. A., Akal, T., Ferla, C., and Jackson, D. R. (**1999**). ''A long-range and variable focus phase-conjugation experiment in shallow water,'' J. Acoust. Soc. Am. **105**, 1597–1604.
Jackson, D. R., and Dowling, D. R. (**1991**). ''Phase conjugation in underwater acoustics,'' J. Acoust. Soc. Am. **89**, 171–181.
Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (**1994**). *Computational Ocean Acoustics* (AIP, New York), pp. 41, 558–560.
Khosla, S. R., and Dowling, D. R. (**1998**). ''Time-reversing array retrofocusing in simple dynamic underwater environments,'' J. Acoust. Soc. Am. **104**, 3339–3350.
Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**1982**). *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York), p. 188.
Kuperman, W. A., Hodgkiss, W. S., Song, H. C., Akal, T., Ferla, C., and Jackson, D. R. (**1998**). ''Phase-conjugation in the ocean: Experimental demonstration of an acoustic time reversal mirror,'' J. Acoust. Soc. Am. **103**, 25–40.
Mordant, N., Prada, C., and Fink, M. (**1999**). ''Highly resolved detection and selective focusing in a waveguide using the DORT method,'' J. Acoust. Soc. Am. **105**, 2634–2642.
Prada, C., and Fink, M. (**1998**). ''Separation of the interfering acoustic scattered signals using invariants of the time-reversal operator. Application to Lamb waves characterization,'' J. Acoust. Soc. Am. **104**, 801–807.
Roux, P., and Fink, M. (**2000**). ''Time reversal in a waveguide: Study of the temporal and spatial focusing,'' J. Acoust. Soc. Am. **107**, 2418–2429.
Song, H. C., Kuperman, W. A., and Hodgkiss, W. S. (**1998**). ''A time-reversal mirror with variable range focusing,'' J. Acoust. Soc. Am. **103**, 3234–3240.
Song, H. C., Kuperman, W. A., Hodgkiss, W. S., Akal, T., and Ferla, C. (**1999**). ''Iterative time reversal in the ocean,'' J. Acoust. Soc. Am. **105**, 3176–3184.
Steinberg, B. D. (**1976**). *Principles of Aperture and Array System Design* (Wiley, New York), Chap. 12.
Tanter, M., Thomas, J.-L., and Fink, M. (**1998**). ''Focusing and steering through absorbing and aberrating layers: Application to ultrasonic propagation through the skull,'' J. Acoust. Soc. Am. **103**, 2403–2410.
Yoo, K., and Yang, T. C. (**1998**). ''Improved vertical array performance in shallow water with a directional noise field,'' J. Acoust. Soc. Am. **104**, 3326–3338.
Ziomek, L. J. (**1993**). ''Three necessary conditions for the validity of the Fresnel phase approximation for the near-field beam pattern of an aperture,'' IEEE J. Ocean Eng. **18**, 73–75.

Noncontact quantitative spatial mapping of stress and flexural rigidity in thin membranes using a picosecond transient grating photoacoustic technique

John A. Rogers,[a] Gregory R. Bogart, and Ron E. Miller
Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974

(Received 23 August 2000; accepted for publication 28 November 2000)

This paper describes a purely optical technique for measuring and spatially mapping out stress and rigidity in thin membranes. Its application to a membrane of aluminum nitride that has significant spatial nonuniformities in its elastic properties demonstrates the method. The attractive features of this technique—fast, noncontacting measurement, good spatial resolution, ability to quantify in-plane anisotropy—make it potentially useful for characterizing elements of microelectromechanical structures, masks for advanced lithography systems, acoustic filters, and other devices in which the mechanical properties of membranes are important. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1342005]

PACS numbers: 43.35.Sx, 43.38.Zp, 43.20.Ye [SLE]

I. INTRODUCTION

Thin membranes are important components of microelectromechanical systems[1] and microfluidic devices. They are also used in masks for advanced lithography systems that use x rays[2] and projected beams of electrons.[3,4] We recently described a purely optical, noncontacting method for evaluating membrane stress and flexural rigidity, two quantities that often determine the utility of these structures in devices.[5] The measurement technique, known as transient grating (TG) photoacoustics or impulsive stimulated thermal scattering, uses crossed laser pulses to excite acoustic modes in the membrane. A continuous wave laser probes the time dependence of these motions. Fitting the TG-measured dispersion of the lowest order acoustic mode (i.e., the variation of its phase velocity with wavelength) to computations for thin plates determines the rigidity and stress. The approach has several attractive features: it is fast (~1 s) and noncontacting, the analysis of data is straightforward, in-plane anisotropies can be characterized easily, and specialized test structures are not required. In this paper we illustrate how the TG method can be used to examine membranes with spatially nonuniform properties. Successful measurement of these samples illustrates the ability (i) to quantify and map out spatial distributions of stress and rigidity and (ii) to evaluate structures that would be difficult or impossible to analyze accurately with conventional mechanical tests that use bulk drumhead motions of membranes (e.g., the resonant frequency method[4,6] or the bulge test[7]). In the following, we describe the TG measurement approach and fabrication procedures for membranes of aluminum nitride. A theory section provides a framework for interpreting data collected from these samples. We summarize the results and conclude by describing factors that affect the accuracy and spatial resolution.

[a]Electronic mail: jarogers@physics.lucent.com

II. EXPERIMENT

A. Transient grating measurements

The TG measurement technique is described in detail elsewhere.[8–11] Briefly, a pair of crossed laser pulses produces an optical interference pattern with a period (Λ) determined by the crossing angle (θ) and the wavelength of the pulses (λ_e) according to

$$\Lambda = \frac{\lambda_e}{2\sin(\theta/2)}. \tag{1}$$

Absorption of excitation light by the sample launches a thermal response and coherent, counterpropagating acoustic waves with wavelengths equal to Λ. (The magnitude of the acoustic wave vector, k, is simply $2\pi/\Lambda$.) Measuring the intensity of light diffracted from a continuous wave laser beam that overlaps the excited region of the sample reveals the time dependence of these motions. See Fig. 1. A fast detector and transient digitizing oscilloscope record the entire material response with each shot of the excitation laser. High quality signal (signal to noise greater than ~100) can be obtained in a few seconds by averaging data collected from several hundred shots. For the experiments described here, a passively Q-switched microchip Nd:YAG laser ($\lambda_e = 1064$ nm, pulse width = 200 ps) was used for excitation and a diode laser (850 nm, 0.2 W) was used for probing. [We note that the heating induced by the excitation pulses can have a small effect on the acoustic response (<~1% shift in frequency). In previous work we demonstrated how to quantify experimentally this shift and to account for it in the analysis.[5] For the purposes of this paper, we ignore this small effect.]

Figure 2 shows typical data from a sample schematically illustrated in Fig. 1. The oscillations in the signal are due to motions associated with the stimulated acoustic mode. Fourier transformation of the time domain data reveals the frequency (ω) of this mode. The power spectrum and a Lorentzian fit are shown in the inset to Fig. 2. The acoustic wavelength, Λ, is $28.56 \pm 0.05\,\mu$m in this case. Measure-

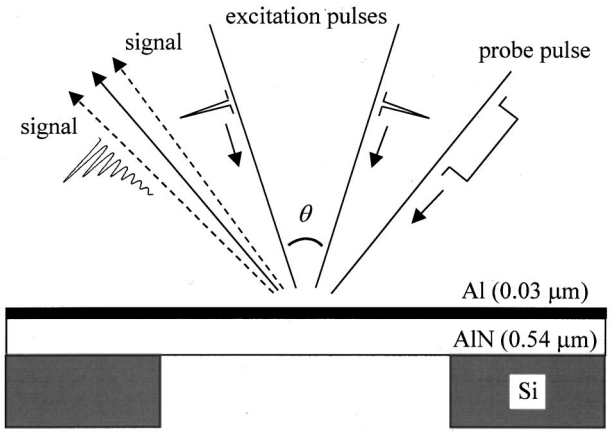


FIG. 1. Schematic illustration of the transient grating measurement and the membrane samples. Crossed picosecond excitation pulses launch coherent counterpropagating acoustic plate modes with wavelengths defined by the optical interference pattern. Measuring the time-dependent diffraction of a probe laser beam from these acoustic motions determines their frequency. The sample consists of a bilayer of Al (0.03 μm)/AlN (0.54 μm) on a silicon wafer. Locally removing the silicon by back side etching produces membranes in selected regions of the wafer.

ments at several wavelengths determine the dependence of the acoustic phase velocity ($v_\varphi = \omega/k$) on the acoustic wavevector, k. We collected data at $\Lambda = 28.56$, 22.85, 18.09, and $14.28 \pm 0.05\ \mu$m.

The size of the excitation region principally determines the spatial resolution of the measurement (provided that this size exceeds the acoustic wavelength).[12] For the experiments described here, the excitation spot was elliptical, with a $\sim$300 μm long axis (perpendicular to the interference fringes) and a $\sim$100 μm short axis (parallel to the fringes). The probe laser was focused to a round spot (diameter of $\sim$100 μm) overlapped with the center of the excitation region. By keeping the positions of the laser beams fixed and translating the sample with $x-y$ stages, different spatial locations on the membrane (rectangular in shape with dimensions of 1.0 cm$\times$0.075 cm, see Sec. II B) were probed. Sequential measurements performed in this way yielded maps of the acoustic properties and, as described in Sec. II B, the stress and rigidity. Using this approach, we measured the acoustic response as a function of position along the long axis of the membrane at its center. In this case, we recorded data in 100 μm steps with the acoustic wavevector aligned along the translation direction (i.e., wavevector aligned parallel to the long axis of the membrane). We performed a similar series of measurements (100 μm steps) along the short axis of the membrane at two different locations along its long axis. Finally, we examined the degree of in-plane anisotropy by comparing the acoustic responses with wavevectors aligned parallel and perpendicular to the long axis of the membrane.

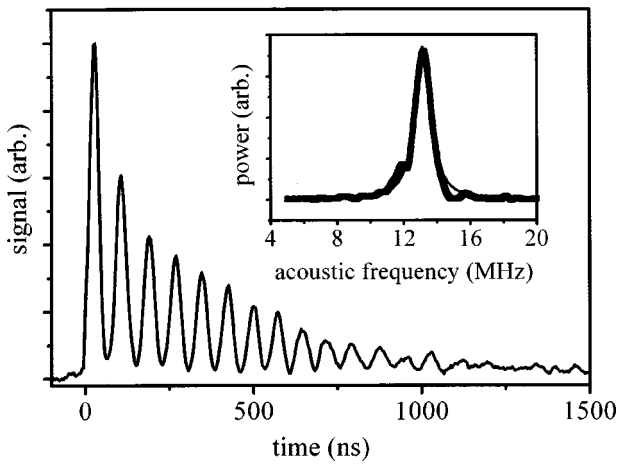


FIG. 2. Typical transient grating photoacoustic data collected from a bilayer membrane of Al (0.03 μm)/AlN (0.54 μm). Crossed excitation pulses arrive at the sample at $t \sim 0$ ns. Acoustic and thermal motions induced by slight absorption of these pulses ripple the surface of the membrane. This ripple causes time-dependent diffraction of a probe laser beam that is overlapped with the excited region of the sample. Oscillations of the signal in this case are produced by motions associated with the lowest order antisymmetric Lamb acoustic mode. The wavelength of this mode ($28.56 \pm 0.05\ \mu$m) is defined by the angle between the excitation beams. The inset shows the power spectrum and a fit to a Lorentzian line shape.

B. Sample fabrication

The fabrication begins with deposition of a film of aluminum nitride (AlN) onto a silicon wafer (200 mm diameter, $\langle 001 \rangle$ orientation) using a plasma vapor deposition tool (Model F$\times$P, Trikon Technologies). The chamber pressure was maintained at 1.6 mT with 20 sccm of both Ar and N_2. A pulsed (100 kHz, positive pulse duration 10 μs) dc potential (2000 W) applied to the Al target (heated to 150 °C) yields a consistent deposition rate of 0.6 nm/s. Films with nominal thickness of 0.5 μm on the front of the wafer (membrane) and 0.3 μm on the back of the wafer (etch mask) were deposited in this manner. A radio frequency substrate bias (65 W) provides control over the film stress. The conditions were optimized to deliver uniform films with low stress.

The membranes were created by removing the silicon in rectangular regions at various locations across the wafer. To generate these membranes, we first patterned a layer of photoresist (Microposit S1818, Shipley) on the back of the wafer. After exposure, the photoresist is developed using a potassium hydroxide based developer that also etches the AlN. The photoresist is then removed and the silicon is etched from the back side in the regions not protected by the aluminum nitride etch mask. The deep silicon etching was performed with a Surface Technology Systems, ASE-ICP tool (Newport, UK) and a variant of the time multiplexed deep etching (TMDE) method.[13,14] The process consists of alternating etching and polymer deposition steps that allow for etched features with high aspect ratios. The etch selectivity between the Si and AlN is greater than 10 000:1, which enables the membrane itself to act as its own etch stop.

The etching part of the procedure was performed with a pulsed bias-TMDE using a 2000 W source power and a 20 W platen bias pulse for 2.5 s followed by a 2 s delay. The bias pulse was followed by a constant 5 W platen bias for the remainder of the $\sim$13 s etch. The etchant gases and flows were 180 sccm SF_6 and 18 sccm O_2 at 42 mTorr. Helium back side cooling was performed at low pressures ($\sim$2 Torr) to avoid damaging the membranes. The polymer deposition part of the TDME process was carried out with 60 sccm of C_4F_8 at 15 mT and a power of 1200 W for 5 s. Membranes

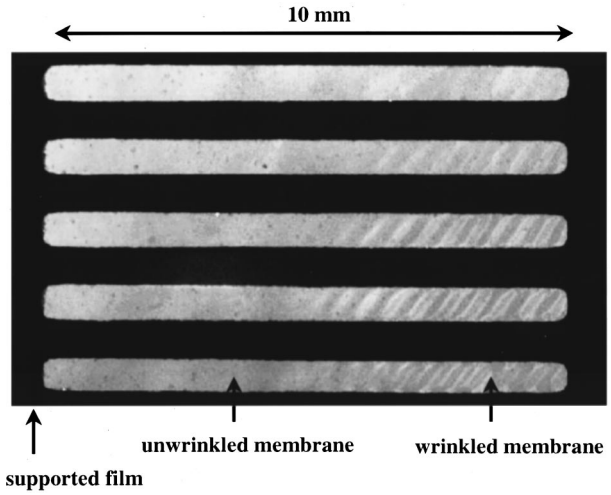


FIG. 3. Optical micrograph of five rectangular bilayer membranes of Al (0.03 μm)/AlN (0.54 μm). The images were recorded by illuminating the sample from the back side. The black areas are silicon (opaque) and the gray regions are the membranes (partially transparent). The right sides of the membranes are wrinkled. In these regions, it is likely that the silicon-supported bilayer is under compressive stress. Removing the silicon allows the membrane to relieve this stress (at least partially) by buckling. On the left sides, the stress is tensile.

formed in this manner show clear indications of stress nonuniformities. In particular, certain parts of these membranes are wrinkled, which suggests that they buckled when the silicon was removed in order to relax, at least partially, compressive stress. Other regions appear flat and taut, which is consistent with a state of tensile stress. The cause of these variations is not currently understood clearly.

In the final step of the fabrication, a thin (30 nm) coating of aluminum was thermally evaporated onto the AlN. This coating absorbs a small fraction of the excitation light; the mild heating associated with this absorption launches the acoustic response. Figure 3 shows an optical micrograph (back side illumination) of an array of five rectangular membranes of Al (0.03 μm)/AlN (0.54 μm) produced using the above-described procedures. The black and gray regions correspond to the silicon substrate (opaque) and the membranes (partially transparent), respectively. The dimensions of the membranes are approximately 1 cm$\times$0.075 cm. The thickness of the AlN was determined adjacent to the membranes by optical reflectometry. These measurements revealed no significant variation in thickness or index of refraction in the films along the lengths of the membranes. Note the wrinkled regions on the right sides of the membranes in Fig. 3. This appearance is consistent with membrane stress that varies from zero on the right to increasingly tensile on the left.

III. THEORY

The TG method has been used in the past to measure the dispersion of Lamb acoustic waveguide modes in thin membranes.[8,15,16] These data can be interpreted with rigorous acoustic waveguide calculations to extract the elastic moduli. In the limit that the wavelengths of the excited acoustic modes are much larger than the thickness of the membrane, it is possible to account for the dispersion of the lowest order Lamb mode (a mode that is typically easy to measure with TG methods)[15] with small-deflection plate theory.[5] The corresponding one-dimensional equation of motion is[17]

$$\frac{D}{h}\frac{\partial^4 u}{\partial x^4}+\rho\frac{\partial^2 u}{\partial t^2}=\sigma\frac{\partial^2 u}{\partial x^2}, \tag{2}$$

where D is the flexural rigidity, ρ is the density, σ is the stress, t is time, u is the out of plane displacement, and x is the position along the membrane. This theory (in one and two spatial dimensions) is commonly used to analyze data from bulge and resonance frequency tests. Its application to analysis of TG measurements is simple because it does not require assumptions about the mechanical nature of the interface between the membrane and its support. Also, the wavevectors of motions excited in a TG experiment are well defined: They are determined by the crossing angle between the excitation pulses according to Eq. (1). Furthermore, the lateral dimensions of the membrane are unimportant in the analysis, provided that they are sufficiently larger than the acoustic wavelength. These features allow a simple expression for interpreting measured variations in the acoustic phase velocity with acoustic wavevector:

$$v_\varphi=\frac{\omega}{k}=\sqrt{\frac{D}{\rho h}k^2+\frac{\sigma}{\rho}}. \tag{3}$$

The flexural rigidity, D, is related to Young's modulus E, Poisson's ratio v, and the film thickness h by

$$D=\frac{Eh^3}{12(1-v^2)}. \tag{4}$$

The validity of Eq. (3) depends on the mechanical properties of the membrane as well as the product kh. For large kh or for acoustic modes other than the lowest one, waveguide effects are important and Eq. (3) cannot be used. At large kh, the lowest order acoustic mode becomes confined to the surface of the membrane and takes on characteristics similar to those of the Rayleigh wave of a free surface. This type of waveguide motion deviates strongly from the simple drumhead vibrations described by the plate theory. Equation (3) generally provides a good approximation, however, for the lowest order mode in films of most materials at $kh<0.5$. In this regime, the square of the phase velocity is linearly related to the square of the wavevector. The intercept of this linear relation defines the ratio of the stress to the density; the slope defines the ratio of the flexural rigidity to the product of the film thickness and the density ($D'=D/\rho h$). Figure 4 compares the dispersion computed for a layer of AlN (0.54 μm) using acoustic waveguide theory (isotropic elastic properties: $E=283$ GPa, $v=0.25$, and $\rho=3250$ kg/m^3)[18] to that defined by Eq. (3). In both cases, $\sigma/\rho=400$ m^2/s^2. These results indicate the range of validity of plate theory for this sample.

The waveguide computations can, in a straightforward way, explicitly include multiple membrane layers. For example, Fig. 4 shows the computed dispersion for the bilayer Al (0.03 μm)/AlN (0.54 μm) (isotropic properties for Al: $E=70.0$ GPa, $v=0.35$, and $\rho=2700$ kg/m^3).[19] The results in this case differ from those for the single layer of AlN by only

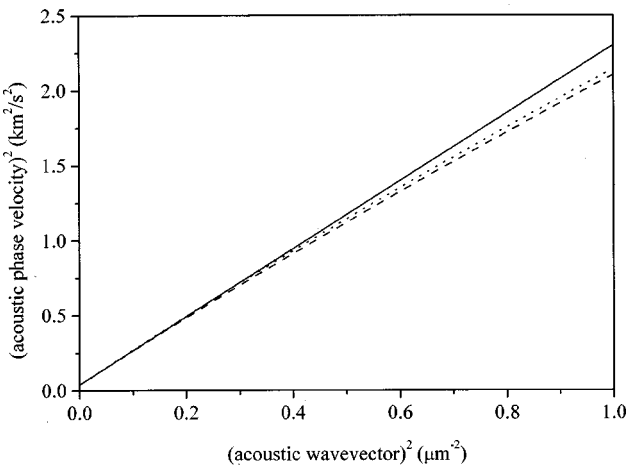


FIG. 4. Dispersion of the lowest order Lamb mode in a membrane of AlN (0.54 μm). The dashed and solid lines correspond to computations that use acoustic waveguide modeling and small-deflection plate theory, respectively. The dotted line corresponds to the dispersion of a bilayer membrane of Al (0.03 μm)/AlN (0.54 μm) calculated using a multilayer waveguide model. These data indicate that the waveguide and plate theory results compare well when the acoustic wavevector is small. They also illustrate the effect of the Al overlayer.

a small amount: The Al increases the slope of the linear part of the dispersion (i.e., D') by ~2%. Multilayered systems can be treated approximately with Eq. (3) by replacing the flexural rigidity, stress, density, and thickness with corresponding composite values. The plate theory described by Eq. (2) does not, of course, include piezoelectric or anisotropic effects that can be significant in materials such as AlN. Piezoelectricity slightly increases the acoustic velocities[8] and, therefore, fitted moduli that are extracted with theory that does not include piezoelectricity. For the samples studied here, the Al overlayer electrically shorts one side of the AlN; this shorting reduces the effects of piezoelectricity on the acoustic response of the membranes. Finally, we note that the above-mentioned theory does not explicitly include the effects of the air on the response. Our previous measurements indicate that these effects are small for the typical range of acoustic wavelengths and frequencies examined with transient grating measurements.[5]

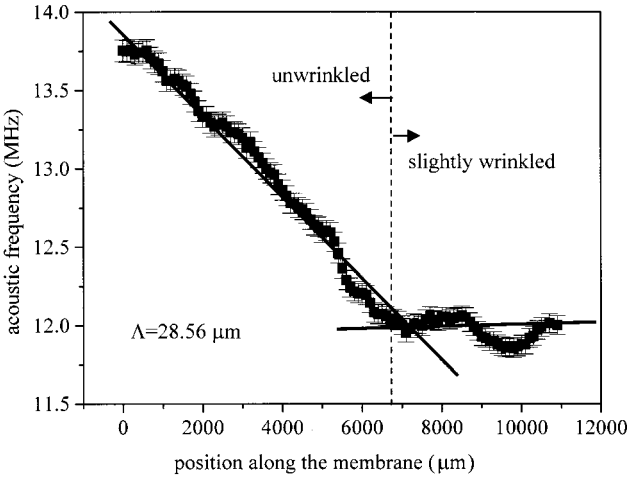


FIG. 5. Frequency of the lowest order antisymmetric Lamb acoustic mode (wavelength=28.56±0.05 μm) as a function of position along the length of a rectangular (1 cm×0.075 cm) bilayer membrane of Al (0.03 μm)/AlN (0.54 μm). The transient grating measurements that yielded these data were performed by translating the sample with respect to the excitation and probing laser beams. The acoustic wavevector lies along the direction of translation. From left to right, the frequency decreases and then levels out at a roughly constant value for the last third of the membrane. Optical inspection of the sample indicates that the onset of wrinkling (~6800 μm) corresponds roughly to the position where the frequency levels out at ~12 MHz. The solid lines are guides to the eye.

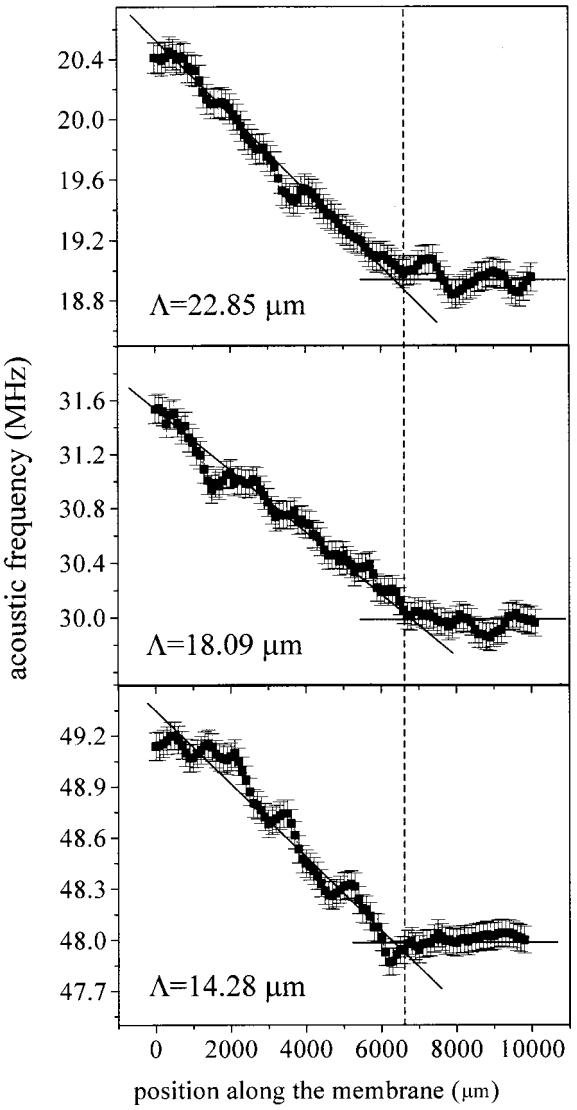


FIG. 6. Frequency of the lowest order antisymmetric Lamb acoustic mode as a function of position along the length of a rectangular (1 cm ×0.075 cm) bilayer membrane of Al (0.03 μm)/AlN (0.54 μm). In the top, middle, and bottom frames, the wavelengths of the acoustic motions are 22.85, 18.09, and 14.28±0.05 μm, respectively. In each case, the acoustic frequency ceases to vary significantly beyond ~6800 μm; this position is close to the onset of visible wrinkling in the membrane. The solid lines are guides to the eye.

IV. RESULTS

Figures 5 and 6 show measured variations in the acoustic frequency along the long axis and at the center of a rectangular membrane similar to those shown in Fig. 3. The acoustic wavelengths are 28.56±0.05 μm (Fig. 5) and 22.85, 18.09, and 14.28±0.05 μm (Fig. 6). Qualitatively, the fre-

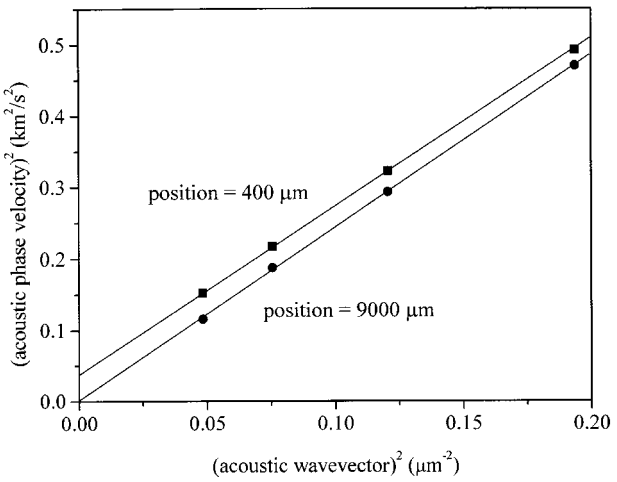


FIG. 7. Measured (symbols) variation of the square of the acoustic phase velocity with the square of the acoustic wavevector at two different spatial locations along a rectangular bilayer membrane of Al (0.03 μm)/AlN (0.54 μm). When classical plate theory is valid, this plot yields a straight line whose intercept defines the ratio of the membrane stress to the density. The slope defines the ratio of the flexural rigidity to the product of the density and the membrane thickness. The lines represent best linear fits to the data. At $x=400\,\mu$m, the membrane is in a state of tensile stress. At $x=9000\,\mu$m, visible wrinkling of the membrane suggests that the stress is close to zero.

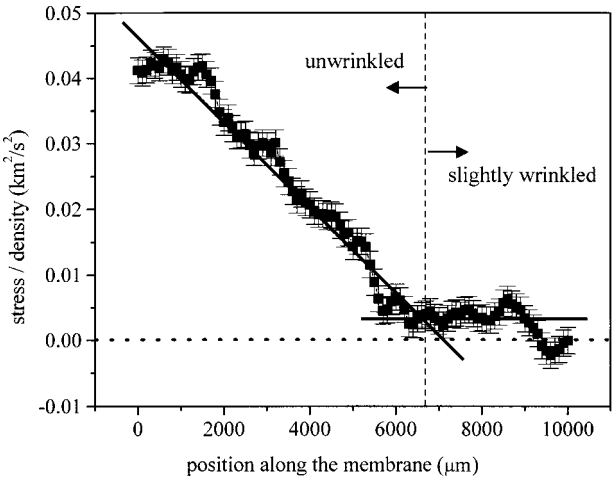


FIG. 8. Ratio of stress to density, measured as a function of position along a rectangular (1 cm$\times$0.075 cm) membrane of Al (0.03 μm)/AlN (0.54 μm). This quantity was determined from the acoustic dispersion of the lowest order antisymmetric Lamb acoustic mode. The stress (tensile) decreases steadily with position up to $\sim$6800 μm. Beyond 6800 μm, the magnitude and variation in the stress are small. Optical inspection of the membrane reveals slight wrinkling at positions greater than $\sim$6800 μm. This observation is consistent with compressive stresses in the silicon-supported bilayer that relax by buckling when the silicon is removed. The solid lines are guides to the eye.

quencies at each of the four wavelengths decrease with increasing position up to $\sim$6800 μm. Beyond this location, they cease to change significantly or systematically. If the frequency variations are dominated by changes in stress, then the results illustrated in Figs. 5 and 6 are consistent with expectations based on the appearance of the membrane: It is visibly wrinkled at positions greater than $\sim$6800 μm. We note that the measurements were performed to within $\sim$100 μm of the edges of the membranes. Even at these close distances, we observed no systematic variations in response frequencies or signal levels. These results are consistent with a spatial measurement resolution that is comparable to the size of the probing spot ($\sim$100 μm).

Analyzing the dependence of the frequencies on acoustic wavevector using Eq. (3) enables both the stress and the flexural rigidity to be determined. Figure 7 shows the dispersion measured at locations where (i) the stress is expected to be tensile ($x=400\,\mu$m) and (ii) where it is expected to be close to zero ($x=9000\,\mu$m, where the membrane is wrinkled). Linear fits to the measurements show that, indeed, the stress (y intercept of the linear fit) is nonzero at $x=400\,\mu$m ($\sigma/\rho=370\pm5\,\mathrm{m^2/s^2}$) and that it is close to zero ($\sigma/\rho=5\pm10\,\mathrm{m^2/s^2}$) at $x=9000\,\mu$m. The results also show that the flexural rigidities at these two locations are, to within $\sim$5%, the same: $D'=2.34\times10^{-6}\pm0.01\times10^{-6}\,\mathrm{m^4/s^2}$ and $2.43\times10^{-6}\pm0.02\times10^{-6}\,\mathrm{m^4/s^2}$ at $x=400$ and 9000 μm, respectively. The value of D' determined from the slope of the linear part of the dispersion calculated by waveguide theory (Fig. 4) is $2.30\times10^{-6}\,\mathrm{m^4/s^2}$, in good agreement with the measured values.

Figures 8 and 9 summarize the variation in the stress and rigidity along the length of the membrane. The stress decreases steadily to a value close to zero at $x=\sim6800\,\mu$m. There is a hint of a slight increase in rigidity with position along the length of the membrane. Neither its variation nor its magnitude change significantly, however, at $x=\sim6800\,\mu$m. The stress and the rigidity appear not to be closely related in this system. [The results in Figs. 8 and 9 use plate theory and dispersion measured at wavelengths greater than 15 μm. Waveguide calculations show that for Al(0.03 μm)/AlN(0.54μm), the plate mode approximations for this range of wavelengths introduce systematic errors that

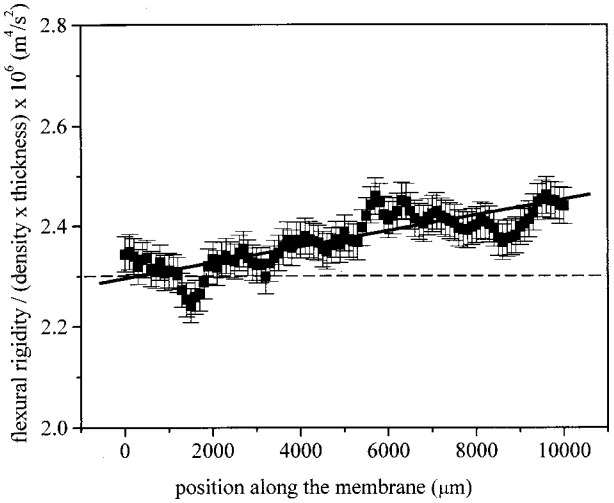


FIG. 9. Flexural ridigity divided by the product of the density and the membrane thickness, measured as a function of position in a rectangular (1 cm$\times$0.075 cm) membrane of Al (0.03 μm)/AlN (0.54 μm). This quantity was determined from the acoustic dispersion of the lowest order antisymmetric Lamb acoustic mode. Although there is a hint of an increase in the rigidity with position, the magnitude of the variation is small ($<\sim$10%). The solid line is a guide to the eye. The dashed line indicates the value computed with waveguide theory and literature values for the elastic properties of Al and AlN.

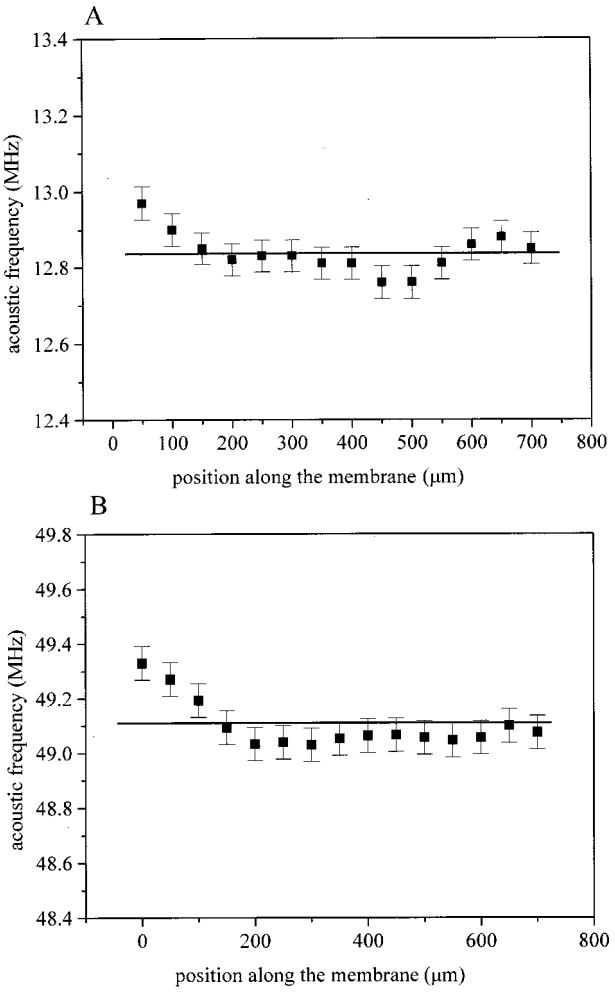


FIG. 10. Variation in acoustic frequency as a function of position across the short axis of a rectangular membrane (1 cm×0.075 cm) of Al (0.03 μm)/AlN (0.54 μm). (a) Measurements at a position $x=5000\ \mu$m along the long axis of the membrane, at an acoustic wavelength of $28.56\pm0.05\ \mu$m. (b) Measurements at $x=400\ \mu$m at an acoustic wavelength of $14.28\pm0.05\ \mu$m. These data do not reveal any significant variation in acoustic response. The solid lines are guides to the eye.

correspond to a ~2% decrease in D' and a $\sim 3\times10^{-4}\ \mathrm{m^2/s^2}$ increase in σ/ρ.]

Figure 10 shows the acoustic frequency as a function of position along the short axis of the membrane. These data do not indicate a significant variation in the acoustic response at these wavelengths (i.e., 28.56 and $14.28\pm0.05\ \mu$m for Figs. 10(a) and 10(b), respectively). Finally, Fig. 11 shows acoustic frequencies (wavelength$=22.85\pm0.05\ \mu$m) measured along the long axis of a membrane similar to the one analyzed previously, with the acoustic wavevector aligned parallel and perpendicular to this axis. To within uncertainties, the data do not reveal any significant in-plane anisotropy in the acoustic response.

V. CONCLUSIONS

This paper illustrates how a purely optical technique, known as transient grating photoacoustics, can be used to spatially map out distributions of stress and flexural rigidity in thin membranes that have nonuniform elastic properties. The measured variations in the stress are qualitatively consistent with expectations, and the flexural rigidity agrees with computations that use literature values for mechanical properties. Uncertainties in the TG measured stress and rigidity are typically determined by the precision with which the acoustic frequency can be determined from the data. This precision, in turn, is often related to the signal to noise ratio of the measurement. In the case of the sample analyzed here, however, the uncertainties are dominated by coherent interference (e.g., heterodyning) of diffracted signal with parasitically scattered light, especially in the regions of the membranes that are wrinkled. This heterodyning qualitatively changes the form of the signal and systematically alters, in subtle ways, the shape of the peak in the power spectrum. It is generally difficult to account accurately for these effects because the phase of the scattered light and, often, its intensity are not known precisely. As a result, this hetereodyning can alter slightly the frequency determined by fitting peaks in the power spectra. To eliminate this source of uncertainty, it is possible to introduce a separate, relatively intense, reference beam for heterodyne detection.[8,20,21] Interference of diffracted light with this beam then dominates the signal so that the effects of slight point to point variations in scattered light are minimized.

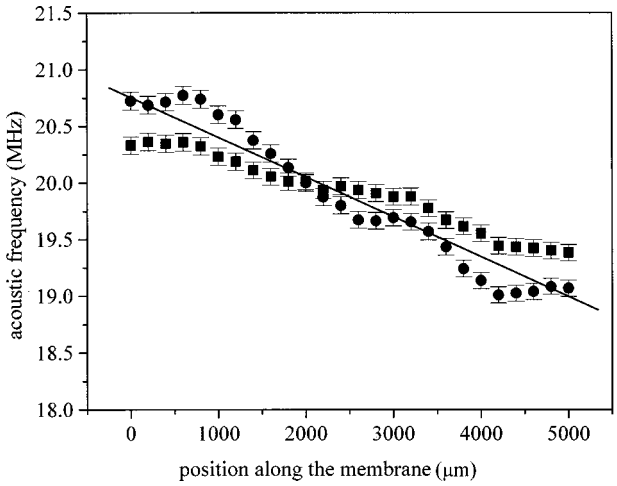


FIG. 11. Variation in acoustic frequency as a function of position along the long axis of a rectangular membrane (1 cm×0.075 cm) of Al (0.03 μm)/AlN (0.54 μm), at its center. The squares and circles correspond to data recorded with the acoustic wavevector (wavelength$=22.85\pm0.05\ \mu$m) aligned parallel and perpendicular to the long axis of the membrane, respectively. These data do not indicate any significant anisotropy in the acoustic response. The solid line is a guide to the eye.

Finally, an important feature of the TG approach is its relatively high spatial resolution. Other mechanical techniques, such as the resonance frequency method and the bulge test, rely on drumhead vibrations and/or static displacements of entire membranes. These types of approaches typically do not allow for spatially resolved measurement: They yield properties that represent averages over the area of the membrane. The good spatial resolution of the TG measurement derives from the small (~100 μm) sizes of the excitation and probing laser beams and from the relatively short wavelengths (~10 μm) of the acoustic motions. The resolution is ultimately limited to values comparable to the acoustic wavelength: Reducing the wavelength improves the

possible resolution. As the wavelength decreases, however, the effects of stress on the response decrease. The plate mode description of the dispersion also becomes incomplete when the wavelength is comparable to or smaller than the membrane thickness. The elastic properties and thickness of the membrane, the precision of the measurement, and the demands of the application will determine the balance between the resolution afforded by short wavelengths and the accuracy and simplicity of data analysis enabled by long wavelengths. The spatial resolution that can be easily achieved with membranes commonly found in microelectromechanical systems, microfluidic devices, and acoustic filters, however, is significant. We believe that the TG approach will be useful for evaluating materials and structures for these and other applications.

[1] G. T. A. Kovacs, *Micromachined Transducers Sourcebook* (McGraw–Hill, New York, 1998).

[2] A. D. Wilson, ''X-ray-lithography in IBM 1980–1992, the development years,'' IBM J. Res. Dev. **37**, 299–318 (1993); H. I. Smith and M. L. Schattenburg, ''X-ray-lithography, from 500 to 30 nm—X-ray nanolithography,'' *ibid.* **37**, 319–329 (1993).

[3] J. A. Liddle *et al.*, ''The scattering with angular limitation in projection electron-beam lithography (SCALPEL) system,'' Jpn. J. Appl. Phys., Part 1 **134**, 6663–6671 (1995); L. R. Harriott, ''Scattering with angular limitation projection electron beam lithography for suboptical lithography,'' J. Vac. Sci. Technol. B **15**, 2130–2135 (1997).

[4] M. P. Schlax, R. L. Engelstad, E. G. Lovell, J. A. Liddle, and A. E. Novembre, ''Stress mapping techniques for the SCALPEL mask membrane system,'' 1999 SPIE Symposium on Emerging Lithographic Technologies III (in press).

[5] J. A. Rogers and G. R. Bogart, J. Mater. Res. (in press).

[6] M. A. Maden, A. Jagota, S. Mazur, and R. J. Farris, ''Vibrational technique for stress measurement in thin films. 1. Ideal membrane behavior,'' J. Am. Chem. Soc. **77**, 625–635 (1994).

[7] J. J. Vlassak and W. D. Nix, ''A new bulge test technique for the determination of Young modulus and Poisson ratio in thin films,'' J. Mater. Res. **7**, 3242–3249 (1992).

[8] J. A. Rogers, A. A. Maznev, M. J. Banet, and K. A. Nelson, ''Optical generation and characterization of acoustic waves in thin films: Fundamentals and applications,'' Annu. Rev. Mater. Sci. **30**, 117–157 (2000).

[9] K. A. Nelson and M. D. Fayer, ''Laser induced phonons: A probe of intermolecular interactions in molecular solids,'' J. Chem. Phys. **72**, 5202–5218 (1980).

[10] H. J. Eichler, P. Gunter, and D. W. Pohl, *Laser-Induced Dynamic Gratings* (Springer, Berlin, 1986).

[11] A. Harata, Q. Shen, and T. Sawada, ''Photothermal applications of lasers: Study of ultrafast photothermal phenomena at metal-liquid interfaces,'' Annu. Rev. Phys. Chem. **50**, 193–219 (1999).

[12] Y. Yan and K. A. Nelson, ''Impulsive stimulated light scattering. 1. General theory,'' J. Chem. Phys. **87**, 6240–6256 (1987).

[13] R. Bosch, US patent #5501893, German Patent DE 4241045C1.

[14] A. A. Ayon, R. Braff, C. C. Lin, H. H. Sawin, and M. A. Schmidt, ''Characterization of a time multiplexed inductively coupled plasma etcher,'' J. Electrochem. Soc. **146**, 339–349 (1999).

[15] J. A. Rogers, L. Dhar, and K. A. Nelson, ''Noncontact determination of transverse isotropic elastic moduli in polyimide thin films using a laser based ultrasonic method,'' Appl. Phys. Lett. **65**, 312–314 (1994); J. A. Rogers and K. A. Nelson, ''Photoacoustic determination of the residual stress and transverse isotropic elastic moduli in thin films of the polyimide PMDA/ODA,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 555–566 (1995).

[16] J. S. Meth, C. D. Marshall, and M. D. Fayer, ''Experimental and theoretical analysis of transient grating generation and detection of acoustic waveguide modes in ultrathin solids,'' J. Appl. Phys. **67**, 3362–3377 (1990).

[17] A. C. Ugural, *Stresses in Plates and Shells*, 2nd ed. (McGraw–Hill, New York, 1999).

[18] G. Carlotti, G. Gubbiotti, F. S. Hickernell, H. M. Liaw, and G. Socino, ''Comparative study of the elastic properties of polycrystalline aluminum nitride films on silicon by Brillouin light scattering,'' Thin Solid Films **310**, 34–38 (1997).

[19] G. Simmons and H. Wang, *Single Crystal Elastic Constants and Calculated Aggregate Properties: A Handbook*, 2nd ed. (MIT, Cambridge, 1971).

[20] J. A. Rogers, M. Fuchs, M. J. Banet, J. B. Hanselman, R. Logan, and K. A. Nelson, ''Optical system for rapid materials characterization with the transient grating technique: Application to nondestructive evaluation of thin films used in microelectronics,'' Appl. Phys. Lett. **71**, 225–227 (1997).

[21] A. A. Maznev, J. A. Rogers, and K. A. Nelson, ''Optical heterodyne detection of laser-induced gratings,'' Opt. Lett. **23**, 1319–1321 (1998).

Bender transducer design and operation

John L. Delany[a)]
Ultra Electronics Ltd., Sonar and Communication Systems, Bridport Road, Greenford, Middlesex UB6 8UA, United Kingdom

(Received 10 June 2000; accepted for publication 28 November 2000)

An empirical study covering a wide range of bender transducer sizes and operating frequencies is reported. A spherical device model is shown to give good account of bender performance, including interaction effects. A set of empirical rules for scaling equivalent circuit parameters according to the device geometry is identified. An effective spherical radius, approximately half the diaphragm radius, is identified for the typical bender. The effects of pressure and drive voltage on performance are described for particular devices. Sensitivity factors for the equivalent circuit parameters to the operating conditions are determined. These are related to sensitivity factors for the coupling coefficient (*Kc*) and electromechanical transformer turns ratio (*N*). Both these parameters are shown to have similar sensitivity responses, decreasing with pressure (planar stress), and increasing with voltage (electric field). The results of high drive tests carried out at Seneca Lake are reported. Values of conventional figures of merit (FOM_V and FOM_M), close to the highest claimed for any underwater transducer, are tabled. When allowance for the operating efficiency is included in the figure of merit definitions, the bender appears to be superior to other device types. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1342002]

PACS numbers: 43.38.Fx, 43.30.Yj [SLE]

I. INTRODUCTION

This paper is concerned with a type of underwater acoustic transducer known as the bender. This is a flexural mode device that derives its name from a bending action that is similar to that in a bimetal thermostat. In the bender, the action is produced electrically by a bilaminar structure that incorporates a piezoelectric plate. The effect has been known since the discovery of piezoelectricity,[1] and was first exploited in the 1930's in a device named the bimorph.[2] A surge of interest in benders for underwater applications followed the development of piezoelectric ceramics, and in particular PZT (PZT is the generic name for ceramic formulations based on lead zirconate titanate) in the early 1950's. Over the next decade, many device designs were produced,[3] with the Navy Underwater Sound Laboratory (New London) playing a lead role. While the basic device mechanism was noted to be simple in concept, its detailed analysis was generally found to be complex. Woollett, however, developed basic treatments for disk[4] and bar[5] benders, providing insight into the influence of the diaphragm support, the choice of materials and the dimensions. Since from the 1970's, there has been a general reduction in the number of published reports on benders, with interest generally switching to flextensional types of transducer.[6]

The bender considered here is a disk type, employing a circular diaphragm structure formed by bonding a piezoelectric ceramic to a metal plate. A short cylindrical section supports a pair of such diaphragms, thereby enclosing an air cavity, as shown schematically in Fig. 1. A typical arrangement is for each ceramic to be bonded with its poled direction such that an applied voltage causes both diaphragms to flex together, inwards or outwards. Encapsulating the assembly within a thin elastomeric coating provides waterproofing for a practical device. Both projector and hydrophone operation is available, the latter being achieved by sensing the electrical output produced by the (pressure wave) induced flexure of the diaphragm.

It has been long recognized that the above type of construction is well suited to applications in which compactness and high efficiency are the paramount design considerations. These are the general requirements for autonomously operating systems, such as sonobuoys, and have been the stimulus for an empirical study into the design, operation, and limiting mechanisms of the bender disk transducer. The current study has covered a wide range of device geometry and resonance frequency. It has demonstrated the usefulness of equivalent circuit modeling to characterize device operation, including nonlinear behavior under extreme conditions. It has also allowed a general design rationale to be developed for identifying optimum drive system configurations. Further, it has led to the development of transducer designs that have achieved among the highest output powers observed from any type of transducer of similar volume or mass.

The following starts with a summary of the range of benders that has been constructed and tested. There is then discussion of an equivalent circuit model for the bender, with examples which support its validity. A set of general scaling rules that relate the equivalent circuit elements to the geometry is presented. Measurements showing the typical effects of pressure and electrical drive are also reported. Finally, examples of the high drive performance observed with particular prototype designs are presented and compared with other transducer types.

II. DEVICE CONSTRUCTION RANGE

Benders have been constructed using a range of diaphragm materials and piezoelectric ceramic types. A selec-

a)Electronic mail: jld@ultra-scs.com

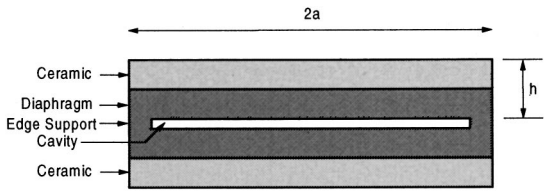


FIG. 1. Cross-section schematic of bender disk transducer.

tion of devices employed by current systems is shown in Fig. 2. For this paper, results and discussion are restricted to devices built using diaphragms of a particular metal, and a ''hard'' type of PZT classified[7] as Navy Type-III ceramic. It should be noted that various grades of the metal have been employed, resulting in a range of ultimate depth capabilities. Also, the ceramics have been procured from a number of manufacturers.

A common configuration, with equal diameters and thickness ratio for the diaphragm components, has been adopted. In all cases, the ceramic has been bonded externally to the supporting diaphragm, thereby maintaining compressive planar stress within the ceramic under hydrostatic loading. Diaphragms of radius (a) ranging from <20 mm to >100 mm have been employed, with the composite thickness ratios (h/a) being within the range 0.1 to 0.3. All diaphragms have been separated by a short, thin-walled cylindrical section, giving a typical air gap length of several millimeters.

The above construction has yielded devices with in-water resonance frequencies between 800 Hz to 20 kHz. In terms of the propagating wavelength (λ) at the resonance frequency (f), all devices have been small, with radii in the range 0.05λ to 0.2λ. In wave number ($k=2\pi/\lambda$) terms, the typical *ka* value is less than unity. All devices have been observed to exhibit omnidirectional acoustic responses over the resonance band and below. The typical resonance has a Q factor of between 10 and 12, and the acoustic output efficiency at low drive is greater than 95%.

III. EQUIVALENT CIRCUIT MODEL

All the constructed benders have exhibited a fundamental resonance response that can be represented in terms of a

FIG. 2. Examples of bender devices currently employed by developmental systems. The scale is shown by the 300-mm ruler.

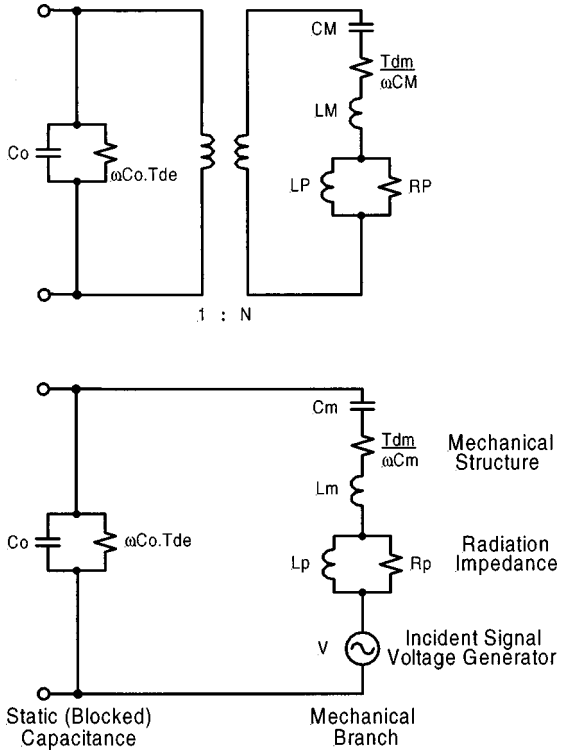


FIG. 3. Equivalent circuit for piezoelectric transducer with spherical radiating surface (radius, r). The upper circuit shows the full transducer model, incorporating an electromechanical transformer (turns ratio N). The lower circuit is a simplified circuit, with transformer effect included in circuit element values (i.e., $Rp=RP/N^2$). The signal generator (V) included in the lower circuit represents the effect of an incident pressure signal (p), and is given by $V=pAD/N$, where A is area, and D is the diffraction constant.

spherical transducer equivalent circuit model, as shown in Fig. 3. The upper circuit includes a transformer to represent the electromechanical coupling produced by the piezoelectric effect. In the lower circuit, this effect is incorporated in the circuit element values. The basic lumped element equivalent circuit is well known,[8] and is a development of the Van Dyke resonator circuit. In this, the mechanical (resonating) branch is split into components representing the transducer mechanical structure and the propagation medium load (radiation impedance). For a spherical source of radius r, the radiation impedance is represented by a parallel combination of fixed resistance (Rp) and inductance (Lp), such that $Lp/Rp=r/c$, where c is the sound velocity in the medium. The standard spherical source equivalent circuit parameters are given by $RP=Z_0A$, $LP=(r/c)RP$, $N=\sqrt{(Z_0A/Rp)}$, $D=1/\pi(1+(kr)2)$, and $A=4\pi r^2$, where Z_0 is the characteristic impedance of the medium. The signal source, shown in series with the radiation impedance, allows for the effect of an incident pressure signal. Electrical and mechanical loss within the transducer is represented in Fig. 3 by conductance and resistance elements defined, respectively, in terms of loss tangents (*Tde*, and *Tdm*) associated with the static and dynamic capacitors (*Co* and *Cm*). It should be noted that the loss parameter definitions are equivalent in the low loss limit, i.e., loss tangent <0.02.

A simple procedure has been developed for determining the equivalent circuit parameters. This is based on measurements of impedance made under low drive voltage (1 V

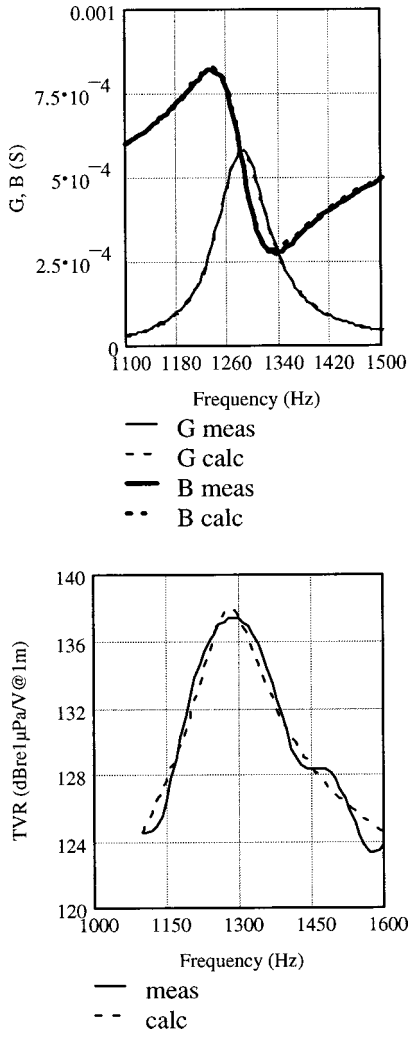


FIG. 4. Typical bender performance characteristics in water at resonance, as measured and as calculated. The upper graph shows the impedance response expressed in terms of conductance (*G*) and susceptance (*B*). The lower graph shows the corresponding transmitting voltage response (TVR).

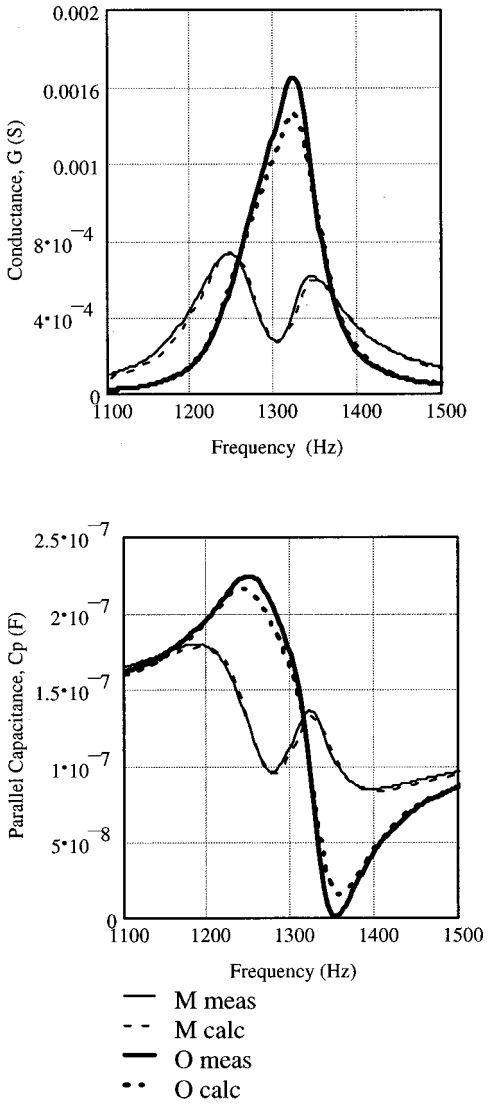


FIG. 5. Impedance responses for a bender pair, showing the effects of mutual interaction at a spacing of 0.2λ Matched polarity is indicated by the letter M, opposed polarity by the letter O; measurement and calculation are as indicated.

rms), using an analyzer type HP4192A or 4194A. Each device is tested in air and in water, and characterized over the resonance band and at a spot low frequency (typically 1/10th the resonance frequency). All data are recorded on floppy disk, and an iterative analysis is followed, using MATHCAD, to derive values for the equivalent circuit elements. A typical impedance response, expressed in terms of conductance (*G*) and susceptance (*B*), is shown in the upper graph of Fig. 4. The solid traces are the measurements; the broken traces are the calculated responses, derived from the equivalent circuit analysis. The agreement between the measurement and calculation is seen to be very close.

Support for the model comes from its ability to reliably predict the device transmitting voltage response, TVR, shown by the lower graph in Fig. 4. The TVR is defined as the transmitted acoustic pressure referenced to a range of 1 m and drive voltage of 1 V rms. The acoustic pressure referenced to 1 m is known as the source level. The device characterized here has a diaphragm diameter of 100 mm, and for this measurement was tested at a water depth of 5 m. The ''ripple'' on the measured acoustic response is attributed to a reverberation effect. This arose here as a consequence of the measurement technique and test configuration. A gated sine wave technique was employed, with a 100-V rms drive pulse of duration 100 ms, with the measurement widow of 40-ms duration, starting 50 ms from the pulse turn on. The separation between test device and a ball hydrophone was 0.5 m, and the measurement depth was 5 m. This caused the hydrophone to record a multipath signal, comprising the direct path transmission and an indirect path signal reflected from the water surface. The broken trace is that calculated using the derived equivalent circuit parameters, and is seen to fall within the measured response.

Perhaps more striking support for the equivalent circuit model comes from its ability to account for the mutual coupling effect between devices. This is demonstrated by the set of measured and calculated impedance responses shown in Fig. 5. A pair of devices, of the type used for Fig. 4, was employed, with a center-to-center spacing of 230 mm (ca. 0.2λ at resonance). These were connected in parallel, first with the polarities matched, and then in opposition. The in-

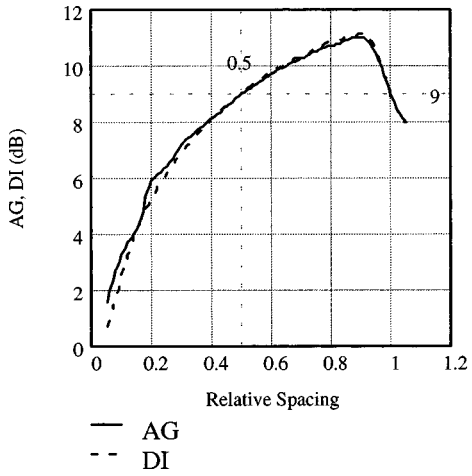


FIG. 6. Array gain (AG) calculated for a uniform 8-element array, plotted as a function of the relative element spacing (d/λ). The AG trace is calculated from the equivalent circuit model, and the Dl trace is derived using Urick's formula (Ref. 9).

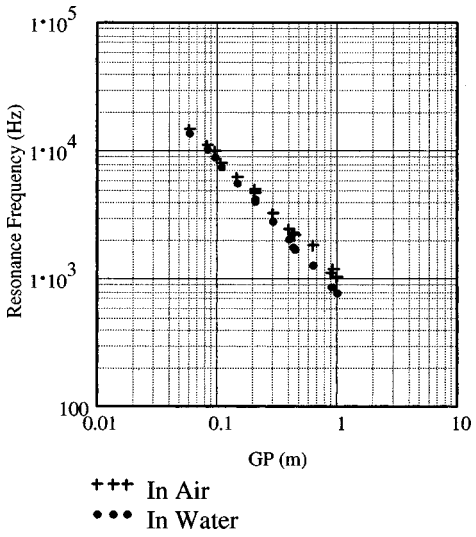


FIG. 7. Bender resonance frequency, measured in air (*fa*) and in water (*fw*), plotted as a function of the device geometry parameter, $GP(=a^2/h)$.

teraction effect is evident from the different impedance responses for these two conditions. In each plot, the solid traces are the measurement and the broken traces the calculation. The calculated responses are based on the circuit parameter sets derived from the individual device impedance responses. In the analysis, the different polarity connections are treated using the electromechanical transformer to define the phase of the drive signal applied to each mechanical branch of the equivalent circuit. Referring to the circuit, the calculation involves solving a pair of simultaneous equations representing the voltages developed across the mechanical branch components. The voltage across each radiation impedance generates the transmitted pressure signal, and the series generator is the voltage produced by the pressure signal from the neighboring element. The latter signal is delayed and attenuated according to the device separation. The agreement between the measured and calculated responses is seen to be close.

Based on knowledge of a typical equivalent circuit parameter set, the interaction effect calculation has been extended to various line-array configurations. Figure 6 shows the array gain (AG) calculated for an array of $N=8$ identical elements, as a function of d/λ (element spacing relative to operating wavelength). It is to be noted that there is close agreement with the directivity index (DI) calculated for a uniform receive array using a standard formula quoted by Urick.[9] In this example, both calculations predict an array gain of 9 dB for half-wavelength spacing. This is the array gain calculated for the zero interaction case, corresponding to infinite element spacing.

IV. DEVICE SCALING

Impedance responses have been recorded for the device set discussed in Sec. II. The resonance frequencies, measured in air (*fa*) and in water (*fw*), are plotted in Fig. 7 in terms of the geometry parameter, GP, defined by a^2/h. From the logarithmic scaling of the plots, it is evident that the resonance frequency in water is nominally directly proportional to the reciprocal of GP (i.e., $fw \propto h/a^2$). While for the in air case, the dependence is slightly sub-linear, given approximately by $fa \propto (h/a^2)^{0.9}$.

Using the equivalent circuit model, sets of circuit parameters (*Co, Cm, Lm, Rp, Lp, Tde,* and *Tdm*) have been determined for each device type. Apart from the loss tangents, the circuit element values are also observed to be related to the device geometry. This is evident from Fig. 8, in which normalized capacitance and inductance values are plotted as functions of GP. The normalization is made using the equivalent circuit parameters derived from the impedance response of the 100-mm diameter device shown in Fig. 4. The data points are seen to be distributed about notional straight lines, thereby indicating simple power-law dependence on GP (i.e., parameter$\propto$GP^m). With the circuit parameter signified by the corresponding subscript, the nominal power laws are defined by $m_{Co}=1.03\pm0.04$, $m_{Cm}=1.18\pm0.06$, $m_{Lm}=0.69\pm0.05$, and $m_{Lp}=1.22\pm0.09$. The general spread of points evident in these plots is attributed to differences in measurement conditions, device material properties, and detail of the diaphragm edge support.

Plots of the equivalent circuit resistance (*Rp*) as a function of GP reveal a significantly greater scatter of points than observed with the other circuit parameters, thereby suggesting a more complex relationship between *Rp* and GP. This view is reinforced when the *Rp* values are expressed in terms of an effective relative radius (r/a), where the effective radius $r(=cLp/Rp)$ is the equivalent spherical radius, and *a* is the actual diaphragm radius. This is shown by the bottom graph in Fig. 8, which indicates that r/a is insensitive to GP, with a mean value of 0.50 ± 0.04. It is noteworthy that the effective spherical radius for a plane circular piston in an infinite baffle[10] is approximately 0.59, in the limit $ka \ll 1$.

V. INFLUENCE OF PRESSURE

The various test benders have been required for systems typically operating with maximum depth limits between 200 and 450 m, corresponding to nominal pressure limits of 2

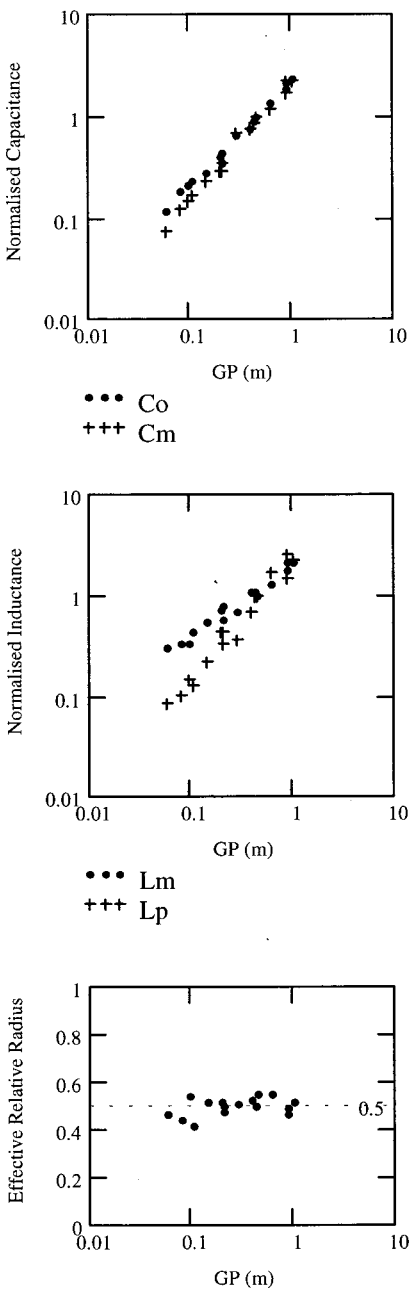


FIG. 8. Normalized equivalent circuit parameter plots as a function of the bender geometry parameter, GP. The top graph shows normalized *Co* and *Cm*; the middle plot shows normalized *Lm* and *Lp*. The bottom plot displays the effective spherical radius relative to the actual diaphragm radius (i.e., r/a).

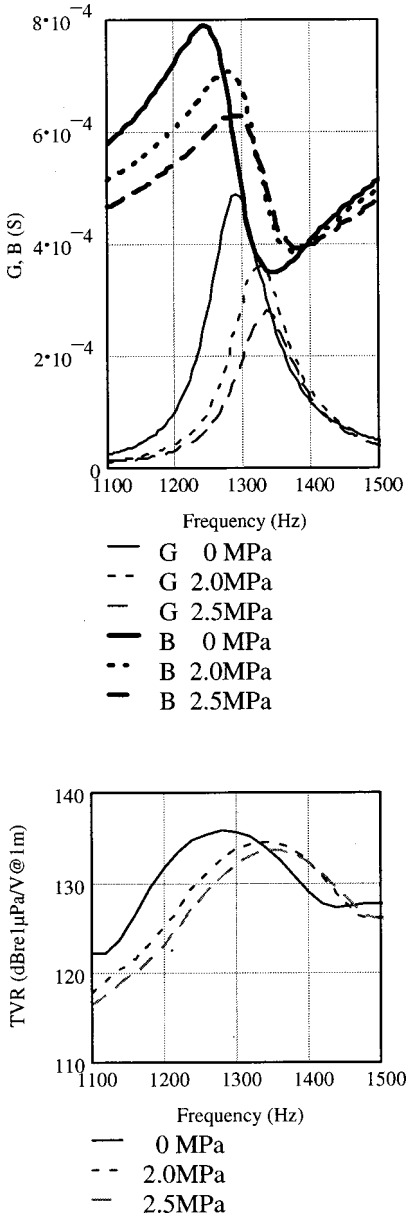


FIG. 9. Plots showing the performance sensitivity to pressure for a typical bender. The top graph gives the impedance response, plotted in terms of conductance (G) and susceptance (B). The lower graph shows the TVR measured within a pressure vessel at shallow depth. Operating pressures (P) are as indicated.

and 4.5 MPa. Accordingly, it has been important to characterize the performance sensitivity to depth, and to identify limits for reliable operation. This has been facilitated by the use of a thin-wall Kevlar pressure vessel, having external diameter of 1.2 m and length of 2 m, available at the UK Acoustic Test Centre (Underwater Acoustic Calibration Centre, Wraysbury Reservoir, Wraysbury, Middlesex TW19 5NW. Operated since 1999 by Neptune Sonar Ltd.). All tested devices have shown a similar sensitivity to pressure, dependent on the bender geometry. Figure 9 shows a selection of impedance (G, B) and transmitting voltage response (TVR) measured for a device of the type shown in Fig. 4. Again, a ''ripple'' effect is evident on the TVR curves, and is attributed to surface reverberation, with an additional contribution from the pressure vessel structure. The measurements show that the peak conductance and transmitting voltage response values decrease with pressure, accompanied by an increase in resonance frequency. These changes have been observed to be reversible below a certain pressure limit, and are attributed to changes in the piezoelectric properties with increasing planar compressive stress within the PZT ceramic. An ultimate limit to the practical operating depth tends to be set by the ceramic cracking, this being induced by plastic deformation (creep) within the supporting diaphragm.

It has been possible to relate the pressure-induced changes of impedance to corresponding changes in the

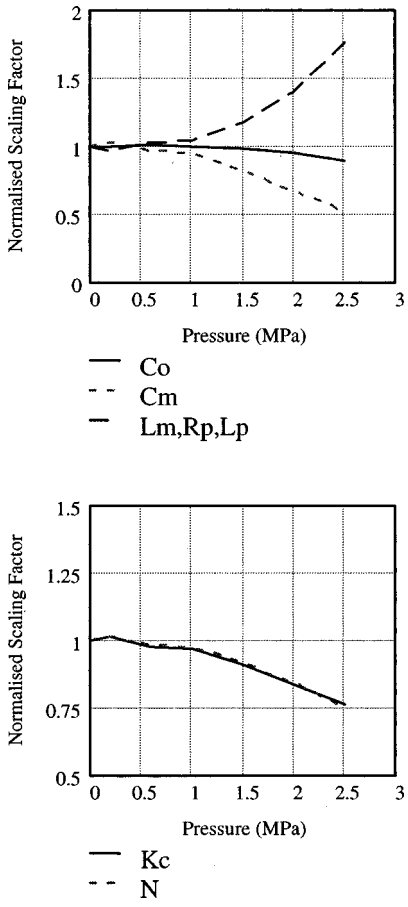


FIG. 10. Normalized scaling factors as a function of pressure, derived from the impedance responses shown in Fig. 9. The upper graph gives the circuit element sensitivities; the lower graph gives the normalized coupling coefficient (*Kc*) and transformer turns ratio (*N*).

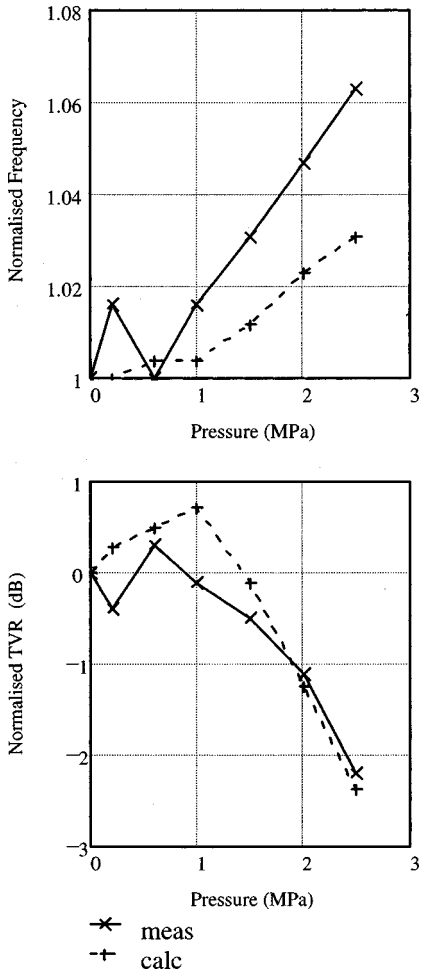


FIG. 11. Normalized resonance frequency and peak TVR as a function of pressure, as measured and calculated using the sensitivity parameters derived in Fig. 10.

equivalent circuit parameters. For modeling purposes, it has been assumed that the primary effect of pressure is to modify the capacitor elements (*Co, CM*), and the electromechanical transformer (*N*). The other components to the right of the transformer are assumed to be insensitive to the pressure, on the basis that *LM* is related to the diaphragm mass, and *RP* and *LP* are related to the diaphragm area. An iterative procedure has been followed, using MATHCAD, to track the measured impedance changes and calculate the corresponding changes to the ambient pressure equivalent circuit parameters. This has allowed a set of normalized sensitivity factors to be computed for each of the circuit components defined in Fig. 3. Sensitivity factors are defined by expressing the circuit parameter values under the test condition (pressure, P) to those under the initial test state, e.g., $\mathrm{Co}(P)/\mathrm{Co}$. The upper graph in Fig. 10 shows the sensitivity factors derived from the impedance changes shown in Fig. 9. These are displayed, in the lower graph, in terms of normalized characteristics of the transformer turns ratio (N), and coupling coefficient (Kc), the latter being defined by $\sqrt{(Cm/(Co+Cm))}$. These parameters exhibit closely matched responses to pressure, decreasing with increasing pressure. There is an accompanying increase in electrical loss and decrease in mechanical loss with pressure. Figure 11 shows corresponding plots of the normalized calculated resonance frequency and transmitting voltage response (TVR) as a function of pressure (P).

Given that the measured acoustic response is complicated by the effect of reverberation, the general agreement between the measurements and prediction is considered to validate the model.

VI. INFLUENCE OF DRIVE LEVEL

An aim has also been to identify safe limits for reliable bender operation under pulsed drive (typically 0.1- to 10-s duration, and 5% duty cycle). Accordingly, the effect of drive has been investigated by testing devices in air and in water. Both types of measurement show a slight reduction of resonance frequency with increasing drive level, typically larger than that expected to arise from internal heating. The type of device design being considered here exhibits a relatively low sensitivity to temperature. For example, raising the temperature from 0 to 60 °C produces a downward shift of resonance frequency in air that is typically $<0.4\%$. This is small compared to the shifts that can be produced under short-pulse, low-duty cycle, operation at relatively low drive levels. Increasing the drive at resonance in air tends to produce changes that are reversible up to a limit, beyond which small but permanent reductions in resonance frequency are typically observed. These frequency changes are generally not accompanied by distortion to the shape of the impedance

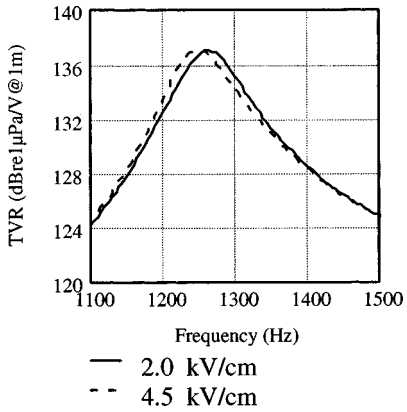


FIG. 12. Transmitting voltage response (TVR) at different drive fields, as indicated, measured at 60-m water depth, for a bender of the type shown in Fig. 4.

response. They also occur with no sudden change to the peak conductance value. This suggests that they may be caused by mechanical damage within the ceramic on a granular scale, rather than by gross fracture or rupture of the glue bond to the metal diaphragm.

High-drive testing in water has been carried out with various devices until the onset of device failure either by electrical or mechanical breakdown. The former is typically characterized by localized electrical punch-through from the ceramic outer surface. Mechanical failure is typically characterized by the extensive surface crazing and cracking of the ceramic.

High-drive measurements have been carried out at the UK Acoustic Test Centre, and at the Sonar Test Facility at Lake Seneca (NY). The Seneca Lake Sonar Test Facility is operated by the Naval Undersea Warfare Center, Newport, RI. The latter facility offers the advantage of permitting deep-water deployment, down to about 160 m. This allows the surface reverberation effects discussed earlier to be readily avoided. In addition, the Seneca Lake test system is able to measure impedance and source level under pulsed conditions at high-drive levels. This has allowed the influence of drive on the bender response to be examined in detail. The TVR response in Fig. 12 shows the typical decrease in resonance frequency with increase in drive voltage, shown here for a device of the same type as in Fig. 4, measured at 60-m depth. It is to be noted that there is no sign of the ripple effect evident in the earlier figure.

A decrease in the resonance frequency with drive is also observed when any device is operated in air. This is to be seen in the upper response of Fig. 13, measured with a large diameter (240-mm) bender operated under short pulse (0.1 s) drive in air. A set of such responses has been analyzed using the transducer equivalent circuit model with the radiation impedance shorted out to represent the air load condition. Normalized values of Lm and $1/Cm$ as a function of drive field are shown in the lower plot of Fig. 13. The closeness of the normalized responses indicates that the changes in these parameters are linked by a common mechanism, related to the electric drive field in the ceramic.

The above 240-mm bender has also been measured at different drive levels in water, at Seneca Lake. Its impedance

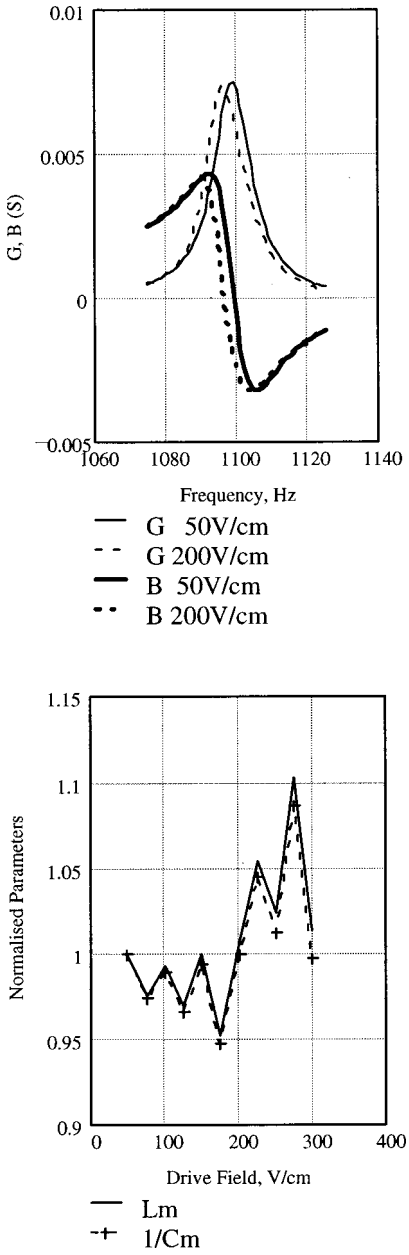


FIG. 13. Plots showing the effect of electrical drive on the in-air resonance response of a 240-mm diameter bender. The top graph shows the conductance (G) and susceptance (B) responses measured at drive levels corresponding to 50 and 200 V/cm. The bottom graph shows the equivalent circuit parameters Lm and $1/Cm$, normalized to the lowest drive field values, as a function of the drive field.

behavior has then been analyzed in a similar fashion to that used to characterize the pressure effect, in Sec. V. That is, it has been assumed that the drive voltage (field) only has direct influence on the transformer turns ratio (N) and the equivalent circuit capacitors (Co and Cm). Accordingly, the normalized coupling coefficient (Kc) and transformer turns ratio (N) sensitivities to drive voltage have been deduced, as shown in Fig. 14. It is evident that the changes in these parameters are closely matched. However, the effect of drive is to increase their magnitudes, the opposite of the pressure effect. The corresponding effect of drive on the electrical and mechanical loss tangents is an increase in the loss. Finally, to validate the model, Fig. 15 compares the measured and cal-

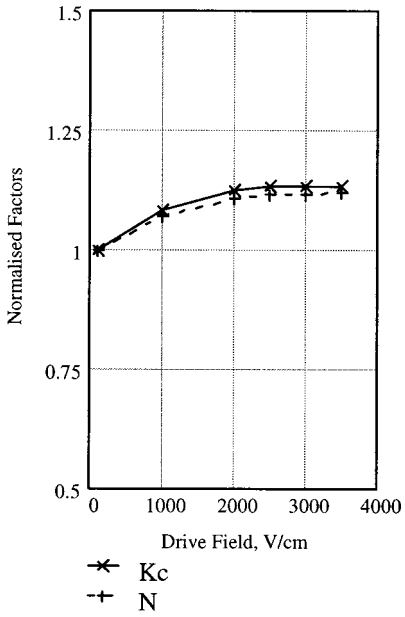


FIG. 14. Normalized coupling coefficient (*Kc*) and electromechanical transformer turns ratio (*N*), as a function of drive field, derived for a 240-mm diameter bender.

culated normalized resonance frequency and peak TVR responses to drive. Given the small scale of the changes, there is seen to be close agreement between the curves, thereby supporting the analysis.

VII. ACOUSTIC PERFORMANCE LIMITS

High-drive tests have been conducted at the Lake Seneca test site for several prototype bender designs. These have demonstrated high acoustic performance capability, as illustrated by Table I. This lists the measured resonance frequency (fw), quality factor (Q), source level (SL), efficiency (η), mass (M), and two figure of merit parameters (FOM_V and FOM_M). These figures of merit are defined in Ref. 6 as follows: FOM_M=radiated power(W)/[mass(kg)×frequency(kHz)×quality factor]; FOM_V=radiated power(W)/[volume(m^3)×frequency(Hz)×quality factor]. Also included in the table are the performance characteristics that have been published for other underwater transducer designs. It is evident that the bender performance is among the highest reported for any type of transducer.

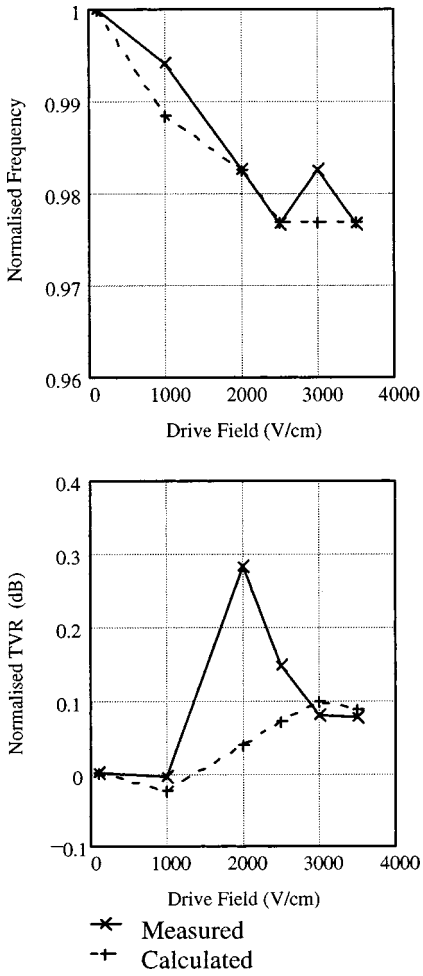


FIG. 15. Drive field sensitivities of normalized resonance frequency and peak transmitting voltage response for a 240-mm diameter bender. Measured and calculated data points are as indicated.

It should be noted that the defined figures of merit do not fully reflect the suitability of a transducer for certain types of acoustic system. In particular, for battery-powered systems, design viability usually depends on being able to meet an acoustic requirement within a defined volume or weight con-

TABLE I. Sonar projector performance Summary. Device data, other than for benders, are taken from Ref. 6. SL units are dB *re* 1 μPa at 1 m, FOM_M units are W/(kg-kHz), and FOM_V units are W/(m^3-Hz). For the benders, the conversion factor relating dB *re* 1 W to dB *re* 1 μPa at 1 m is 170.5, corresponding to a fresh-water sound speed of 1423 m/s.

Projector description	fw (Hz)	Q	SL(dB)	η (%)	M (kg)	FOM_M	FOM_V
Bender (240-mm diam.)	860	10	212.3	80	12.4	140	680
Bender (100-mm diam.)	1 730	10	203.6	80	0.82	140	645
DREA PZT Barrel-stave large Class I	790	3.6	204.3	81	16.7	47	162
NUWC/Sanders Terfenol-D ''dogbone'' Class VII	930	4.7	212.0	46	15.4	210	710
Allied Signal DX835E PMN Class IV	2 500	4.6	210.2	43	10	83	185
Raytheon copolymer projector	14 600	2.4	214.0	26	1.0	80	331

straint. In this situation, the available system efficiency is also a key factor in the selection of a suitable transducer. Multiplying the conventional FOM by the device efficiency (η) provides a means for taking this factor into account. On this basis, the bender is seen to offer a higher system performance than the listed alternatives.

VIII. CONCLUSIONS

The reliance of the fundamental resonance of the bender on diaphragm flexure provides a simple mechanism for achieving a low resonance frequency from a device that is small compared to the propagating wavelength. The small relative size accounts for the success of the spherical transducer equivalent circuit model. This has provided a reliable means for predicting the acoustic response from impedance measurements. Accordingly, with measurements made under effective free-field conditions, the impedance response has been found to provide a reliable means for checking the accuracy of hydrophone calibrations.

The range of device designs studied here has provided a simple set of scaling rules for the equivalent circuit parameters. The identification of an effective spherical radius (r) that is nominally half the diaphragm radius (a) indicates that the behavior approaches that of a circular piston in an infinite baffle. Given the scaling model, it has been possible to predict the performances available from a wide range of device geometries, and accordingly to identify geometry options for particular performance requirements.

Measurements of devices under different operating conditions have revealed performance sensitivities to pressure and drive voltage. It has been found that the effect of pressure can be modeled by assuming that its direct effect is restricted to the equivalent circuit capacitors and transformer turns ratio. On this basis, the pressure sensitivities of the coupling coefficient and transformer turns ratio have been observed to match each other, decreasing with increasing pressure. A similar model has been employed to deduce the device sensitivity to drive voltage. The coupling coefficient and transformer turns ratio sensitivities to drive voltage have also been found to match each other. In this case, these parameters increase with drive level.

Given the figure of merit values observed for the tested devices, it is not surprising that the bender transducer has proved to be an important device type for self-powered acoustic system applications. In particular, the ability of the bender to fit within a defined diameter, and stack compactly, one on top of the other, generally provides for space-efficient packaging. Finally, given the relatively high operating efficiency of the bender under high drive, the power requirements of underwater systems can generally be kept to a minimum by their use.

ACKNOWLEDGMENTS

Thanks are expressed to the Directors of Ultra Electronics Ltd. for their permission to publish this work. Thanks are also expressed to DERA (Winfrith) for their sponsorship of various projects to build and test bender devices, and in particular the 240-mm bender discussed here. Special thanks are given to Dr. Geoff Williams of DERA for his general encouragement and support in related areas of study over a number of years. Tribute is also paid to the staff of the Seneca Lake and Wraysbury test facilities for their assistance in making measurements at the respective sites. Finally, sincere thanks are given to Andy Carter (Ultra) for his care and attention in constructing the majority of the benders reported in this paper.

[1] W. G. Cady, *Piezoelectricity* (Dover, New York, 1964), p. 181.
[2] C. P. Germano, ''Flexure Mode Piezoelectric Transducers,'' IEEE Trans. Audio Electroacoust. **AU-19**, 6–12 (1971).
[3] R. S. Woollett, ''Trends and Problems in Sonar Transducer Design,'' IEEE Trans. Ultrason. Eng. **AU-10**, 116–124 (1963).
[4] R. S. Woollett, ''Theory of the Piezoelectric Flexural Disk Transducer with Application to Underwater Sound,'' USL Research Report No. 490 (1960).
[5] R. S. Woollett, *The Flexural Bar Transducer* (NUSC, New London, CT, 1986).
[6] D. F. Jones and J. F. Lindberg, ''Recent Transduction Developments in Canada and the United States,'' Proceedings of the Institute of Acoustics, Vol. 17, Pt. 3, 1995, pp. 15–33.
[7] *Piezoelectric Ceramics for Sonar Transducers*, MIL-STD-1376 (SHIPS), Dec. 1970 (US Government Printing Office, Washington, DC).
[8] *Notes on Transducer Design*, A.U.W.E. Publication 20730, Jan. 1969 (Admiralty Underwater Weapons Establishment, Portland, UK). See Chap. 2.
[9] R. J. Urick, *Principles of Underwater Sound*, 2nd ed. (McGraw-Hill, New York, 1983). See the formula for ''Line of n elements of equal spacing d,'' p. 43.
[10] L. L. Beranek, *Acoustics* (American Institute of Physics, 1986). See Eq. (5.1), p. 118.

Estimation of broadband acoustic power due to rib forces on a reinforced panel under turbulent boundary layer-like pressure excitation. I. Derivations using string model

M. L. Rumerman
Naval Surface Warfare Center, Carderock Division, Signatures Directorate (Code 7204), 9500 MacArthur Boulevard, West Bethesda, Maryland 20817-5700

(Received 28 October 1999; revised 9 October 2000; accepted 13 October 2000)

This paper shows that, when the attachment forces on a rib-reinforced panel subjected to turbulent boundary layer (TBL) excitation can be considered to radiate independently, the rib-related acoustic power in a broad (e.g., one-third octave) frequency band can be estimated as the product of the average mean-squared force, the real part of the radiation admittance of an attachment force, and the number of ribs. Using a simple model of a string with point mass or spring attachments, an approach is developed for estimating the average mean-squared force in broad frequency bands. The results are in a form that can be applied to ribbed plates and shells. The following paper establishes the condition under which the ribs can be considered to radiate independently, and presents the results of validating calculations for steel plates in water. [DOI: 10.1121/1.1331111]

PACS numbers: 43.40.Hb, 43.40.Cw [CBB]

I. INTRODUCTION

Exterior portions of many structures of interest may be considered large flat or curved panels or shells reinforced by ribs that are parallel and nominally identical, and often equally spaced. When subjected to turbulent flow over its surface, the panel will vibrate due to the turbulent boundary layer (TBL) pressure acting upon it, and radiate into the ambient acoustic medium. Calculation of the acoustic power radiated requires specification of the frequency-wave number spectrum (or cross-correlation function) of the driving TBL pressure, a description of the panel properties, and a method for calculating the panel's response to the driving pressure. Analyses of the response of finite panels to TBL pressure excitation typically treat rectangular panels, and formulate the solution in terms of *in vacuo* modal series. References 1–3 discuss finite rectangular panels that are otherwise homogeneous. Davies[1] evaluates the radiated power by ignoring modal coupling and including only the contributions of modes resonant in selected frequency bands, effectively obtaining a smoothed broadband radiated power spectrum. Graham[2] also ignores modal coupling, but evaluates the resulting series for radiated power by summing all terms at small frequency intervals, obtaining a narrow-band radiated power spectrum. Mkhitarov[3] presents a full modal formulation, but then appears to also neglect modal coupling in computing radiated pressure directivities at particular frequencies and a narrow-band spectrum for pressure radiated into a particular direction. In Ref. 4, Mkhitarov presents an analysis of a finite rectangular panel reinforced by several parallel ribs. The solution is developed in terms of modal series for each bay, and the imposition of continuity requirements at adjacent bays results in an infinite set of simultaneous algebraic equations for the coefficients. The formulation is very complicated, but produces narrow-band radiation spectra. Howe and Shah,[5] using a different approach, first consider the TBL excitation of two semi-infinite plates joined by a simply supported or clamped boundary condition, which represents an infinite impedance rib. They then extend the method to a periodic array of finite plate sections joined by clamped boundary conditions. Explicit but complicated expressions are given for the radiated pressure frequency spectral density. Shah and Howe[6] also present an alternative approach for the infinite plate with one clamped or simply supported discontinuity.

The methods of Refs. 4 and 5 spend a large portion of the computational effort in producing narrow-band results that may be given a significance not warranted by the accuracy of the descriptions of the narrow-band response characteristics of the structure. Estimates of radiated power in broad frequency bands may be more appropriate when no ''point'' frequency is of particular interest. This paper presents an approach for directly estimating the broadband radiated power spectrum (in contrast to integrating narrow-band results), when the rib forces on a reinforced panel driven by TBL pressures can be considered to radiate independently. Subject to this condition, a method is given for calculating the average (over all the ribs) of the mean-squared force that each rib exerts on the otherwise uniform panel within a frequency band. The power radiated is then found as the product of this average mean-squared force, the real part of the radiation admittance of a rib force, and the number of ribs. A following paper[7] demonstrates the conditions under which a set of ribs can be considered to radiate independently under TBL excitation.

This approach is intended to be applicable to ribbed plates or shells in water when the dominant response is in bending. In order to derive explicit results and keep expressions simple, it will be discussed in this paper in the context of the one-dimensional problem of a string, with mass or spring attachments, in a light acoustic medium. Although the derivations may be complicated, the expressions that emerge,

and the relationships among responses, will be shown to have simple physical interpretations that can be extended to plates and shells. Reference 7 presents validating calculations for a ribbed steel plate in water.

The paper is organized as follows: In Sec. II, a generic TBL pressure spectrum is applied to a lightly damped string without attachments, and the string's mean-squared velocity is calculated in broad frequency bands. The intent is to show how various calculations are performed, generate results needed in Sec. III, and provide a basis for viewing the effects of an attachment. In Sec. III, a centrally located attachment is added and the mean-squared velocity is recalculated along with the mean-squared force acting between the string and attachment. In Sec. IV, expressions are made specific for discrete mass or spring attachments (which are stand-ins for a beam in the case of a ribbed plate), the relationships among the various responses are discussed, and terms in the calculated expressions are given physical interpretations. In Sec. V, the results are generalized to a noncentrally located attachment. Section VI presents results of sample calculations that validate the approximate forms developed for the single attachment. In Sec. VII, the results for one noncentral attachment are used to formulate the approximation for a string with multiple attachments, and the results of validating calculations are presented. Section VIII summarizes the results and shows how they can be obtained from a spatially averaged equivalent of the inhomogeneous string. This interpretation provides the guidance for applying the results to ribbed plates and shells. Section IX presents conclusions.

II. STRING WITHOUT ATTACHMENT

A. Description of structure and excitation

The ''structure'' to be considered is a string of length L oriented along the x axis and fixed at its ends $x=\pm L/2$. The string has mass per unit length m, and is under a tension such that the speed of free waves in the lossless string *in vacuo* is c_0. At radian frequency ω, the wave number of the lossless string is $k_0=\omega/c_0$, and if the string is given a small spatially uniform loss factor η, the complex wave number k is found from the relation $k^2=k_0^2(1+i\eta)$, where $\exp(-i\omega t)$ time dependence is implicit. If the kinetic energy in the ambient acoustic medium is small compared to that of the string and radiation from the ends of the string is ignored, then, at radian frequency ω, the transverse velocity v_u of the uniform finite string due to a driving force distribution over its length may be given by Eq. (1). The term $v_0(x)$ is the response, for $|x|<L/2$, of a comparable string of infinite length to the same driving force

$$v_u(x)=v_0(x)-\frac{1}{2}[v_0(+L/2)+v_0(-L/2)]\frac{\cos(kx)}{\cos(kL/2)} - \frac{1}{2}[v_0(+L/2)-v_0(-L/2)]\frac{\sin(kx)}{\sin(kL/2)}. \quad (1)$$

Because the driving force distribution will represent a TBL pressure field, it is random in space and time; therefore, Eq. (1) represents a sample function in the ensemble of responses of a specific string to the various manifestations of the driving force distribution.

The driving force distribution will be assumed temporally stationary and spatially homogeneous, and will be specified by its wave number-frequency spectrum, $\phi(\alpha,\omega)$, taken to be in the form of Eq. (2) in which α is the wave number

$$\phi(\alpha,\omega)=A(\alpha,\omega)+B(\alpha,\omega;k_c). \quad (2)$$

In Eq. (2), the function B represents the convection peak at convection wave number k_c, and the function A represents the slowly varying lower portion of the spectrum at wave numbers less than k_c. Although the contribution of the convection peak may be important in aeroacoustic applications where the flow speed is so large that the convection wave number can be comparable to the structural wave number, it does not usually have a significant effect in hydroacoustic applications where the flow speed is much lower and the convection wave number much higher. The B term will, therefore, be neglected and the forcing spectrum represented by the A term alone which will now be assumed wavenumber white, so that $A(\alpha,\omega)=S(\omega)$.

B. Calculation of spatial mean-squared velocity

The spatial mean-squared velocity for one sample function of the uniform string, $\overline{v_u^2}$, is defined by Eq. (3)

$$\overline{v_u^2}=\frac{1}{L}\int_{-L/2}^{+L/2}|v_u^2(x)|dx, \quad (3)$$

in which $|v_u^2(x)|$ is formed by multiplying the right-hand side of Eq. (1) by its complex conjugate. This gives a preliminary deterministic result. The statistical mean-squared velocity, $\langle|v_u^2|\rangle$, is found by using the autocorrelation function $R_{v_0}(x_1-x_2)=\langle v_0(x_1)v_0^*(x_2)\rangle$ (where * represents the complex conjugate) to form the statistical average of the various velocity quadratics. When $\eta\ll 1$, the autocorrelation function is given by Eq. (4a). Its derivation is outlined in Appendix A, along with those of some other results in this section.

$$R_{v_0}(\varepsilon)=\frac{\pi k_0 S(\omega)}{\eta\omega^2 m^2}\left[\cos(k_0\varepsilon)+\frac{\eta}{2}\sin(k_0|\varepsilon|)\right] \times\exp(-\eta k_0|\varepsilon|/2). \quad (4a)$$

If the string is lightly damped, so that $\eta k_0 L/2\ll 1$, Eq. (4a) may be approximated by Eq. (4b)

$$R_{v_0}(\varepsilon)\approx\frac{\pi k_0 S(\omega)}{\eta\omega^2 m^2}\cos(k_0\varepsilon)+\frac{\pi k_0 S(\omega)}{2\omega^2 m^2} \times[\sin(k_0|\varepsilon|)-k_0|\varepsilon|\cos(k_0\varepsilon)]. \quad (4b)$$

The right-hand side of Eq. (1) is inserted into the integral of Eq. (3), and Eq. (4b) is used to evaluate $\langle v_0(x_1)v_0^*(x_2)\rangle$. The integration over x leads to the following integrals which, for $k_0L\gg 1$ and $\eta k_0 L\ll 1$, are evaluated as indicated:

$$\frac{1}{L}\int_{-L/2}^{+L/2}|\cos^2(kx)|dx\approx 1/2,$$

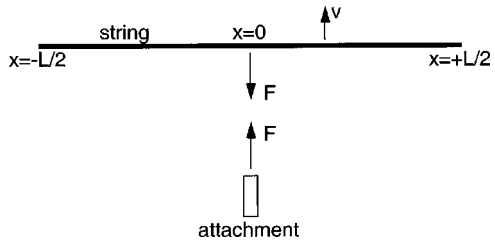


FIG. 1. String coordinates and sign conventions.

$$\frac{1}{L}\int_{-L/2}^{+L/2}|\sin^2(kx)|dx\approx 1/2,$$

$$\frac{1}{L}\int_{-L/2}^{+L/2}\sin(kx)\cos^*(kx)dx\rightarrow O(1/kL),$$

which will be neglected in comparison to 1/2.

$$\frac{1}{L}\int_{-L/2}^{+L/2}\sin^*(kx)\cos(kx)dx\rightarrow O(1/k_0L),$$

which will be neglected in comparison to 1/2. With these approximations to the integrals, the mean-squared velocity is given by Eq. (5)

$$\langle|v_u^2|\rangle\approx\frac{\pi k_0 S(\omega)}{8\omega^2 m^2}k_0L\left[\frac{1}{|\sin(kL/2)|^2}+\frac{1}{|\cos(kL/2)|^2}\right]\approx\frac{\pi k_0 S(\omega)}{2\omega^2 m^2}\frac{k_0L}{|\sin(kL)|^2}. \tag{5}$$

Equation (5) represents the point frequency spectrum of the mean-squared velocity. For broadband excitation, it will be dominated by peaks that occur at resonances of the string. The mean-squared velocity due to one resonance may be found by integrating Eq. (5) over the resonance bandwidth. This integral may be approximated in terms of its residue, and the result is $[S(\omega_n)\pi^2]/(\eta L\omega_n m^2)$, where ω_n is the nth resonance frequency. In a frequency band $\Delta\omega$, there will be approximately $L/\pi c_0\Delta\omega$ resonances; if this number is at least 2, the total mean-squared velocity in $\Delta\omega$ is approximated by Eq. (6), in which $\bar{\omega}$ is a representative resonance frequency within the band, and $S(\omega)$ is assumed to be slowly varying

$$\langle|v^2|\rangle_{\Delta\omega}\approx\frac{\pi S(\bar{\omega})}{\eta\bar{\omega}c_0m^2}\Delta\omega. \tag{6}$$

This result is identical to the mean-squared velocity of the infinite string in $\Delta\omega$.

III. STRING WITH CENTRALLY LOCATED ATTACHMENT

A. Modification to response of string

Consider an attachment to the string at $x=0$, which can exert a normal force on the string and in turn respond to the string's reaction force on it. Using the sign convention shown in Fig. 1, the mutual force developed between the string and the attachment is given by Eq. (7), in which $v_u(0)$ is the velocity of the uniform string (without attachment) at $x=0$, $\bar{Y}_a(\omega)=-iY_a(\omega)$ is the attachment's input admittance relating the force and velocity at its connection to the string, and $\bar{Y}_i(\omega)=-iY_i(\omega)+Y_r(\omega)$ is the input admittance of the string for a force applied at $x=0$

$$F=\frac{v_u(0)}{Y_i(\omega)+Y_a(\omega)}. \tag{7}$$

In writing the input admittance, radiation due to the action of the force on the string is included through the radiation admittance, $Y_r(\omega)$, which is defined so that $|F^2|\mathrm{Re}(Y_r)$ is the radiated power. Acoustic added mass and radiation from the ends will again be ignored. The ''structural'' part of the admittance is $-iY_i$. The effect of the acoustic field due to the attachment force will be assumed confined to a small neighborhood of the drive point, so that its influence on the string's velocity at $x=\pm L/2$ is negligible. With these assumptions, the input admittance is approximated by Eq. (8), in which the first term on the right-hand side is the *in vacuo* admittance for the string fixed at its ends. The second term, $\bar{Y}_r(0;\omega)$, is the acoustic component of the input admittance, where $F\bar{Y}_r(x;\omega)$ is the acoustic component of the string's velocity. For the purpose of this simplified model, $\bar{Y}_r(0;\omega)$ will be assumed real and positive

$$\bar{Y}_i(\omega)=-\frac{i}{2mc}\frac{\sin(kL/2)}{\cos(kL/2)}+\bar{Y}_r(0;\omega). \tag{8}$$

Combination of Eqs. (1), (7), and (8), and rearrangement of terms produces Eq. (9) for the force F

$$F=\frac{iv_0(0)-(i/2)[v_0(+L/2)+v_0(-L/2)]/\cos(kL/2)}{Y_i(\omega)+Y_a(\omega)+iY_r(\omega)}, \tag{9}$$

in which

$$Y_i(\omega)=\frac{1}{2mc}\frac{\sin(kL/2)}{\cos(kL/2)} \tag{10a}$$

and

$$Y_r(\omega)=\bar{Y}_r(0;\omega). \tag{10b}$$

The contribution of the force to the velocity of the string is given by Eq. (11)

$$v_F(x)=\frac{iF}{2mc}\frac{\sin[k(L/2-|x|)]}{\cos(kL/2)}+F\bar{Y}_r(x;\omega). \tag{11}$$

B. Modification of spatial mean-squared velocity

The total velocity of the string with the centrally located attachment is found by adding the right-hand side of Eq. (11) to the right-hand side of Eq. (1), and is given by Eq. (12)

$$v(x)=v_u(x)+v_F(x). \tag{12}$$

The spatial mean-squared velocity of a sample function is given by Eq. (13)

$$\overline{v^2}=\frac{1}{L}\int_{-L/2}^{+L/2}|v_u^2(x)|dx+\frac{1}{L}\int_{-L/2}^{+L/2}|v_F^2(x)|dx+\frac{1}{L}\int_{-L/2}^{+L/2}[v_u(x)v_F^*(x)+v_u^*(x)v_F(x)]dx. \tag{13}$$

The statistical average of the first term on the right-hand side of Eq. (13) has already been evaluated, and the result for a frequency band $\Delta\omega$ is given by Eq. (6). The statistical average of the other two terms is evaluated by using the velocity autocorrelation function given by Eq. (4b). The leading terms are given by expressions (14), in which it has been assumed that the contribution of the acoustic component of velocity is localized and negligible

$$\frac{\pi S(\omega)L}{8m^2c_0^2}\,\psi, \tag{14a}$$

$$\Psi = \frac{1}{|\cos(kL/2)|^2}\left|\frac{1}{2mc\cos(kL/2)}\right|^2\left|\frac{1}{Y_i(\omega)+Y_a(\omega)+iY_r(\omega)}\right|^2 - \frac{\sin(k_0L/2)}{|\cos(kL/2)|^2}\left[\frac{1}{2mc\cos(kL/2)}\right] \times\left[\frac{1}{Y_i(\omega)+Y_a(\omega)+iY_r(\omega)}\right] - \frac{\sin(k_0L/2)}{|\cos(kL/2)|^2}\left[\frac{1}{2mc\cos(kL/2)}\right]^*\left[\frac{1}{Y_i(\omega)+Y_a(\omega)+iY_r(\omega)}\right]^*. \tag{14b}$$

Expressions (14) represent part of the point frequency spectrum of the mean-squared velocity of the string with the centrally located attachment. For broadband excitation, it will be dominated by peaks which correspond to resonances of even modes of the string without the attachment, which are identified by the condition $\cos(k_nL/2)=0$, and by peaks which correspond to resonances of the new even modes of the string with the attachment at $x=0$, identified by the condition $Y_i(\omega)+Y_a(\omega)+iY_r(\omega)=0$. The contribution of a peak to the mean-squared velocity is obtained by integrating Eq. (14) across its bandwidth. It is found that the contribution of the original even modes is canceled, a result which is expected because they no longer represent resonances of the string. The contribution of a new even mode having resonance frequency ω_m is found to be approximated by Eq. (15a), in which $Y'(\omega)=dY/d\omega$, $Y'_{iR}=\mathrm{Re}(Y'_i)$, and $Y'_{aR}=\mathrm{Re}(Y'_a)$. It has been assumed that $\eta k_0L\ll 1$, $|Y'_r(\omega_m)/Y'_i(\omega_m)|\ll 1$, and that the attachment is lightly damped.

$$\frac{\pi^2S(\omega_m)L}{4m^2c_0^2}\left|\frac{1}{2mc\cos^2(k_mL/2)}\right|^2\cdot\left[\frac{1}{Y'_{iR}(\omega_m)+Y'_{aR}(\omega_m)}\right] \cdot\left[\frac{1}{2\omega_m\,\mathrm{Im}[Y'_i(\omega_m)]+2\omega_m|\mathrm{Im}[Y'_a(\omega_m)]|+2Y_r(\omega_m)}\right]. \tag{15a}$$

It follows from the assumptions that $2\,\mathrm{Im}[Y'_i(\omega_m)] \approx \eta Y'_{iR}(\omega_m)$. From Eq. (10a), $\cos^2(k_mL/2)Y'_i(\omega_m) = mL/(2mc)^2 \approx \cos^2(k_mL/2)Y'_{iR}(\omega_m) \approx mL/(2mc_0)^2$; therefore, Eq. (15a) may be rewritten as Eq. (15b)

$$\frac{\pi^2S(\omega_m)}{Lm^2}\left[\frac{1}{1+Y'_{aR}(\omega_m)/Y'_{iR}(\omega_m)}\right]\cdot\left[\frac{1}{\eta\omega_m+2\omega_m|\mathrm{Im}[Y'_a(\omega_m)]|/Y'_{iR}(\omega_m)+2Y_r(\omega_m)/Y'_{iR}(\omega_m)}\right]. \tag{15b}$$

In a sufficiently wide frequency band $\Delta\omega$, there will be approximately $[(L/2\pi c_0)+N_a(\omega)]\Delta\omega$ such resonances, where N_a is the modal density of the attachment. Upon multiplying Eq. (15b) by this number, and adding the result to the mean-squared velocity due to the unaltered odd modes of the string, Eq. (16) is obtained for the statistical average of the spatial mean-squared velocity of the string with attachment in frequency band $\Delta\omega$.

$$\langle|v^2|\rangle_{\Delta\omega} \approx \frac{\pi S(\bar\omega)}{2\eta\bar\omega c_0m^2}\Delta\omega + \frac{\pi S(\bar\omega)}{2\bar\omega c_0m^2}\left[\frac{1+2\pi c_0N_a(\bar\omega)/L}{1+Y'_{aR}(\bar\omega)/Y'_{iR}(\bar\omega)}\right]\cdot\left[\frac{1}{\eta+2|\mathrm{Im}[Y'_a(\bar\omega)]|/Y'_{iR}(\bar\omega)+2Y_r(\bar\omega)/[\bar\omega Y'_{iR}(\bar\omega)]}\right]\Delta\omega. \tag{16}$$

C. Calculation of mean-squared attachment force and velocity

The attachment force is given by Eq. (9), and, after applying the velocity autocorrelation function of Eq. (4b), the dominant term of the mean-squared value is found to be given by Eq. (17)

$$\langle|F^2|\rangle \approx \frac{\pi S(\omega)L}{4m^2c_0^2}\,\frac{1}{|\cos(kL/2)|^2}\left|\frac{1}{Y_i(\omega)+Y_a(\omega)+iY_r(\omega)}\right|^2. \tag{17}$$

This is a point frequency spectrum; to get the broadband spectrum, this must be integrated across the resonance bandwidths. The only resonance contributions occur at $Y_i(\omega_m)+Y_a(\omega_m)+iY_r(\omega_m)=0$, and the result of integrating over a single resonance bandwidth is given by Eq. (18), in which the previous approximations have been applied

$$\frac{2\pi^2 S(\omega_m)}{m\omega_m}\frac{1}{Y'_{iR}(\omega_m)}\left[\frac{1}{1+Y'_{aR}(\omega_m)/Y'_{iR}(\omega_m)}\right]\cdot\left[\frac{1}{\eta+2|\mathrm{Im}[Y'_a(\omega_m)]|/Y'_{iR}(\omega_m)+2Y_r(\omega_m)/[\omega_m Y'_{iR}\omega_m)}\right]. \tag{18}$$

Upon multiplying by the modal density, Eq. (19) results for the mean-squared force within a bandwidth $\Delta\omega$

$$\langle|F^2|\rangle_{\Delta\omega}\approx\frac{\pi S(\bar{\omega})L}{\bar{\omega}mc_0}\frac{1}{Y'_{iR}(\bar{\omega})}\left[\frac{1+2\pi c_0 N_a(\bar{\omega})/L}{1+Y'_{aR}(\bar{\omega})/Y'_{iR}(\bar{\omega})}\right]\cdot\left[\frac{1}{\eta+2|\mathrm{Im}[Y'_a(\bar{\omega})]|/Y'_{iR}(\bar{\omega})+2Y_r(\bar{\omega})/[\bar{\omega}Y'_{iR}\bar{\omega})}\right]\Delta\omega. \tag{19}$$

The mean-squared velocity at the attachment is calculated in a similar manner and found to be given by Eq. (20)

$$\langle|v_a^2|\rangle_{\Delta\omega}\approx|Y_a(\bar{\omega})|^2\langle|F^2|\rangle_{\Delta\omega}. \tag{20}$$

IV. INTERPRETATION OF RESULTS

Because the attachment is centrally located on the string, it does not change the character and responses of the odd modes; therefore, attention is restricted to the altered even modes of the string. Consideration is given to the responses in a broad frequency band $\Delta\omega$.

A. Properties of the attachment

The attachment to the string will now be assumed to be either a concentrated lumped mass or a spring. For a mass, $Y_a(\omega)=(i+\eta_a)/\omega\tau$ or $Y_a(\omega)=(-1+i\eta_a)/\omega\tau$, where τ is the size of the mass and η_a is a loss factor associated with its velocity. For $\eta_a\ll 1$, its total (kinetic) energy when driven by a force F is $E_a=(\tau/2)|F^2|/(\tau^2\omega^2)=|F^2|/2\tau\omega^2$. For a spring, $Y_a(\omega)=[(-i+\eta_a)\omega]/K$ and $Y_a(\omega)=[(1+i\eta_a)\omega]/K$, where K is the spring constant and η_a is a loss factor associated with the spring's strain rate. For $\eta_a\ll 1$, its total (potential) energy when driven by a force F is $E_a=(K/2\omega^2)(\omega^2|F^2|/K^2)=|F^2|/2K$. It follows that $|F^2|Y'_{aR}(\omega)=2E_a$ for both the mass and the spring. This result also applies to the string, namely that $Y'_{iR}(\omega)$ is equal to twice the total (kinetic+potential) energy, E_s, that the force F puts into the string; at a system resonance, this amounts to all the energy in the string. Because of these relationships, the ratio $2|\mathrm{Im}[Y'_a(\bar{\omega})]|/Y'_{iR}(\bar{\omega})$ appearing in Eqs. (16) and (19) is equal to $2\eta_a Y'_{aR}(\bar{\omega})/Y'_{iR}(\bar{\omega})$, in which the ratio $Y'_{aR}(\bar{\omega})/Y'_{iR}(\bar{\omega})$ is approximately equal to the ratio of the energy in the mass or spring attachment to the energy in the string in bandwidth $\Delta\omega$. (It is an approximation because there are several resonances within the bandwidth, each with its own ratio; the equality is exact for any single resonance.)

B. Relationships between mean-squared force and velocity

Comparison of Eqs. (16) and (19) shows that the relationship between the mean-squared velocity and the mean-squared attachment force for the even modes is given by Eq. (21)

$$\langle|F^2|\rangle_{\Delta\omega}\approx\frac{2mL\langle|v^2|\rangle_{\Delta\omega}}{Y'_{iR}(\bar{\omega})} \tag{21}$$

The string's kinetic energy is $mL/2\langle|v^2|\rangle_{\Delta\omega}$, which is half its total energy; therefore, $\langle|F^2|\rangle_{\Delta\omega}Y'_{iR}(\bar{\omega})$ is equal to twice the total energy in the string, as expected.

C. Power balance

There is no additional modal density associated with a mass or spring attachment, so that $N_a=0$. From Eq. (16), and upon applying the relationships just derived, the mean-squared velocity due to the even modes of the string with mass or spring attachment may be given by Eq. (22)

$$\langle|v^2|\rangle_{\Delta\omega}\approx\frac{\pi S(\bar{\omega})}{2\bar{\omega}c_0 m^2}\left[\frac{1}{1+Y'_{aR}(\bar{\omega})/Y'_{iR}(\bar{\omega})}\right]\cdot\left[\frac{1}{\eta+2\eta_a E_a/E_s+2Y_r(\bar{\omega})/[\bar{\omega}Y'_{iR}(\bar{\omega})]}\right]\Delta\omega. \tag{22a}$$

After rearrangement and multiplication of both sides by mL, Eq. (22b) results

$$\eta mL\langle|v^2|\rangle_{\Delta\omega}+2\eta_a E_a\frac{mL\langle|v^2|\rangle_{\Delta\omega}}{E_s}+Y_r(\bar{\omega})\frac{2mL\langle|v^2|\rangle_{\Delta\omega}}{\bar{\omega}Y'_{iR}(\bar{\omega})}\approx\frac{\pi S(\bar{\omega})}{2\bar{\omega}c_0 m^2}\left[\frac{mL}{1+Y'_{aR}(\bar{\omega})/Y'_{iR}(\bar{\omega})}\right]. \tag{22b}$$

Because $E_s=mL\langle|v^2|\rangle_{\Delta\omega}$ and the force F is given by Eq. (21), Eq. (22b) may be written as Eq. (23)

$$\bar{\omega}\eta E_s+2\bar{\omega}\eta_a E_a+Y_r\langle|F^2|\rangle_{\Delta\omega}\approx\frac{\pi LS(\bar{\omega})}{2mc_0}\left[\frac{1}{1+Y'_{aR}(\bar{\omega})/Y'_{iR}(\bar{\omega})}\right]. \tag{23}$$

The three terms on the left-hand side of Eq. (23), respectively, represent the power dissipated in the string, the power dissipated in the mass or spring attachment, and the power radiated due to the attachment force. The right-hand side should, therefore, equal the power input to the modified even modes, which can be verified by an independent calculation beginning with the integral

$$\mathrm{Real}\int_{-L/2}^{+L/2}v(x)f^*(x)dx,$$

where $f(x)$ is a sample function of the TBL driving force per unit length acting on the string, forming the statistical average and integrating over the resonance bandwidths.

D. Simplified expression for the mean-squared force

When the radiation, structural, and attachment damping are each small, the relationship defining a resonance becomes $2mc_0\sin(\omega L/2c_0)+Y_{aR}(\omega)\cos(\omega L/2c_0)=0$, resulting in $(2mc_0)^2\cos^2(k_mL/2)=Y_\infty^2+Y_{aR}^2(\omega_m)$, where $Y_\infty=1/(2mc_0)$ is the input admittance, *in vacuo*, of the equivalent uniform infinite string, ω_m is a resonance frequency, and $k_m=\omega_m/c_0$. It follows that $Y'_{iR}(\omega_m)=mL[Y_\infty^2+Y_{aR}^2(\omega_m)]$, and the mean-squared force in $\Delta\omega$ may be given by Eq. (24), in which $Y_0^2(\overline{\omega})=Y_\infty^2+Y_{aR}^2(\overline{\omega})=|Y_\infty-iY_{aR}|^2$, where Y_0 is the magnitude of the *in vacuo* input admittance of the loaded infinite string at the attachment point (neglecting η_a)

$$\langle|F^2|\rangle_{\Delta\omega}\approx\frac{\pi S(\overline{\omega})}{\overline{\omega}m^2c_0}\frac{1}{Y_0^2(\overline{\omega})}\left[\frac{1+2\pi c_0N_a(\overline{\omega})/L}{1+Y'_{aR}(\overline{\omega})/[mLY_0^2(\overline{\omega})]}\right]\cdot\left[\frac{1}{\eta+2\eta_aY'_{aR}(\overline{\omega})/[mLY_0^2(\overline{\omega})]+2Y_r(\overline{\omega})/[\overline{\omega}mLY_0^2(\overline{\omega})]}\right]\Delta\omega. \tag{24a}$$

It is shown in Appendix B that, for a mass or spring attachment, $[Y'_{aR}(\omega)]/[mLY_0^2(\omega)]\ll 1$, when $k_0L\gg 1$. Therefore, with $N_a=0$, Eq. (24a) can be further approximated by Eq. (24b)

$$\langle|F^2|\rangle_{\Delta\omega}\approx\frac{\pi S(\overline{\omega})}{\overline{\omega}m^2c_0}\frac{1}{Y_0^2(\overline{\omega})}\frac{\Delta\omega}{\eta+2\eta_aY'_{aR}(\overline{\omega})/[mLY_0^2(\overline{\omega})]+2Y_r(\overline{\omega})/[\overline{\omega}mLY_0^2(\overline{\omega})]}. \tag{24b}$$

E. Effects of significant structural damping

The derivations to this point have assumed ''light'' structural damping in the string, so that $\eta k_0L\ll 1$. If the string is ''significantly'' damped so that $\eta k_0L/2\gg 1$, then $\exp(-\eta k_0L/4)\ll 1$, and the velocity autocorrelation in the form of Eq. (4a) must be used. Additionally, $|\cos(kL/2)|\gg 1$, so that $\langle|v_u(0)|^2\rangle$, where $v_u(0)$ is the velocity at $x=0$ of the string without the attachment, is approximated as $R_{v_0}(0)$; waves generated at the ends of the string are damped out upon arrival at $x=0$. In this case, the point spectrum of the mean-squared attachment force, as calculated from Eq. (7), is given by Eq. (25a)

$$\langle|F^2|\rangle\approx\frac{\pi S(\omega)}{\eta\omega m^2c_0}\frac{1}{|\overline{Y_i}(\omega)+\overline{Y_a}(\omega)|^2}. \tag{25a}$$

From Eq. (8), with $\eta k_0L\gg 1$ but $\eta\ll 1$, $\overline{Y_i}(\omega)\approx Y_\infty+Y_r(\omega)$, and Eq. (25a) may be rewritten as Eq. (25b)

$$\langle|F^2|\rangle\approx\frac{\pi S(\omega)}{\eta\omega m^2c_0}\frac{1}{|Y_\infty+Y_r(\omega)-iY_a(\omega)|^2}. \tag{25b}$$

For a mass or spring attachment, $Y_a(\omega)$ is a smooth function of frequency, and the mean-squared force in a band $\Delta\omega$ is approximated by Eq. (26)

$$\langle|F^2|\rangle_{\Delta\omega}\approx\frac{\pi S(\overline{\omega})}{\eta\overline{\omega}m^2c_0}\frac{\Delta\omega}{|Y_\infty+Y_r(\overline{\omega})-iY_a(\overline{\omega})|^2}. \tag{26}$$

This is the result for a nonresonant system and is to be compared to Eq. (24b), which gives the mean-squared force for a resonant system with mass or spring attachment. Although its derivation has assumed a lightly damped system, the results of Appendix B show that, if ηk_0L is arbitrarily allowed to become large, Eq. (24b) will be approximated by Eq. (27)

$$\langle|F^2|\rangle_{\Delta\omega}\approx\frac{\pi S(\overline{\omega})}{\eta\overline{\omega}m^2c_0}\frac{1}{Y_0^2(\overline{\omega})}\Delta\omega. \tag{27}$$

Recalling that $Y_0^2(\overline{\omega})=Y_\infty^2+Y_{aR}^2(\overline{\omega})$, Eqs. (26) and (27) are seen to be similar but not identical. The effects of the radiation admittance Y_r and the attachment loss factor η_a are not included in the denominator of Eq. (27), but it has been assumed that $2mc_0Y_r\ll 1$ and $\eta_a\ll 1$, so these effects are small. Equation (24), which was derived for very light string damping, will yield a good approximation to the mean-squared attachment force, even if used for large string damping. Its accuracy for arbitrary intermediate string damping will be tested in the calculations described later.

V. SINGLE NON-CENTRAL ATTACHMENT

A. Mean-squared force

In Secs. III and IV, a centrally located attachment was considered, which only interacted with the even modes of the string. Because the interaction of multiple attachments with the string involves both odd and even modes, an expression for the mean-squared attachment force within a frequency band $\Delta\omega$, due to a noncentral attachment, will be developed. Following the procedures used above, the mean-squared force due to an attachment at $x=\xi$, for a particular sample function and frequency, is given by Eq. (28)

$$\langle|F^2(\omega)|\rangle=\frac{\langle|v_u(\xi)|^2\rangle}{|\overline{Y_i}(\omega)+\overline{Y_a}(\omega)|^2}. \tag{28}$$

The quantity $v_u(\xi)$ is the velocity of the finite string in the absence of the attachment, and is given by Eq. (1), with ξ replacing x. The statistical average of the point frequency spectrum of the mean-squared force is found by using the autocorrelation function of v_0, given by Eqs. (4), to form the averages of the various quadratic pairs which form $|v_u(\xi)|^2$. These operations yield the point frequency spectrum; the broadband mean-squared force spectrum, $\langle|F^2(\omega)|\rangle_{\Delta\omega}$, is estimated by integrating Eq. (28) over a resonance bandwidth, and multiplying by the modal density (including both even and odd modes) $L/(\pi c_0)$. The result is given by Eq. (29) for a mass or spring attachment, for which $N_a(\omega)=0$

$$\langle |F^2| \rangle_{\Delta\omega} \approx \frac{2\pi S(\bar{\omega})L}{\bar{\omega} m c_0} \left[\frac{1}{1 + Y'_{aR}(\bar{\omega})/Y'_{iR}(\bar{\omega})} \right] \cdot \left[\frac{1}{\eta Y'_{iR}(\bar{\omega}) + 2\eta_a Y'_{aR}(\bar{\omega}) + 2Y_r(\bar{\omega})/\bar{\omega}} \right] \Delta\omega. \quad (29)$$

For a mass attachment, $Y'_{aR}(\omega) = 1/(\tau\omega^2)$, and for a spring, $Y'_{aR}(\omega) = 1/K$.

For arbitrary force location ξ and light damping, $Y_{iR}(\omega)$ is given by Eq. (30a), and its derivative with respect to ω, $Y'_{iR}(\omega)$, is given by Eq. (30b)

$$Y_{iR}(\omega) = \frac{1}{2mc_0} \frac{\cos(2k_0\xi) - \cos(k_0 L)}{\sin(k_0 L)}, \quad (30a)$$

$$Y'_{iR}(\omega) = \frac{1}{2mc_0^2} \times \frac{L[1 - \cos(k_0 L)\cos(2k_0\xi)] - 2\xi \sin(k_0 L)\sin(2k_0\xi)}{\sin^2(k_0 L)}. \quad (30b)$$

For a lightly damped system, the resonance condition is given by $Y_{iR}(\omega_m) + Y_{aR}(\omega_m) = 0$, where ω_m is a resonance frequency; this leads to Eq. (31a) and then to Eq. (31b), in which $\Gamma(k_m L, \xi/L)$ is given by Eq. (31c)

$$\frac{1}{\sin^2(k_m L)} = \frac{2(mc_0)^2 \left[\left(\frac{1}{2mc_0} \right)^2 + Y_{aR}^2(\omega_m) \right]}{[1 - \cos(k_m L)\cos(2k_m\xi)] - \frac{1}{2}\sin^2(2k_m\xi)}, \quad (31a)$$

$$Y'_{iR}(\omega_m) = mL \left[\left(\frac{1}{2mc_0} \right)^2 + Y_{aR}^2(\omega_m) \right] \Gamma(k_m L, \xi/L), \quad (31b)$$

$$\Gamma(k_m L, \xi/L) = \frac{[1 - \cos(k_m L)\cos(2k_m\xi)] - \frac{2\xi}{L}\sin(k_m L)\sin(2k_m\xi)}{[1 - \cos(k_m L)\cos(2k_m\xi)] - \frac{1}{2}\sin^2(2k_m\xi)}. \quad (31c)$$

When $\xi = 0$, $\Gamma(k_m L, \xi/L)$ is identically equal to unity and $Y'_{iR}(\omega_m)$ is equal to that given earlier for the centrally located attachment. For ξ not equal to zero, no explicit evaluation is possible, but setting Γ equal to 2 leads to Eq. (32), in which $Y_0^2(\bar{\omega}) = [(1/2mc_0)^2 + Y_{aR}^2(\bar{\omega})]$

$$\langle |F^2| \rangle_{\Delta\omega} \approx \frac{\pi S(\bar{\omega})}{\bar{\omega} m^2 c_0} \frac{1}{Y_0^2(\bar{\omega})} \left[\frac{1}{1 + Y'_{aR}(\bar{\omega})/[2mLY_0^2(\bar{\omega})]} \right] \cdot \left[\frac{1}{\eta + \eta_a Y'_{aR}(\bar{\omega})/[mLY_0^2(\bar{\omega})] + Y_r(\bar{\omega})/[\bar{\omega} mLY_0^2(\bar{\omega})]} \right] \Delta\omega. \quad (32a)$$

For a mass or spring attachment, $Y'_{aR}(\bar{\omega})/[mLY_0^2(\bar{\omega})] \ll 1$; therefore, Eq. (32a) is further approximated by Eq. (32b)

$$\langle |F^2| \rangle_{\Delta\omega} \approx \frac{\pi S(\bar{\omega})}{\bar{\omega} m^2 c_0} \frac{1}{Y_0^2(\bar{\omega})} \left[\frac{1}{\eta + \eta_a Y'_{aR}(\bar{\omega})/[mLY_0^2(\bar{\omega})] + Y_r(\bar{\omega})/[\bar{\omega} mLY_0^2(\bar{\omega})]} \right] \Delta\omega. \quad (32b)$$

This form has the following two attractive properties:

(1) When there is no string damping ($\eta = 0$), the mean-squared force is twice that of the centrally located attachment. This is expected because both odd and even modes contribute to the noncentral force; that is, the effective modal density is doubled.
(2) When the string damping is more effective than the attachment damping or radiation loss, i.e., $\eta \gg \eta_a Y'_{aR}(\bar{\omega})/[mLY_0^2(\bar{\omega})]$ and $\eta \gg Y_r(\bar{\omega})/[mLY_0^2(\bar{\omega})]$, the mean-squared force is equal to that of the centrally located attachment. As shown in Appendix B, these inequalities result when $\eta k_0 L \gg 1$. Under this condition, reflections from the ends of the string are effectively dissipated before reaching the attachment which then appears to be centrally located on an infinite string.

B. Mean-squared velocity

The total velocity of the string for a particular sample function is given by Eqs. (33)

$$v(x) = v_u(x) - \frac{F}{mc_0} \frac{\sin[k(L/2 + \xi)]\sin[k(L/2 - x)]}{\sin(kL)}; \quad x > \xi, \quad (33a)$$

$$v(x) = v_u(x) - \frac{F}{mc_0} \frac{\sin[k(L/2 - \xi)]\sin[k(L/2 + x)]}{\sin(kL)}; \quad x < \xi. \quad (33b)$$

The statistically averaged spatial mean-squared velocity is found by forming $|v^2(x)|$, using Eqs. (4) to form the statistical average of the quadratic pairs, integrating over x, and retaining only leading terms in kL. The result is given by Eq. (34a)

$$\langle |v^2| \rangle_{\Delta\omega} \approx \frac{\pi S(\overline{\omega})}{\overline{\omega} c_0 m^2} \left[\frac{1}{1 + Y'_{aR}(\overline{\omega})/Y'_{iR}(\overline{\omega})} \right] \left[\frac{Y'_{iR}(\overline{\omega})}{\eta Y'_{iR}(\overline{\omega}) + 2\eta_a Y'_{aR}(\overline{\omega}) + 2Y_r(\overline{\omega})/\overline{\omega}} \right] \Delta\omega \tag{34a}$$

Again, using Eq. (31b) with $\Gamma = 2$ results in Eq. (34b) for a mass or spring attachment

$$\langle |v^2| \rangle_{\Delta\omega} \approx \frac{\pi S(\overline{\omega})}{\overline{\omega} c_0 m^2} \left[\frac{1}{\eta + \eta_a Y'_{aR}(\overline{\omega})/[mLY_0^2(\overline{\omega})] + Y_r(\overline{\omega})/[\overline{\omega} mLY_0^2(\overline{\omega})]} \right] \Delta\omega. \tag{34b}$$

Comparison of Eqs. (32b) and (34b) leads to Eq. (35)

$$\langle |F^2| \rangle_{\Delta\omega} \approx \frac{\langle |v^2| \rangle_{\Delta\omega}}{Y_0^2(\overline{\omega})}. \tag{35}$$

VI. SAMPLE CALCULATIONS FOR SINGLE ATTACHMENT

Equations (24), for the centrally placed attachment, and (32), for the noncentral attachment, will be applied to estimating the mean-squared attachment force in one-third-octave frequency bands due to a mass and a spring. The results will be compared to those obtained by direct integration over frequency of the mean-squared forces given by Eqs. (17) and (28), respectively. The parameters do not correspond to any particular system of interest, but were selected to test the robustness of the approximation. Calculations are made over a frequency range of 10 to 10 000 Hz. The string's mass per unit length is taken to be 100 kg/m and its length, L, and wave speed, c_0, are chosen to be 8 m and 100 m/s, respectively. These provide that $k_0 L = 5$ at 10 Hz, and that there are two even modes of the uniform string in the 100-Hz one-third-octave band so that the wideband approximation is valid for 100 Hz and above. The radiation admittance, Y_r, is chosen to be $0.1/(2mc_0)$ independent of frequency. Calculations are performed for string loss factors, η, of 0, 0.001, and 0.01. For these loss factors, the values of $\eta k_0 L$ at 1000 Hz are 0, 0.5, and 5, respectively, and are proportional at other frequencies. The wave number white spectral density of the driving force is taken to be $S(\omega) = 1\ \mathrm{N^2 s/m}$, independent of frequency. Results are given for a central attachment ($\xi = 0$ m) and a noncentral attachment ($\xi = 0.5$ m).

A. Mass attachment

The mass, τ, is chosen to be 3.2 kg. This yields $2mc_0 = \omega\tau$ at 1000 Hz, so that the attachment admittance is larger than $1/(2mc_0)$ at lower frequencies and is smaller at higher frequencies. Calculations will be presented for an associated loss factor, η_a, equal to 0.

B. Spring attachment

The spring constant, K, is chosen to be 12.5×10^7 N/m, so that $2mc_0 = K/\omega$ at 1000 Hz. The attachment admittance is smaller than $1/(2mc_0)$ at lower frequencies and is larger at higher frequencies. Calculations will be presented for an associated loss factor, η_a, equal to 0.1.

C. Comparison of results

Figures 2(a) and (b) present the mean-squared mass attachment force in one-third-octave frequency bands for $\xi = 0$ m and $\xi = 0.5$ m, respectively. In each figure, results are shown for string loss factors, η, of 0, 0.001, and 0.01. The open circles, squares, and triangles represent the approximation of Eqs. (24) or (32), and the solid lines represent the result of numerically integrating Eqs. (17) or (28). Figures 3(a) and (b) contain corresponding results for the spring attachment force. For all combinations of loss factors tested, the agreement between the approximation and the numerical

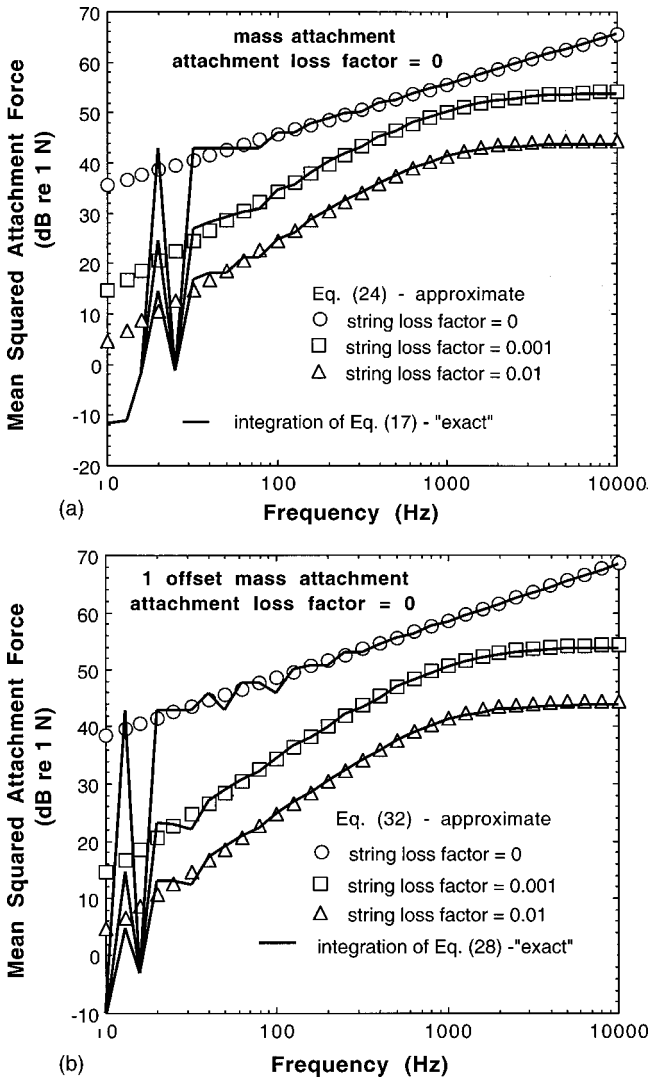


FIG. 2. (a) Mean-squared force for single mass attachment ($\xi = 0$ m); (b) mean squared force for single mass attachment ($\xi = 0.5$ m).

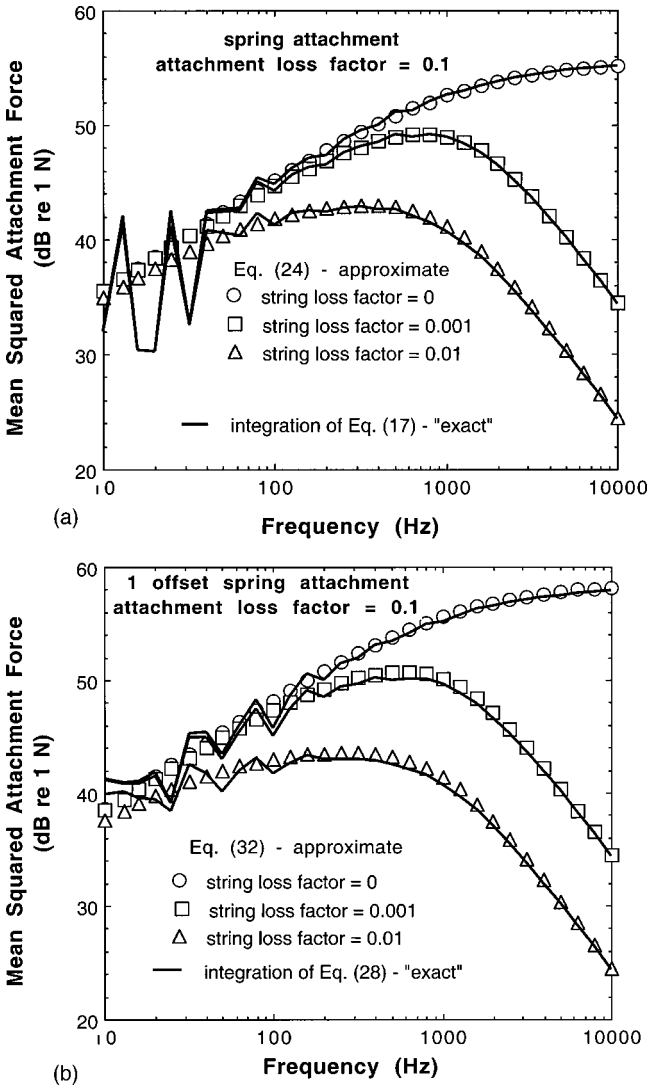


FIG. 3. (a) Mean-squared force for single spring attachment (ξ=0 m); (b) mean-squared force for single spring attachment (ξ=0.5 m).

integration is excellent for frequency bands centered at 100 Hz and above (where expected), and continues for a few lower bands as well.

VII. MULTIPLE ATTACHMENTS

A. Application of single attachment solution

Equation (34b), for a mass or spring attachment, may be rewritten as Eq. (36)

$$\eta\bar{\omega}mL\langle|v^2|\rangle_{\Delta\omega}+\eta_a\bar{\omega}\ Y'_{aR}(\bar{\omega})\frac{\langle|v^2|\rangle_{\Delta\omega}}{Y_0^2(\bar{\omega})}+Y_r\frac{\langle|v^2|\rangle_{\Delta\omega}}{Y_0^2(\bar{\omega})}=\frac{\pi LS(\bar{\omega})}{c_0m}. \tag{36}$$

Use of Eq. (35), and recognition that $mL\langle|v^2|\rangle_{\Delta\omega}=E_s$, where E_s is the energy in the string, and $\langle|F^2|\rangle_{\Delta\omega}Y'_{aR}(\bar{\omega})=2E_a$, where E_a is the energy in the attachment, leads to Eq. (37)

$$\eta\bar{\omega}E_s+2\,\eta_a\bar{\omega}E_a+\langle|F^2|\rangle_{\Delta\omega}Y_r=\frac{\pi LS(\bar{\omega})}{c_0m}. \tag{37}$$

The three terms on the left-hand side of Eq. (37) are, respectively, the power dissipated in the string, the power dissipated in the attachment, and the power radiated by the attachment force. These should sum to the power input to the system and it can be independently verified that the right-hand side of Eq. (37) represents that input power. The quantity $Y'_{aR}(\bar{\omega})/[2mLY_0^2(\bar{\omega})]$ has been neglected in comparison to unity. The input power is insensitive to the parameters of the mass or spring attachment.

Suppose that there are a number N mass or spring attachments. It can be expected that the input power is insensitive to N. The total power dissipated must again equal the input power, but power is now radiated and lost at N locations in addition to the dissipation over the length of the string. If $\langle|F^2|\rangle_{\Delta\omega}$ is now allowed to represent the average over all attachments of the mean-squared attachment force in $\Delta\omega$ for each individual attachment, the power balance is given by Eq. (38a) or by Eq. (38b)

$$\eta\bar{\omega}E_s+2\,\eta_aN\bar{\omega}E_a+N\langle|F^2|\rangle_{\Delta\omega}Y_r=\frac{\pi LS(\bar{\omega})}{c_0m}, \tag{38a}$$

$$\eta\bar{\omega}mL\langle|v^2|\rangle_{\Delta\omega}+\eta_aN\bar{\omega}Y'_{aR}(\bar{\omega})\frac{\langle|v^2|\rangle_{\Delta\omega}}{Y_0^2(\bar{\omega})}+NY_r\frac{\langle|v^2|\rangle_{\Delta\omega}}{Y_0^2(\bar{\omega})}=\frac{\pi LS(\bar{\omega})}{c_0m}. \tag{38b}$$

Finally, upon solving for $\langle|v^2|\rangle_{\Delta\omega}$, using Eq. (35), and letting $d=L/N$ be the *average* spacing between attachments, Eq. (39) results for the overall mean-squared force in the frequency band $\Delta\omega$

$$\langle|F^2|\rangle_{\Delta\omega}\approx\frac{\pi S(\bar{\omega})}{\bar{\omega}m^2c_0}\frac{1}{Y_0^2(\bar{\omega})}\frac{\Delta\omega}{\eta+\eta_aY'_{aR}(\bar{\omega})/[mdY_0^2(\bar{\omega})]+Y'_r(\bar{\omega})/[\bar{\omega}mdY_0^2(\bar{\omega})]}. \tag{39}$$

B. Exact solution

Equation (39) will be compared to the result obtained by numerically integrating the exact solution for a string of length L having N identical attachments of admittance $-iY_a(\omega)$ at positions $x=\xi_n$, $n=1,N$. Let $v_u(x)$ be a sample function of the velocity of the uniform string; then, in order to maintain continuity between the string and each attachment, the attachment forces $F_n(\omega)$ must satisfy Eq. (40), in

which $-iY_s(\xi_i,x_n;\omega)+Y_r(\omega)\delta_{in}$ is the strings response at x_n due to a force at ξ_i, and δ_{in} is the Kronecker delta function signifying that the acoustic component of the response is assumed localized to the drive point.

$$\sum_{i=1}^{N} F_i(\omega)\{Y_s(\xi_i,x_n;\omega)+[Y_a(\omega)+iY_r(\omega)]\delta_{in}\} = iv_u(x_n); \quad n=1,N. \tag{40}$$

These equations may be solved for the forces in the form of Eqs. (41), in which $Z_{ni}(\omega)$ is an impedance matrix

$$F_i(\omega)=\sum_{n=1}^{N} iv_u(x_n)Z_{ni}(\omega); \quad i=1,n. \tag{41}$$

It follows that

$$|F_i^2(\omega)|=\sum_{n=1}^{N} v_u(x_n)Z_{ni}(\omega)\sum_{k=1}^{N} v_u^*(x_k)Z_{ki}^*(\omega), \tag{42a}$$

and the statistical average of the mean-squared force of the ith attachment as a function of frequency is given by Eq. (42b), in which $R_{v_0}(\varepsilon)$ is given by Eqs. (4)

$$\langle|F_i^2(\omega)|\rangle=\sum_{n=1}^{N}\sum_{k=1}^{N} R_{v_0}(x_n-x_k)Z_{ni}(\omega)Z_{ki}^*(\omega). \tag{42b}$$

The right-hand side of Eq. (42b) may be integrated over frequency to obtain $\langle|F_i^2(\omega)|\rangle_{\Delta\omega}$, and the average over all attachments is given by Eq. (43), in which $\overline{\omega}$ is the band center frequency

$$\langle|F^2(\overline{\omega})|\rangle=\frac{1}{N}\sum_{i=1}^{N}\langle|F_i^2(\omega)|\rangle_{\Delta\omega}. \tag{43}$$

C. Calculations

Figure 4(a) presents a comparison of the overall mean-squared attachment force in one-third-octave frequency bands for a string having four mass attachments located at $x=-3.0$, -1.3, $+0.7$, $+2.8$. Results are shown for $\eta=0$, 0.001, and 0.01, with $\eta_a=0$ for each case. Figure 4(b) presents a similar comparison for four spring attachments at the same locations and for the same three string loss factors, but with $\eta_a=0.1$. In both cases, very good agreement is obtained down to frequencies below 100 Hz.

As a check on whether the approximation is applicable to uniformly spaced attachments, calculations were done for eight masses located at $x=\pm3.5$, ±2.5, ±1.5, and ±0.5, with $\eta=0$ and $\eta_a=0.1$. Results are given in Fig. 5(a) which again shows very good agreement between the approximation and exact solution down to well below 100 Hz. This is surprising because it is expected that the exact solution would have shown dropouts influenced by stop bands. The explanation for the absence of dropouts is that the particular choice of parameters results in a high density of stop-/passbands, and any one-third-octave band enclosing more than one stop band must also possess at least one passband that contains the resonance required for the approximation.

A different situation is illustrated by Fig. 5(b) for another set of parameters: $m=39$ g/cm, $L=400$ cm eight

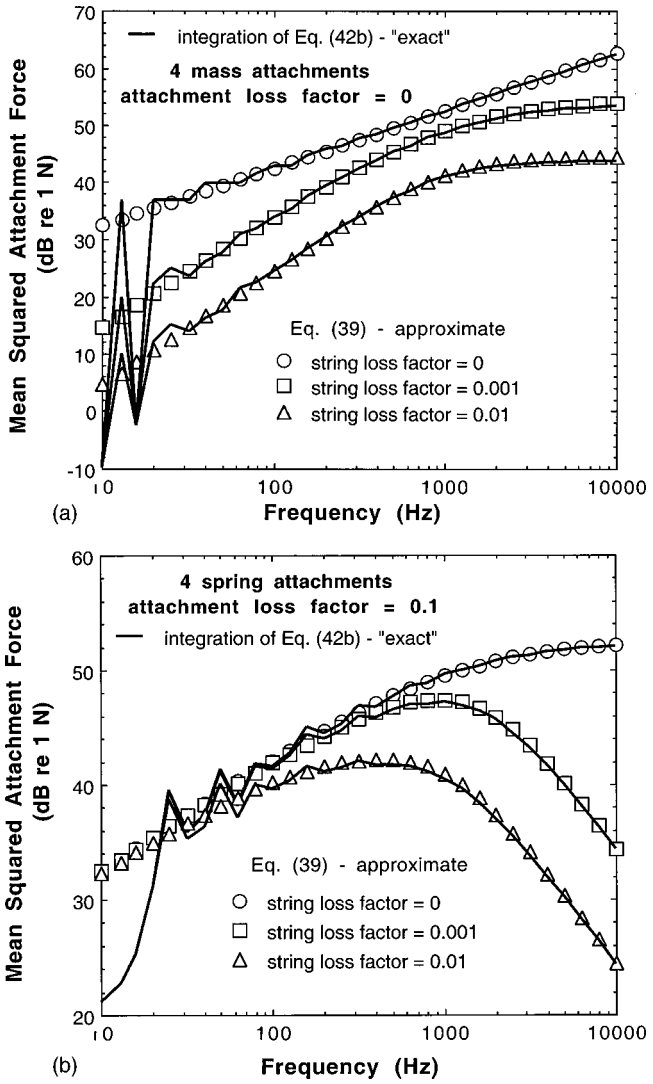


FIG. 4. (a) Mean-squared force for four mass attachments; (b) mean-squared force for four spring attachments.

masses of $\tau=585$ g located at $x=\pm175$ cm, ±125 cm, ±75 cm, ±25 cm. The wave speed, c_0, is chosen here to be frequency dependent so that the string mimics the bending characteristics of a 5-cm-thick steel beam: $c_0=340\sqrt{mf}$, where c_0 is in cm/s, m is in g/cm, and f is the frequency in Hz. The loss factors are again $\eta=0$ and $\eta_a=0.1$. For these parameters, some stop bands span an entire one-third-octave band, and dropouts result. At frequencies above about 280 Hz, the minima in the exact solution are due to stop bands, which cannot be identified by the approximate solution. Below about 150 Hz, the exact solution for the finite string departs from the approximate solution, not because of stop bands, but because some one-third-octave bands do not contain a resonance.

VIII. INTERPRETATION AND APPLICATION

The purpose of this section is to examine the explicit results for the string, and give the various combinations of parameters a robust interpretation. This will permit equivalent quantities and results for more complicated systems to be calculated without a detailed analysis. It will be shown

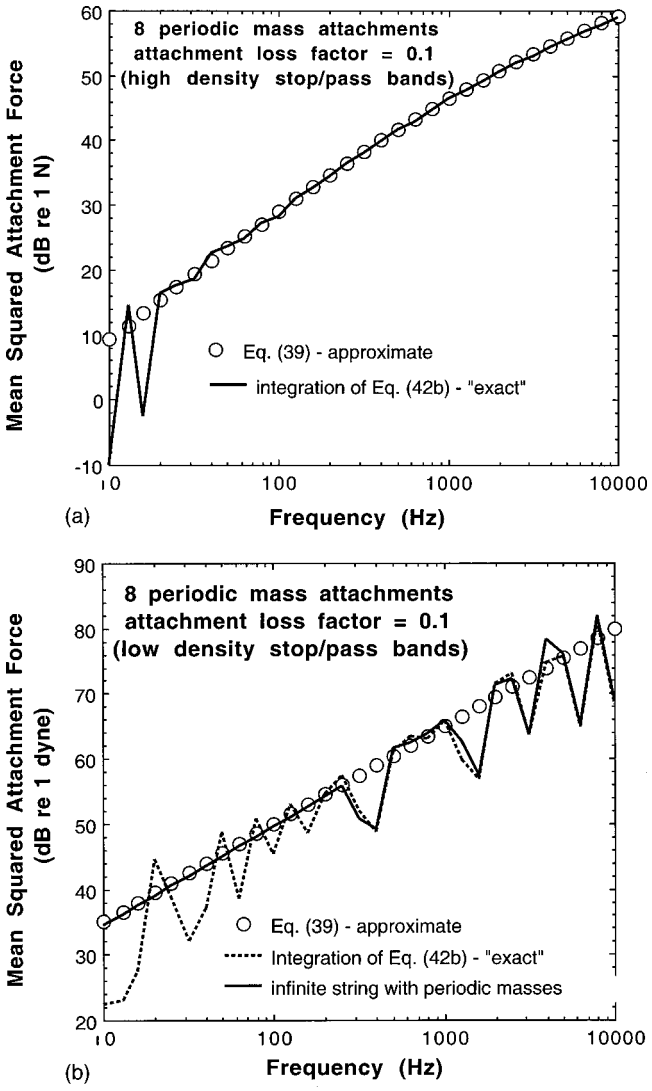


FIG. 5. (a) Mean-squared force for eight uniformly spaced mass attachments (high density stop-/passbands); (b) mean-squared force for eight uniformly spaced mass attachments (low density stop/pass bands).

that the mean-squared velocity of the string is equivalent to that of an infinite uniform string having spatially averaged dissipation properties related to the losses in the vibrating attachments, radiation by the attachment forces, and structural dissipation in the string itself. The mean-squared attachment force follows from enforcing continuity of velocity between the attachment and the string.

Equation (38b) may be rewritten as Eq. (44)

$$\eta\bar{\omega}m\langle|v^2|\rangle_{\Delta\omega}+\frac{\eta_a\bar{\omega}\bar{Y}'_{aR}(\bar{\omega})}{dY_0^2(\bar{\omega})}\langle|v^2|\rangle_{\Delta\omega}+\frac{Y_r}{dY_0^2(\bar{\omega})}\langle|v^2|\rangle_{\Delta\omega}=\frac{\pi S(\bar{\omega})}{c_0 m}. \tag{44}$$

The right-hand side is the input power per unit length. The three terms on the left-hand side are, respectively, the power dissipated per unit length by string damping, the *spatially, averaged* power dissipated per unit string length by the attachment damping, and the *spatially averaged* power radiated per unit string length by the attachment forces. The spatially averaged values are used to define uniform viscous damping coefficients, β_a and β_r, for the attachment dissipation and radiation, respectively.

$$\beta_a=\frac{\eta_a\bar{\omega}\bar{Y}'_{aR}(\bar{\omega})}{dY_0^2(\bar{\omega})}, \tag{45a}$$

$$\beta_r=\frac{Y_r}{dY_0^2(\bar{\omega})}. \tag{45b}$$

These coefficients can be used to form a differential equation for the response of an infinite string having smoothed dissipation properties

$$-T(1-i\eta)u(x)_{xx}-m\omega^2u(x)-i\omega(\beta_a+\beta_r)u(x)=f(x). \tag{46}$$

In Eq. (46), T is the string tension, $u(x)$ is the string displacement as a function of x, u_{xx} is its second derivative with respect to x, f is the driving force per unit length, and exp $(-i\omega t)$ time dependence is assumed and suppressed. If $f(x)=\bar{f}_0\exp(i\alpha x)$, the solution of Eq. (46) is $\bar{u}_0\exp(i\alpha x)$, where $\bar{u}_0$ is given by Eq. (47)

$$\bar{u}_0=\bar{f}_0[T(1-i\eta)\alpha^2-M\omega^2-i\omega(\beta_a+\beta_r)]^{-1}. \tag{47}$$

The mean-squared string velocity at frequency ω, due to a spatially random driving force represented by a slowly varying wave-number-white mean-squared spectral density $S(\omega)$, is given by Eq. (48)

$$\langle|v^2|\rangle=S(\omega)\omega^2\int_{-\infty}^{+\infty}|T(1-i\eta)\alpha^2-M\omega^2-i\omega(\beta_a+\beta_r)|^{-2}d\alpha. \tag{48}$$

The integral can be evaluated in terms of its residues, and, for small loss factors, the mean-squared velocity in a bandwidth $\Delta\omega$ is given by Eq. (49)

$$\langle v^2\rangle_{\Delta\omega}=\frac{\pi S(\bar{\omega})}{\bar{\omega}c_0m^2}\frac{1}{\eta+(\beta_a+\beta_r)/(\bar{\omega}m)}. \tag{49}$$

This result is identical to that of Eq. (48). If the mean-squared attachment force is calculated by assuming no interaction among attachments, then $\langle|F^2|\rangle_{\Delta\omega}=\langle|v^2|\rangle_{\Delta\omega}/|\bar{Y}_i(\bar{\omega})+\bar{Y}_a(\bar{\omega})|^2$. When loss factors are much smaller than unity, $|Y_i(\bar{\omega})+Y_a(\bar{\omega})|^2=Y_\infty^2(\bar{\omega})+Y_{aR}^2(\bar{\omega})=Y_0^2(\bar{\omega})$, and Eq. (35) is recovered.

This result can also be obtained by a heuristic argument. Assume that $\langle|v^2|\rangle_{\Delta\omega}$ is the resulting mean-squared velocity of the string with attachments. Again assuming each attachment force to be independent of the others, continuity requires that $\langle|F^2|\rangle\approx\langle|v^2|\rangle/Y_0^2$. The spatially averaged power radiated per unit length of string is $\langle|F^2|\rangle(Y_r/d)=\langle|v^2|\rangle\times(Y_r/d)/Y_0^2$. The power dissipated by mass attachments per unit length of string is given by $\langle|F^2|\rangle(\eta_a/\omega\tau d)=\langle|v^2|\rangle\times(\eta_a/\omega\tau d)/Y_0^2$. For spring attachments, the power dissipated per unit length is $\langle|F^2|\rangle(\omega\eta_a/Kd)=\langle|v^2|\rangle\times(\omega\eta_a/Kd)/Y_0^2$. Each of these can be written as $\langle|v^2|\rangle\omega\eta_aY'_{aR}(\omega)/[dY_0^2(\omega)]$. The power dissipated through structural string damping per unit length is $\eta\omega m\langle|v^2|\rangle$. When these are equated to the power input per unit length, Eq. (49) is reproduced.

IX. CONCLUSION

The objective of this and the following paper is to develop a formalism for estimating the broadband acoustic power radiated by a rib-reinforced panel driven by a spatially homogeneous and temporally stationary random forcing field representative of a TBL excitation. In order to simplify the derivations and resulting expressions, the structure considered for analysis in this paper was a finite length string having identical mass or spring attachments. It was shown that, if the acoustic parts of the attachments' interactions with the string can be considered equal and independent of one another, an equivalent uniform string can be defined. The uniform model can then be used to estimate the overall mean-squared attachment force, $\langle|F^2|\rangle_{\Delta\omega}$, in broad frequency bands, without accounting for interactions among attachments. The power radiated in $\Delta\omega$ is then given as the product $NY_r(\omega)\langle|F^2|\rangle_{\Delta\omega}$, where N is the number of attachments and Y_r is the power radiated per unit force. Exceptions occur in those frequency bands in which the response of the structure is dominated by a stop band. In these cases, the approximation produces an overestimate of the mean-squared force and the resulting rib-related radiated power. For typical steel or aluminum panels in water, the severity of stop bands may be mitigated by communication among panel bays via the water path.

ACKNOWLEDGMENT

This work was funded through the Office of Naval Research Hydroacoustics Program: L. Patrick Purtell, Program Manager.

APPENDIX A

1. Derivation of Eq. (4a)

The mean-squared spectral density of the velocity of the infinite string is related to that of the driving pressure by Eq. (A1)

$$\Phi_{v_0}(\alpha,\omega)=\frac{\Phi(\alpha,\omega)}{\left|-i\omega m\left(1-\frac{\alpha^2}{k_0^2(1+i\eta)}\right)\right|^2}. \tag{A1}$$

The autocorrelation function of this velocity, as a function of frequency, is found by integrating the spectral density over wave number. For a wave-number-white driving spectrum, and $\eta^2 \ll 1$, Eq. (A2) is obtained

$$R_{v_0}(\varepsilon)=S(\omega)\frac{k_0^4}{\omega^2 m^2}\times\int_{-\infty}^{+\infty}\frac{\exp(i\varepsilon\alpha)}{[\alpha^2-k_0^2(1+i\eta)][\alpha^2-k_0^2(1-i\eta)]}d\alpha. \tag{A2}$$

The integral may be evaluated by using contour integration. The integrand has four poles located at $\alpha=\pm k_0(1\pm i\eta/2)$. For $\varepsilon>0$, the contour is closed in the upper half plane; for $\varepsilon<0$, it is closed in the lower half plane. Application of the theory of residues leads to Eq. (4a).

2. Derivation of Eq. (5)

The square of the absolute value of the right-hand side of Eq. (1) is formed, which leads to expressions such as the three given below, in which the asterisk signifies the complex conjugate

$$v_0(x)v_0^*(x);$$

$$\tfrac{1}{4}[v_0(+L/2)+v_0(-L/2)][v_0^*(+L/2)+v_0^*(-L/2)]\times\frac{\cos(kx)}{\cos(kL/2)}\frac{\cos(k^*x)}{\cos(k^*L/2)};$$

$$-\frac{1}{2}v_0(x)[v_0^*(+L/2)+v_0^*(-L/2)]\frac{\cos(k^*x)}{\cos(k^*L/2)}-\frac{1}{2}v_0^*(x)[v_0(+L/2)+v_0(-L/2)]\frac{\cos(kx)}{\cos(kL/2)}.$$

The statistical average of the velocity products are formed and expressed in terms of the autocorrelation function, $R_{v_0}(\varepsilon)$. Sample evaluations are given by Eqs. (A3)

$$v_0(x)v_0^*(x)=R_{v_0}(x-x)=R_{v_0}(0), \tag{A3a}$$

$$v_0(+L/2)v_0^*(+L/2)=R_{v_0}(L/2-L/2)=R_{v_0}(0), \tag{A3b}$$

$$v_0(+L/2)v_0^*(-L/2)=R_{v_0}(L/2+L/2)=R_{v_0}(L), \tag{A3c}$$

$$v_0(x)v_0^*(+L/2)=R_{v_0}(x-L/2), \tag{A3d}$$

$$v_0^*(x)v_0(-L/2)=R_{v_0}(-L/2-x)=R_{v_0}(x+L/2). \tag{A3e}$$

When the form of the autocorrelation function given by Eq. (4b) is used to evaluate the right-hand side of Eqs. (A3), and for all such combinations, Eq. (5) is obtained.

3. Derivation of Eq. (6)

Equation (5) represents the point frequency spectral density of the mean-squared velocity. The broadband frequency spectrum is obtained by integrating the point spectral density across resonance bandwidths. Let $\delta\omega$ be a narrow frequency band containing a resonance frequency, ω_n, defined by $kL=(\omega_nL/c_0)(1+i(\eta/2))=n\pi$. The integration of the right-hand side of Eq. (5) over $\delta\omega$ can be approximated by Eq. (A4)

$$\frac{\pi L(\omega_n/c_0)^2S(\omega_n)}{2\omega_n^2m^2}\times\int_{\delta\omega}\frac{d\omega}{\sin\left[\frac{\omega L}{c_0}\left(1+i\frac{\eta}{2}\right)\right]\sin\left[\frac{\omega L}{c_0}\left(1-i\frac{\eta}{2}\right)\right]}. \tag{A4}$$

The integrand has poles defined by $(\omega L/c_0)(1-i(\eta/2))=n\pi$ and $(\omega L/c_0)(1+i(\eta/2))=n\pi$, respectively, corresponding to the first and second terms in its denominator. The value of the integral can be approximated by contour integration, closing the contour in the upper half plane to include the second pole. Application of the theory of residues to the integral results in Eq. (A5)

$$\frac{2\pi i}{\sin[n\pi(1+i\eta)]\dfrac{L}{c_0}\cos(n\pi)}=\frac{2\pi i}{\dfrac{L}{c_0}\cos^2(n\pi)\sin(\eta n\pi i)}$$

$$=\frac{c_0}{L}\frac{2\pi i}{\eta n\pi i}=\frac{2c_0}{\eta nL}. \tag{A5}$$

For $\eta \ll 1$, $\omega_n \approx n\pi c_0/L$. Upon combining the right-hand side of Eq. (A5) with the factor to the left of the integral in Eq. (A4), Eq. (A6) is obtained

$$\frac{\pi^2 S(\omega_n)}{\eta L \omega_n m^2}. \tag{A6}$$

Finally, multiplying Eq. (A6) by the modal density, $L/\pi c_0$, and the bandwidth, $\Delta\omega$, and letting $\bar{\omega}$ be a representative frequency in $\Delta\omega$, Eq. (6) is obtained.

APPENDIX B

For a mass,

$$\frac{Y'_{aR}(\omega)}{mLY_0^2(\omega)}=\frac{1/(\tau\omega^2)}{mL\left[\left(\dfrac{1}{2mc_0}\right)^2+\left(\dfrac{1}{\tau\omega}\right)^2\right]}$$

$$=\frac{2}{k_0L}\frac{(2mc_0/\tau\omega)}{[1+(2mc_0/\tau\omega)^2]}<\frac{1}{k_0L}.$$

For a spring,

$$\frac{Y'_{aR}(\omega)}{mLY_0^2(\omega)}=\frac{1/K}{mL\left[\left(\dfrac{1}{2mc_0}\right)^2+\left(\dfrac{\omega}{K}\right)^2\right]}$$

$$=\frac{2}{k_0L}\frac{(2\omega mc_0/K)}{[1+(2\omega mc_0/K)^2]}<\frac{1}{k_0L}.$$

Therefore, the left-hand side of each expression is much smaller than 1 when $k_0L \gg 1$. It follows that for both a mass and a spring,

$$\frac{\eta_a}{\eta}\frac{Y'_{aR}(\omega)}{mLY_0^2(\omega)}<\frac{\eta_a}{\eta k_0L}\ll 1 \quad \text{when} \quad \eta k_0L\gg 1 \quad \text{and} \quad \eta_a<1.$$

Finally, writing $Y_r=a/(2mc_0)$,

$$\frac{Y_r(\omega)}{\eta\omega mLY_0^2(\omega)}=\frac{a}{2\eta\omega m^2c_0L\left[\left(\dfrac{1}{2mc_0}\right)^2+Y_{aR}^2(\omega)\right]}$$

$$=\frac{2a}{\eta k_0L}\frac{1}{[1+(2mc_0Y_{aR})^2]}<\frac{2a}{\eta k_0L}.$$

Therefore, if $\eta k_0L \gg 1$ and $2a$ is unity or less, the left-hand side is much smaller than 1.

[1] H. G. Davies, ''Sound from turbulent-boundary-layer-excited panels,'' J. Acoust. Soc. Am. **49**, 878–889 (1971).

[2] W. R. Graham, ''Boundary layer induced noise in aircraft. I. The flat plate model,'' J. Sound Vib. **192**, 101–120 (1996).

[3] R. A. Mkhitarov, ''Sound radiation from a rectangular plate vibrating under the action of pressure fluctuations in a turbulent boundary layer,'' Sov. Phys. Acoust. **19**, 367–371 (1974).

[4] R. A. Mkhitarov, ''Sound radiation from a bounded thin inhomogeneous plate reinforced with *N* ribs and driven by boundary-layer pressure fluctuations,'' Sov. Phys. Acoust. **20**, 528–533 (1975).

[5] M. S. Howe and P. L. Shah, ''Influence of mean flow on boundary layer generated interior noise,'' J. Acoust. Soc. Am. **99**, 3401–3411 (1996).

[6] P. L. Shah and M. S. Howe, ''Sound generated by a vortex interacting with a rib-stiffened elastic plate,'' J. Sound Vib. **197**, 103–115 (1996).

[7] M. L. Rumerman, ''Estimation of broadband acoustic power due to rib forces on a reinforced panel under TBL-like pressure excitation. II. Applicability and validation,'' J. Acoust. Soc. Am. **109**, 576–582 (2000).

Estimation of broadband acoustic power due to rib forces on a reinforced panel under turbulent boundary layer-like pressure excitation. II. Applicability and validation

M. L. Rumerman
Naval Surface Warfare Center, Carderock Division, Signatures Directorate (Code 7204), 9500 MacArthur Boulevard, West Bethesda, Maryland 20817-5700

(Received 28 October 1999; revised 9 October 2000; accepted 13 October 2000)

The previous paper showed that, when the attachment forces on a rib-reinforced panel subjected to turbulent boundary layer excitation can be considered to radiate independently, the rib-related acoustic power in a broad (e.g., one-third octave) frequency band can be estimated as the product of the average mean squared force, the real part of the radiation admittance of an attachment force, and the number of ribs. This paper shows that the radiation condition is always approximated when the acoustic wavelength is less than twice the rib spacing of a periodically reinforced panel, and generally applies at lower frequencies where the acoustic wavelength is less than four times the rib spacing. The procedure is used to estimate the broadband acoustic power radiated per rib of an infinite periodically reinforced membrane and plate in water, and the results are shown to agree with those of ''exact'' calculations. [DOI: 10.1121/1.1331112]

PACS numbers: 43.40.Dx, 43.40.Rj [CBB]

I. INTRODUCTION

The random turbulent boundary layer (TBL) pressure field generated on a rib-reinforced panel due to flow over its surface causes the panel to vibrate and forces to be exerted between the panel and each rib that generate radiated sound power. If there are competing loss mechanisms, a mathematical model must be able to properly apportion power radiated with that lost to structural damping and other mechanisms in order to determine the panel response and the resulting radiated power. The calculation of a narrow-band radiated power spectrum requires a complete solution for the system because the phase relationships among the various rib forces are needed to determine the net radiation. Such a solution for a realistic panel driven by random excitation, as given, for example, in Ref. 1, is very complicated because it requires the inclusion of rib interactions and the determination of second-order statistics. Because of the uncertainties in specifying the driving TBL pressure and panel properties, such a detailed solution may not be warranted, and an approximate formulation of the broadband radiated power spectrum due to the rib forces may be sufficient.

An approach to directly estimating the broadband radiated power spectrum due to rib forces, without integrating a narrow-band spectrum, was developed in Ref. 2. Its derivation was based on the assumption that the rib forces can be considered to radiate independently and with equal power per unit force. In order to develop explicit results, a mass- or spring-loaded string in a light acoustic medium was used as the structural model. The purpose of this second paper is to establish the criterion under which the radiation assumption is justified, and to show that the formalism is applicable to ribbed metal (e.g., steel or aluminum) plates in water.

Many authors have contributed to the understanding of the vibration, radiation, and scattering of fluid-loaded ribbed panels. Representative citations include Leppington,[3] Maidanik and Dickey,[4] Cooper and Crighton,[5] Howe and Shah,[6] Mead and Mallik,[7] and Mace.[8]

II. APPLICABILITY OF RADIATION ASSUMPTIONS

A. Idealized model

The TBL pressure excitation is broadband in frequency but, for panels that are lightly or moderately damped, the significant responses, including power radiated, occur at resonance frequencies. At a resonance, the rib forces are correlated and have relative phases principally determined by the wave numbers of dominant structural waves on the fluid-loaded ribbed panel. These panels are often constructed having identical ribs and uniform rib spacing. The response of such a panel may be estimated by considering its infinite extension, for which the relevant wave numbers are those of the Bloch (or Floquet) waves of the system. These are infinite sets of wave numbers, $\pm\alpha_n$, $-\infty<n<\infty$, having elements that are related by $\alpha_n=\alpha_0+2\pi n/s$, where s is the rib spacing and α_0 is defined as that element having a positive imaginary part and a real part whose magnitude does not exceed π. (It is noted that the symbol for rib spacing is here given by ''s,'' in contrast to the notation ''d'' used in Ref. 2.)

Consider an otherwise uniform infinite membrane having line masses attached at $x=ms$, $-\infty<m<\infty$, each with mass per unit length μ. Assume that the excitation and response are invariant in the panel direction parallel to the line attachments. The side of the membrane to which the masses are attached is vacuum-backed, and the other side is in contact with a semi-infinite acoustic medium. Following the derivation in Ref. 9, but applying it to a membrane rather than a plate, accounting for the acoustic medium, and assuming no variation in response in the direction parallel to an attachment, the dispersion relationship for free wave num-

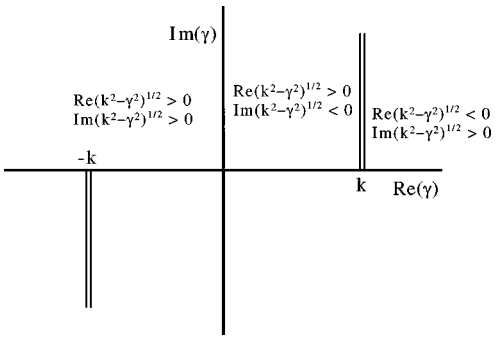


FIG. 1. Branch cuts used in determination of poles.

bers in the system is found to be given by Eq. (1), in which $Z(\gamma)$ is given by Eq. (2):

$$1-\frac{i\omega\mu}{s}\sum_{m=-\infty}^{\infty}\frac{1}{Z(\alpha+2\pi m/s)}=0, \tag{1}$$

$$Z(\gamma)=-i\omega M[1-(\gamma/k_p)^2]+\frac{\rho\omega}{\sqrt{k^2-\gamma^2}}. \tag{2}$$

$Z(\gamma)$ is the wave number impedance of the fluid-loaded uniform membrane in which γ is the wave number for the direction transverse to the attachments, and ω is equal to 2π times the frequency. The membrane *in vacuo* is represented by the first expression on the right hand side of Eq. (2), in which M is the membrane's mass per unit area, and k_p is the wave number of free waves on the uniform membrane *in vacuo*. The fluid impedance is given by the second expression in which ρ is the mass density of the fluid, and k is the acoustic wave number. Time dependence in the form $\exp(-i\omega t)$ has been assumed.

For the membrane *in vacuo*, the identification of the Bloch wave numbers is unique. For the membrane in an acoustic medium, the identification of waves through poles is not unique, but is dependent upon the choice of branch cut and may result in multiple independent sets of wave numbers. The branch cuts used here are shown in Fig. 1, along with the corresponding evaluation of $\sqrt{k^2-\gamma^2}$. This choice tends to maximize the number of significant wave responses that are explicitly represented by pole contributions.

B. Power radiated per rib

1. Infinite number of ribs

Let each of the attachments exert a complex-valued normal force on the membrane, with F_m specifying the force at $x=ms$. The acoustic power, P_∞^{rad}, radiated by one of these forces in the presence of the others is given by Eq. (3), in which the asterisk represents the complex conjugate:

$$P_\infty^{\mathrm{rad}}=\frac{1}{2\pi}\sum_{m=-\infty}^{\infty}\mathrm{Real}(F_0F_m^*)\int_{-k}^{k}\frac{\cos(ms\gamma)}{|Z(\gamma)|^2}\frac{\rho\omega}{\sqrt{k^2-\gamma^2}}d\gamma. \tag{3}$$

Under TBL excitation, the response is random, and the radiated power must be written as an expected value, $\langle P_\infty^{\mathrm{rad}}\rangle$. It can be shown that, if the response is dominated by residues of Bloch wave poles and the excitation is wave number-white, then

$$\langle\mathrm{Real}(F_0F_m^*)\rangle=\langle|F_0^2|\rangle\cos(ms\alpha_{0r})\exp(-ms\alpha_{0i}), \tag{4}$$

where the asterisk signifies the complex conjugate, $\langle|F_0^2|\rangle$ is the mean squared force at any rib and $\alpha_{0r}=\mathrm{Real}(\alpha_0)$ and $\alpha_{0i}=\mathrm{Imag}(\alpha_0)>0$. With these simplifications, the summation over m can be performed explicitly, and the average acoustic power radiated per rib (all ribs are equivalent) is given by Eq. (5a):

$$\langle P_\infty^{\mathrm{rad}}\rangle=\frac{1}{2\pi}\langle|F_0^2|\rangle\int_0^k\frac{d\gamma}{|Z(\gamma)|^2}\frac{\rho\omega}{\sqrt{k^2-\gamma^2}}\times[g(\gamma+\alpha_0)+g(\gamma-\alpha_0)], \tag{5a}$$

where

$$g(\gamma\pm\alpha_0)=\frac{\sinh(s\alpha_{0i})}{\cosh(s\alpha_{0i})-\cos[(\gamma\pm\alpha_{0r})s]}. \tag{5b}$$

2. Finite number of ribs

A panel of finite length with a finite number of attachments can be addressed approximately by examining an infinite panel having a finite number of attachments. The issues are the variation in radiated power among a finite number of ribs and how well each is represented by $\langle P_\infty^{\mathrm{rad}}\rangle$. Assuming the Bloch wave numbers are still applicable, the power $\langle P_{n,N}^{\mathrm{rad}}\rangle$ radiated by the nth rib on an infinite membrane having N ribs is given by Eq. (6a):

$$\langle P_{n,N}^{\mathrm{rad}}\rangle=\frac{\langle|F_0^2|\rangle}{2\pi}\sum_{m=1}^{N}\cos[(n-m)s\alpha_{0r}]\exp[-|n-m|s\alpha_{0i}]\times\int_{-k}^{k}\frac{\cos[(m-n)s\gamma]}{|Z(\gamma)|^2}\frac{\rho\omega}{\sqrt{k^2-\gamma^2}}d\gamma. \tag{6a}$$

The average over all ribs, $\langle P_N^{\mathrm{rad}}\rangle$, is given by Eq. (6b):

$$\langle P_N^{\mathrm{rad}}\rangle=\frac{1}{N}\sum_{n=1}^{N}\langle P_{n,N}^{\mathrm{rad}}\rangle. \tag{6b}$$

In what follows, the operator $\langle\,\rangle$ will no longer be explicitly shown, and the expected value will be implied in all powers and quadratic quantities.

C. Calculated results

Although the panel is modeled as a membrane, its *in vacuo* wave speed will be made frequency-dependent, so that the panel properties can mimic those of a thin plate in bending. Results will be presented for a membrane panel representing a 5.0 cm thick steel plate having mass density 7.8 g/cm^3 ($M=39.0$ g/cm^2). The pseudo-bending wave speed *in centimeters per second* is approximated by $950(hf)^{1/2}$, where h is the panel thickness *in centimeters* and f is the frequency *in Hertz*. The ambient acoustic medium is assumed to be water for which $\rho=1.0$ g/cm^3 and $c=150\,000$ cm/s. These parameters result in a nominal bending-acoustic coincidence frequency of 5000 Hz. The ribs are assumed to be line mass attachments spaced 50 cm

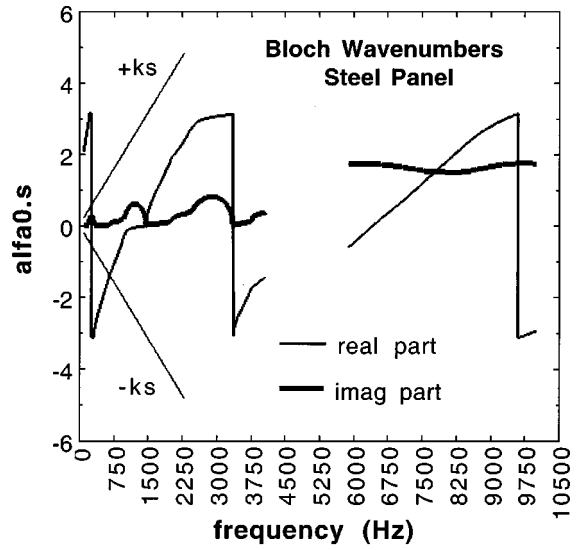


FIG. 2. Bloch wave numbers for steel panel.

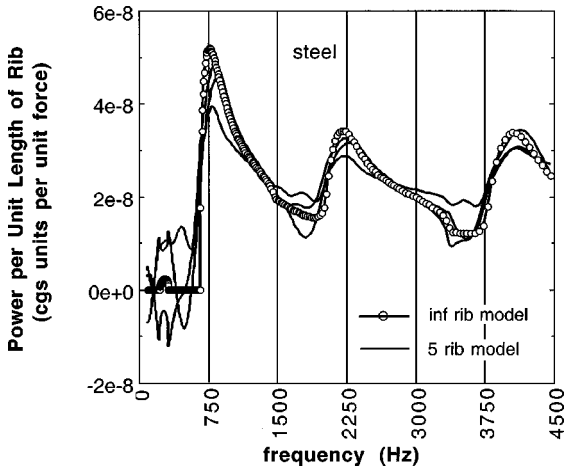


FIG. 3. Variability among radiated powers for five rib steel panel.

apart and having mass per unit length equivalent to a 75 cm^2 cross section of the same material as the panel. This results in $\mu = 585.0$ g/cm.

1. Bloch wave numbers

Figure 2 presents a graph of $\alpha_0 s$ vs frequency for the steel panel, which is shown in two frequency regions. In the lower region, the Bloch wave number is derived from the subsonic wave pole of the uniform fluid-loaded panel. In the higher frequency region, the Bloch wave is derived from the leaky (supersonic) wave pole of the uniform panel. No results are shown for the frequency interval in which the subsonic and leaky waves may have comparable contributions, or branch cut contributions are thought to be comparable to the residue contributions. Also shown are graphs of $\pm ks$ vs frequency. Although the Bloch wave dispersion for panels *in vacuo* shows pure stop bands, communication between bays via the fluid path prevents this from happening for fluid-loaded panels, except at the lowest frequencies. Figure 2 indicates a stop band in the frequency range 220–320 Hz, and quasi-stop bands in the approximate interval 1000–1500 Hz and in the neighborhood of 3000 Hz.

2. Power radiated per rib

Figure 3 compares P_∞^{rad} with $P_{3,5}^{\mathrm{rad}}$, $P_{4,5}^{\mathrm{rad}}$, and $P_{5,5}^{\mathrm{rad}}$ for a steel panel with five ribs, where the nth rib is located at $x = ns$. These were calculated using Eqs. (5) and (6a) and the wave numbers of Fig. 2. This figure shows a steep rise in the power radiated per rib in a narrow frequency interval just below 750 Hz. Above 750 Hz, P_∞^{rad} is representative of the others. Below 750 Hz there is large variability among the ribs, and the power attributed to a particular rib may be negative, indicating that some locations are radiated power sinks and some are sources. Their sum, however, is a positive radiated power. The steep rise in P_∞^{rad} occurs at 663 Hz. At higher frequencies, $|s\alpha_{0r}|$ is smaller than ks, and the set of Bloch waves has at least one radiating member. At lower frequencies, the Bloch waves are all subsonic and nonradiating, and contributions from the ribs nullify one another, except in the low frequency stop band where the communication among ribs is attenuated.

3. Power radiated by isolated force

The power P_1^{rad} radiated by a single isolated force F_0 is given by Eq. (7):

$$P_1^{\mathrm{rad}} = \frac{|F_0^2|}{\pi} \int_0^k \frac{\rho\omega}{\sqrt{k^2-\gamma^2}} \frac{1}{|Z(\gamma)|^2} d\gamma. \tag{7}$$

Consider the behavior of the functions $g(\gamma \pm \alpha_0)$ in Eqs. (5). For $s\alpha_{0i} \gg 1$, $g \to 1$, and the right hand side of Eq. (5a) approaches that of Eq. (7). When $s\alpha_{0i} \ll 1$, $g(\gamma \pm \alpha_0) \to 0$, except when $\cos[(\gamma \pm \alpha_{0r})s] \to 1$, and g has the integral property[10] shown in Eq. (8):

$$\begin{aligned} \lim_{s\alpha_{0i}\to 0} & \int_{2n\pi-\varepsilon}^{2n\pi+\varepsilon} \frac{\sinh(s\alpha_{0i})}{\cosh(s\alpha_{0i})-\cos(s\beta)} d(s\beta) \\ &= \lim_{s\alpha_{0i}\to 0} \int_{-\varepsilon}^{\varepsilon} \frac{\sinh(s\alpha_{0i})}{\cosh(s\alpha_{0i})-\cos(s\beta)} d(s\beta) \\ &= \lim_{s\alpha_{0i}\to 0} 2\tan^{-1}\left[\frac{\tan(s\beta/2)}{\tanh(s\alpha_{0i}/2)}\right]_{-\varepsilon}^{\varepsilon} = 2\pi. \end{aligned} \tag{8}$$

Therefore, g has a sifting property similar to that of a delta function, and its application to Eq. (5a) yields Eq. (9), in which the summation is over those integers j for which $(\alpha_{0r} - 2\pi j/s)^2 < k^2$:

$$\begin{aligned} \lim_{s\alpha_{0i}\to 0} P_\infty^{\mathrm{rad}} = \frac{|F_0^2|}{s} \sum_j & \frac{\rho\omega}{\sqrt{k^2-(\alpha_{0r}-2\pi j/s)^2}} \\ & \times \frac{1}{|Z(\alpha_0 - 2\pi j/2)|^2}. \end{aligned} \tag{9}$$

If the frequency is high enough so that $ks > \pi$, then, because $-\pi < s\alpha_{0r} < \pi$, there will be at least one term in the series of Eq. (9). As ks increases to multiples of π, more terms will be included, and Eq. (9) becomes a discrete approximation to P_1^{rad} as given by Eq. (7).

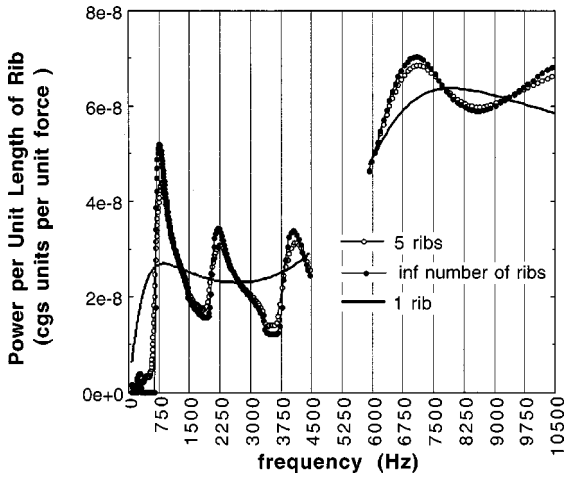


FIG. 4. Comparison of $P_{\infty}^{\mathrm{rad}}$, P_{1}^{rad}, and P_{5}^{rad} for the steel panel.

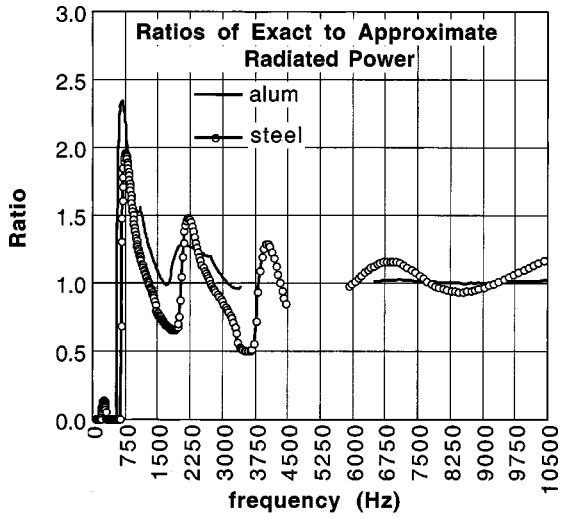


FIG. 5. Ratio of $P_{\infty}^{\mathrm{rad}}$ to P_{1}^{rad} for steel and aluminum panels.

4. Comparisons

Figure 4 compares $P_{\infty}^{\mathrm{rad}}$, P_{1}^{rad}, and P_{5}^{rad}. For the parameters chosen in the example calculations, ks is greater than π at frequencies above 1500 Hz. Figure 4 indicates that, in this frequency range, $P_{\infty}^{\mathrm{rad}}$ and P_{5}^{rad} are almost identical to one another. Because the Bloch wave number tends to remain within the radiating band, $-k<\alpha_{0r}<k$, for about an octave below $ks=\pi$, agreement is maintained down to $ks=\pi/2$ (750 Hz in this example). For as few as five attachments, the agreement between $P_{\infty}^{\mathrm{rad}}$ and P_{5}^{rad} is within about 25% at 750 Hz and improves at higher frequencies. Additionally, P_{1}^{rad} appears to be a good approximation to $P_{\infty}^{\mathrm{rad}}$ above $ks=\pi/2$.

D. Evaluation

It is shown in Ref. 2 that when the rib forces can be assumed to radiate independently and identically, the panel's radiation resistance due to the ribs is proportional to the radiated power per unit rib force. The intent is to ultimately use the approximation of the rib-associated radiation resistance to estimate the acoustic power radiated by a ribbed panel in one-third octave frequency bands under broadband TBL excitation. If the broadband power put into the panel by the TBL excitation is insensitive to the radiation characteristics of the panel, the radiated power is related to the input power P^{in} by Eq. (10). R_{rad} is the effective radiation resistance due to the ribs and R_{diss} is the resistance due to structural dissipation and everything else:

$$P^{\mathrm{rad}}=\frac{R_{\mathrm{rad}}}{R_{\mathrm{rad}}+R_{\mathrm{diss}}}P^{\mathrm{in}}. \tag{10}$$

If R_{diss} is much smaller than the actual radiation resistance, a relative error in estimating R_{rad} amounting to as much as a factor of 2 has little effect on the estimate of P^{rad}. If R_{diss} is larger than R_{rad}, a factor of 2 error (in either direction) in estimating R_{rad} leads to a maximum 3 dB error in P^{rad}. This is a tolerable error for estimates of responses to TBL excitation in one-third octave frequency bands. Therefore, the approximation to the power radiated by a rib is adequate if it is within a factor of 2 (in either direction) of the actual value. Figure 5 presents the ratio of $P_{\infty}^{\mathrm{rad}}$ to P_{1}^{rad} for the steel membrane panel. The factor of 2 criterion is met for frequencies above 750 Hz, corresponding to $ks>\pi/2$. Also shown in the figure is the corresponding result for an all aluminum ribbed panel of identical geometry. The mass density of aluminum is 2.7 g/cm^3 so that $M=13.5$ g/cm^2 and $\mu=202.5$ g/cm, and the pseudo-bending wave speed is the same as for the steel membrane panel.

III. VALIDATION OF APPROXIMATION FOR MEMBRANE

A ''steel'' ribbed membrane, having the properties and dimensions given above, will be used for illustration. The membrane is vacuum-backed, but the other side faces a semi-infinite region of water.

A. Approximate solution

Analyses and discussions in Ref. 2, particularly those in Sec. VIII, outline the procedure for approximating the rib-related radiated power, when the rib forces can be considered to radiate independent of one another. Whereas the developments in Refs. 2 did not include fluid-loading except for specifying a radiation component for the input admittance, fluid-loading will now be explicitly accounted for in the calculations. The steps are:

(1) Compute $Y_{\infty}(\omega)$, the input admittance for a line force acting on the infinite, uniform fluid-loaded membrane, and $Y_r(\omega)$, the acoustic power radiated per unit line force:

$$Y_{\infty}(\omega)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\left[i\frac{T}{\omega}(1-i\eta)\alpha^2-iM\omega+\frac{\rho\omega}{\sqrt{k^2-\alpha^2}}\right]^{-1}d\alpha, \tag{11a}$$

$$Y_r(\omega)=\frac{1}{2\pi}\int_{-k}^{+k}\frac{\rho\omega}{\sqrt{k^2-\alpha^2}}\left|\frac{T}{\omega}(1-i\eta)\alpha^2-M\omega-\frac{i\rho\omega}{\sqrt{k^2-\alpha^2}}\right|^{-2}d\alpha. \tag{11b}$$

$T=Mc_m^2$ is the membrane tension per unit length, η is the membrane's loss factor, and $k=\omega/c$. A time dependence $\exp(-i\omega t)$ is assumed.

(2) Compute β_a and β_r, the spatially averaged equivalent damping coefficients for the attachment dissipation and radiation, respectively. These are obtained from Eqs. (45) of Ref. 2:

$$\beta_a=\frac{\eta_a\omega\dfrac{d}{d\omega}\mathrm{Re}[iY_a(\omega)]}{s|Y_\infty(\omega)+Y_a(\omega)|^2}, \tag{12a}$$

$$\beta_r=\frac{Y_r(\omega)}{s|Y_\infty(\omega)+Y_a(\omega)|^2}. \tag{12b}$$

The admittance of the line attachment is given by $Y_a(\omega)$, *a change in notation from* Ref. 2, with associated loss factor η_a. For a line mass attachment,

$$Y_a(\omega)=\frac{\eta_a+i}{\omega\mu} \tag{13a}$$

$$\frac{d}{d\omega}\mathrm{Re}[iY_a(\omega)]=\frac{1}{\mu\omega^2}. \tag{13b}$$

(3) Find the mean squared velocity of the equivalent panel. From Eq. (48) of Ref. 2, modified for fluid-loading, Eq. (14) is obtained:

$$\langle v^2\rangle=S(\omega)\int_{-\infty}^{+\infty}\left|\frac{T}{\omega}(1-i\eta)\alpha^2-M\omega-i(\beta_a+\beta_r)-\frac{i\rho\omega}{\sqrt{k^2-\alpha^2}}\right|^{-2}d\alpha. \tag{14}$$

The mean squared wave number-frequency spectral density of the excitation, S, is wave number-white, and a function of ω but not α.

(4) Find the mean squared attachment force. From the discussion of Ref. 2, continuity between the membrane and an isolated attachment requires that the attachment force be given by Eq. (15):

$$\langle|F^2(\omega)|\rangle=\frac{\langle|v^2(\omega)|\rangle}{|Y_\infty(\omega)+Y_a(\omega)|^2}. \tag{15}$$

(5) Find the acoustic power radiated in the frequency band per rib by considering the mean squared rib attachment force in isolation. This is given by Eq. (16), in which $\Delta\omega$ is the bandwidth. When evaluating the quantities in Eqs. (11)–(16), ω is taken to be the band's center frequency:

$$\langle\mathbf{P}_{\mathrm{rad}}\rangle=\langle|F^2(\omega)|\rangle Y_r(\omega)\Delta\omega. \tag{16}$$

B. Exact solution

The ''exact'' solution is generated from the results of Ref. 9, after modification for a one-dimensional vice two-dimensional problem, application to a membrane rather than a plate, and the addition of fluid-loading. An equivalent to Eq. (20) of Ref. 9, which represents the Fourier transform of the plate normal velocity due to arbitrary surface excitation, is written along with a corresponding form for its complex conjugate. The two are multiplied and the surface excitation is assumed to be a spatially homogeneous random function having a wave number-white spectrum. Appropriate statistics and limits are taken to arrive at an expression for the mean squared spectral density in wave number space of the normal velocity of the membrane with attachments. This spectral density has the form of Eq. (17):

$$S_v(\alpha)=S_v^0(\alpha)+S_v^{0a}(\alpha)+S_v^a(\alpha). \tag{17}$$

The first term on the rhs represents the velocity wave number spectrum which the excitation would produce on the uniform membrane without attachments. The third term represents the velocity wave number spectrum produced on the uniform membrane by the induced attachment forces. The second term represents that part of the wave number spectrum corresponding to interaction between the velocity caused by the external excitation and that due to the attachment forces.

The *total* acoustic power radiated per ''bay'' (i.e., portion of the membrane between two adjacent attachments) within the frequency band is given by Eq. (18):

$$\mathbf{P}_{\mathrm{rad}}=s\int_{\Delta\omega}d\omega\int_{-k}^{+k}S_v(\alpha)\frac{\rho\omega}{\sqrt{k^2-\alpha^2}}d\alpha. \tag{18}$$

The acoustic power radiated within the frequency band per bay by that part of the velocity field due to the attachment forces alone, and equivalent to the power radiated per attachment, is given by Eq. (19):

$$\mathbf{P}_{\mathrm{rad}}^a=s\int_{\Delta\omega}d\omega\int_{-k}^{+k}S_v^a(\alpha)\frac{\rho\omega}{\sqrt{k^2-\alpha^2}}d\alpha. \tag{19}$$

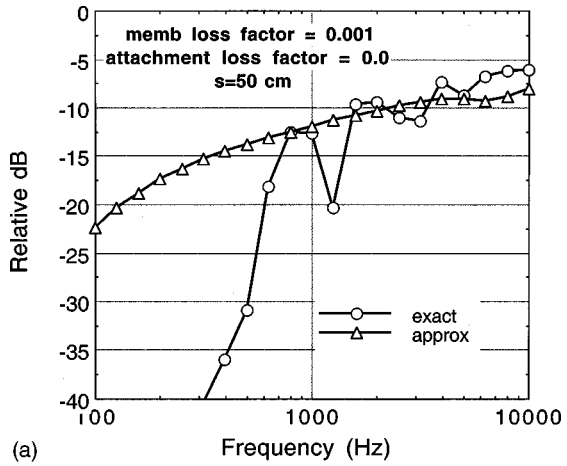
The results of Eq. (19) will be compared to those of Eq. (16) for the fluid-loaded membrane with attachments described above.

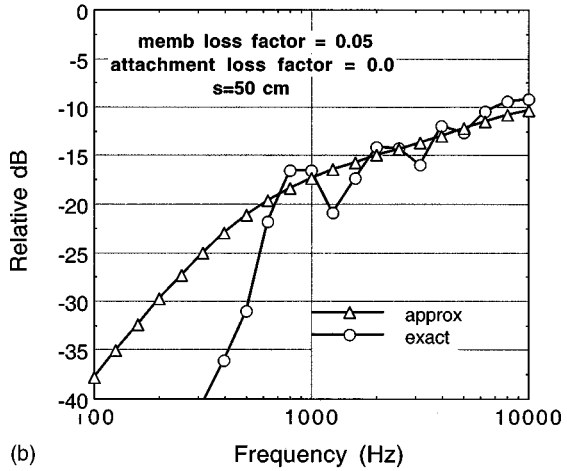
C. Comparisons

Calculations of the approximate and exact attachment-related radiated power per membrane bay (or per attachment), in one-third octave frequency bands, were made for the configuration described above and for the following three sets of loss factors: (1) $\eta=0.001$ and $\eta_a=0$, (2) $\eta=0.05$ and $\eta_a=0$, (3) $\eta=0.001$ and $\eta_a=0.05$. The excitation spectral density, $S(\omega)$, was assumed to have unit magnitude in the cgs system. The exact and approximate results for each set are shown in Figs. 6(a)–(c). The graphs show that for frequencies above 750 Hz ($ks>\pi/2$), the approximation is within 3 dB of the exact result, with the exception of the 1260 Hz band. This is a frequency band controlled by one quasi-stop band, as shown in the graph of Bloch wave numbers for the same system given in Fig. 2. The exact radiated power is reduced because less power can be put into the membrane in such a band.

IV. VALIDATION OF APPROXIMATION FOR PLATE

Although the illustrative analyses have considered a string (or a membrane), the conclusions are applicable to a plate in bending, if the torque between the plate and reinforcing beam is neglected. Because a line torque (moment) drive on the plate is generally a weaker radiator than a line force drive, its omission does not significantly affect the power





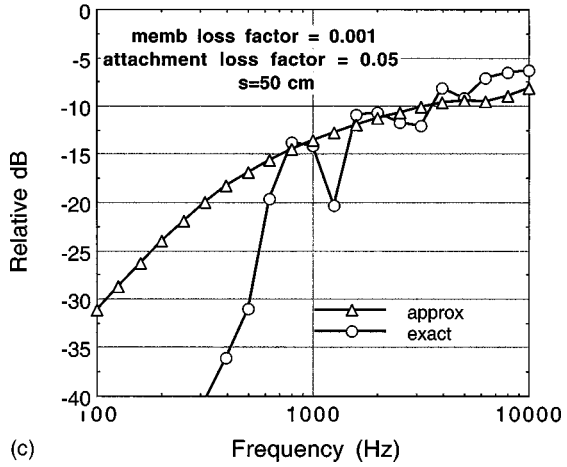


FIG. 6. Comparisons of calculated radiated powers per rib for ''steel'' membrane; $s=50$ cm [(a) $\eta=0.001$, $\eta_a=0$; (b) $\eta=0.05$, $\eta_a=0$; (c) $\eta=0.001$, $\eta_a=0.05$].

radiated. It will affect resonance frequencies, but their precise locations are not important for broadband power estimates. The formalism for the plate is identical to that for the membrane, but requires a minor change in the integrals that define the mean squared velocity and line admittances. In the integrals of Eqs. (11) and (14), the second-order wave number impedance of the membrane is replaced by the fourth-order impedance of the plate, and the membrane's tension per unit length, T, is replaced by the plate's bending rigidity, D. Equations (20) and (21) replace Eqs. (11) and (14), and Eqs. (12), (15), and (16) remain applicable as is. Equations (13) describes the line mass attachments:

$$Y_\infty(\omega)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\left[\,i\frac{D}{\omega}(1-i\eta)\alpha^4-iM\omega+\frac{\rho\omega}{\sqrt{k^2-\alpha^2}}\right]^{-1}d\alpha, \tag{20a}$$

$$Y_r(\omega)=\frac{1}{2\pi}\int_{-k}^{+k}\frac{\rho\omega}{\sqrt{k^2-\alpha^2}}\times\left|\frac{D}{\omega}(1-i\eta)\alpha^4-M\omega-\frac{i\rho\omega}{\sqrt{k^2-\alpha^2}}\right|^{-2}d\alpha, \tag{20b}$$

$$\langle v^2\rangle=S(\omega)\int_{-\infty}^{+\infty}\left|\frac{D}{\omega}(1-i\eta)\alpha^4-M\omega-i(\beta_a+\beta_r)-\frac{i\rho\omega}{\sqrt{k^2-\alpha^2}}\right|^{-2}d\alpha. \tag{21}$$

A. Description of plate

Calculations will be presented for two infinitely extended steel plates in water, each having an infinite number of equally spaced line mass attachments having cross sectional areas of 75 sq cm. The physical parameters for steel and water given above will be used. The first plate is 1 cm thick with a bending loss factor of 0.05 and a rib spacing of 50 cm. The second plate is 5 cm thick with a loss factor of 0.005 and a rib spacing of 100 cm. The driving spectral density is again assumed to be wave number-white and frequency independent, i.e., $S(\omega)=S_0$.

B. Comparison of results

Figure 7(a) compares approximate and exact radiated power per bay for the 1 cm thick plate with $s=50$ cm, $\eta=0.05$, and $\eta_a=0$. Good agreement is obtained at frequencies of 800 Hz and above ($ks>\pi/2$). Figure 7(b) presents a comparison for the 5 cm thick plate with $s=100$ cm, $\eta=0.005$, and $\eta_a=0$. Good agreement is achieved at 400 Hz and above, again corresponding to $ks>\pi/2$.

V. DISCUSSION

Reference 2 developed an approach to estimating broadband (e.g., one-third octave frequency band) levels of acoustic power radiated due to rib/panel interaction under TBL-like excitation. The approximation is generally valid when the frequency band contains at least two resonances of the finite panel and each of the ribs can be considered to radiate the same amount of acoustic power per unit rib force. The first part of this paper showed that the assumption regarding rib radiation is justified at frequencies at which at least one Bloch wave of a periodically (or approximately periodically) reinforced panel is supersonic. This condition is always met when the acoustic wavelength is smaller than two times the rib spacing ($ks>\pi$), but will usually be satisfied when the acoustic wavelength is smaller than four times the rib spac-

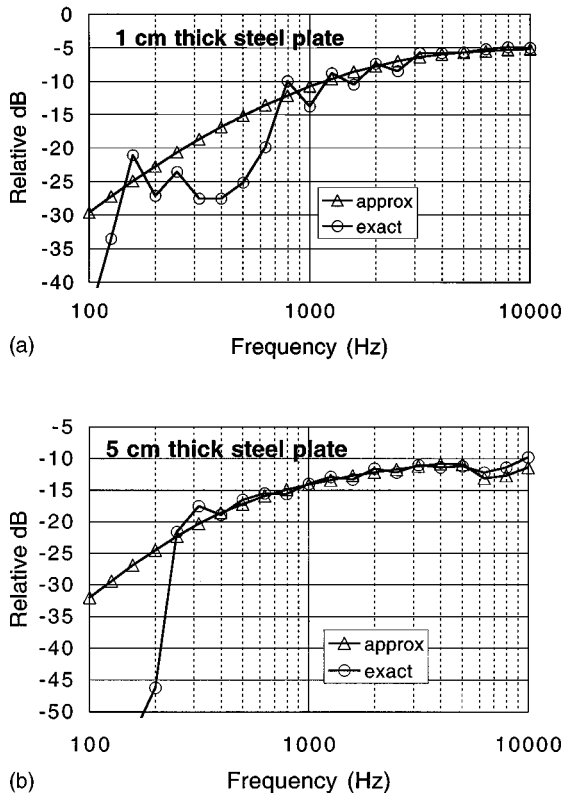


FIG. 7. Comparisons of calculated radiated powers per rib for steel plate [(a) $s=50$ cm, $\eta=0.05$, $\eta_a=0$; (b) $s=100$ cm, $\eta=0.005$, $\eta_a=0$].

ing ($ks>\pi/2$). The condition of independent radiation allows the ribbed panel to be represented by a uniform panel with spatially averaged radiation properties. An average mean squared rib force within a frequency band can then be calculated, and the acoustic power radiated can be estimated by calculating the power radiated by a single line force and multiplying by the number of ribs. This is important because it means that a Bloch wave number model, or any other model requiring explicit inclusion of rib interactions, is not needed.

In the second part of this paper, the formalism was tested by considering the power radiated by an infinite membrane with uniformly spaced identical line mass attachments, and comparing the approximations of the power radiated per attachment to an exact calculation. The excitation was considered invariant in the direction parallel to the attachments, so that only the zero wave number component for that direction was considered. The agreement was within 3 dB when the spacing wavelength criterion was met, except in a frequency band controlled by one quasi-stop band of the reinforced panel. In such a band, the power accepted by the uniformly reinforced membrane is reduced, and will be overestimated by the approximation. The formalism was then applied to a ribbed steel plate in water by modifying the wave number impedance that describes the uniform panel. Similar agreement with exact results was achieved.

In practice, the panels are finite and the excitation is not uniform along the attachment; therefore, modes of the panel associated with the direction parallel to a rib (cross modes) must be considered in estimating the radiated power. This can have the effect of mitigating the effects of quasi-stop bands on the approximation because the admittance of the reinforcing member, say a beam, will be dependent on the cross mode wave number. The bandwidth of a stop band depends on the relationship between the panel properties and the attachment admittance, and will change as the cross mode wave number changes. This can reduce the net discrepancy between exact and approximate calculated radiated power.

ACKNOWLEDGMENT

This work was funded through the Office of Naval Research Hydroacoustics Program, L. Patrick Purtell, Program Manager.

[1] R. A. Mkhitarov, ''Sound radiation from a bounded thin inhomogeneous plate reinforced with *N* ribs and driven by boundary-layer pressure fluctuations,'' Sov. Phys. Acoust. **20**, 528–533 (1975).
[2] M. L. Rumerman, ''Estimation of broadband acoustic power due to rib forces on a reinforced panel under TBL-like pressure excitation: I. Derivations using string model,'' J. Acoust. Soc. Am. **109**, 563–575 (2000).
[3] F. G. Leppington, ''Acoustic scattering by membranes and plates with line constraints,'' J. Sound Vib. **58**, 319–332 (1978).
[4] G. Maidanik and J. Dickey, ''Quadratic and energy estimates of the partial response of ribbed panels,'' J. Acoust. Soc. Am. **94**, 1435–1444 (1993).
[5] A. J. Cooper and D. G. Crighton, ''Transmission of energy down periodically ribbed elastic structures under fluid loading: Spatial periodicity in the pass bands,'' Proc. R. Soc. London, Ser. A **454**, 2893–2909 (1998).
[6] M. S. Howe and P. L. Shah, ''Influence of mean flow on boundary layer generated interior noise,'' J. Acoust. Soc. Am. **99**, 3401–3411 (1996).
[7] D. J. Mead and A. K. Mallik, ''An approximate theory for the sound radiated from a periodic line-supported plate,'' J. Sound Vib. **61**, 315–326 (1978).
[8] B. R. Mace, ''Sound radiation from a plate reinforced by two sets of parallel stiffeners,'' J. Sound Vib. **71**, 435–441 (1980).
[9] M. L. Rumerman, ''Vibration and wave propagation in ribbed plates,'' J. Acoust. Soc. Am. **57**, 370–373 (1975).
[10] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 4th ed. (Academic, New York, 1980), p. 148, Sec. 2.553, Eq. (3).

Behavior of first guided wave on finite cylindrical shells of various lengths: Experimental investigation

Lionel Haumesser, André Baillard, Dominique Décultot, and Gérard Maze
Laboratoire d'Acoustique Ultrasonore et d'Electronique UMR CNRS 6068, Université du Havre, Place Robert Schuman, 76610 Le Havre, France

(Received 17 May 2000; revised 15 October 2000; accepted 18 October 2000)

Acoustic backscattering from elastic cylindrical shells of finite lengths, immersed in water, is investigated. These objects, characterized by the ratio of length over diameter ($L/2a=9.76$, 4.88, 2.44, a: outer radius), are excited by an obliquely incident plane acoustic wave. In the three cases studied here, the radii ratio b/a (b: inner radius) is fixed at 0.97. The investigated dimensionless frequency range extends over $10 \leqslant k_1 a \leqslant 50$ (k_1: wave number in water). The first guided wave, T_0, is of particular interest here. The influence of the shell's length on the backscattered pressure is experimentally observed in the time-angle and frequency-angle representations. In support of this experimental study, a time-domain representation is used by extending a theoretical model that provides a geometrical description of the helical propagation of the surface waves around the shell [Bao, J. Acoust. Soc. Am. **94**, 1461–1466 (1993)]. Theoretical results on cylindrical shells considered as infinitely long, with identical characteristics, are compared with both experimental representations. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1332382]

PACS numbers: 43.40.Fz, 43.20.Fn, 43.30.Gv [CBB]

I. INTRODUCTION

A number of authors have dealt with the acoustic scattering of a plane wave from infinitely long cylindrical shells.[1–4] Their works, both theoretical and experimental, describe the propagation of surface waves, generated at any incidence, and the influence of these waves on the scattered pressure spectra. More recent publications concentrate on shorter cylindrical shells. Brill *et al.*[5] only consider geometrical diffraction. The exact theoretical approach to the scattering phenomenon by a finite cylindrical shell is more complex. Harari *et al.*[6] develop an analytical model of pressure radiation from a finite cylindrical shell closed by plates on both ends. Morse *et al.*[7,8] investigate the frequency-angle domain above the critical angles of the S_0 and T_0 membrane waves (far from normal incidence), beyond the coincidence frequency region. A hybrid method combining full elasticity theory and the Kirchhoff diffraction integral is employed. The effect of extremity discontinuities, at critical incidences, is examined by Kaduchak *et al.*[9] and Gipson *et al.*[10] (the latter paper deals with a solid cylinder). Other authors pay particular attention to the analysis of target's limited length effects. Rumerman[11] shows how the evolution of vibration modes becomes discontinuous in the frequency-angle domain. Moreover, the same vibration mode of a given wave is experimentally identified at various frequencies for the same incidence angle by Lecroq *et al.*[12]

At oblique incidences, surface waves generated by an acoustic beam follow helical paths along the finite shell's axis.[13] They are continuously re-emitted along the travel path in directions related to respective characteristic angles. Moreover, they are partially reflected according to the ''Snell–Descartes'' law when encountering the cylindrical shell end: the reflection modulus is high due to impedance break at the extremity (steel/water interface). Thus, surface waves undergo one or more partial reflections at the end of the cylindrical shell. The geometrical model developed in Ref. 14 associates echo arrival times to helical paths followed by surface waves. The derived equations[14] depict these paths and corresponding times of travel in water, in the case of paths comprising of one end reflection. The amplitude of each echo, which is a function of the distance covered, is associated with the arrival time. Signals obtained from such calculations are compared with experimental measurements, at several incidence angles. It emerges from this study that echo arrival times tend to increase with deviation from normal incidence.

In this paper, the first guided wave T_0, also labeled $p=1$,[15] is studied. The index linking of ''whispering-gallery wave'' (l) and ''guided wave'' (p) define the order of appearance of the resonances for circumferential and guided waves, respectively, as the frequency increases. Certain theoretical results on infinite cylindrical shell are recalled in order to complement discussions. Time-signal recordings in the azimuthal plane enclosing their axis are successively made on three finite cylindrical shells of various lengths. The signals are recorded between $\beta=90°$ (normal incidence) and $\beta=60°$, at 1-deg intervals. In order to analyze them, the approached theoretical model presented by Bao[14] is partially adopted and developed: travel paths with several end reflections are taken into account. The analysis of results is achieved in the time-angle and frequency-angle domains, where effects of the length of the cylindrical shell are particularly highlighted.

II. INFINITE CYLINDRICAL SHELL: A BRIEF THEORETICAL RECALL

In this section, theoretical results on infinite cylindrical shells are recalled[15,16] in order to enable comparison with experimental results to be made. Full 3D elasticity theory is applied to an immersed empty infinite cylindrical shell. Expressions of the corresponding matrix elements are found in

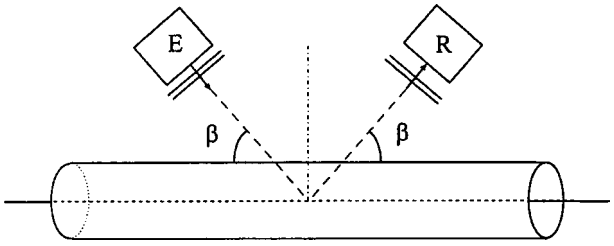


FIG. 1. Bistatic setup considered in theoretical calculations.

Ref. 12. Calculations yield the far-field form function. They are achieved by considering a bistatic configuration (Fig. 1); the emitter and receiver transducers are set in the Snell–Descartes configuration. The insonification direction is defined by the incidence angle β. The parameters used in the computations are equal to the values of the characteristics of the experimentally investigated targets; they are given in Sec. III.

Inverse fast Fourier transform (FFT) of the form function is performed over the reduced frequency band $0 \leqslant k_1 a \leqslant 70$. A time-signal representation, dependent on the incidence angle β, is obtained (Fig. 2). In the considered frequency band, only three surface waves are observable: (a) the S_0 wave, from normal incidence to angle β_{S0} given by the velocity at low frequencies of the first longitudinal mode for plates; (b) the first guided wave T_0, from normal incidence to the critical angle β_{T0}; (c) The Scholte–Stoneley or *A* wave, throughout the angular range.

The second figure concerning the infinite cylindrical shell (Fig. 3) presents the theoretical resonance spectra in the frequency range $10 \leqslant k_1 a \leqslant 50$, versus the incidence angle. The spectra are obtained by FFT calculations of the theoretical echo wave form signals over a duration of 1000 μs, without taking into account the direct reflection echo. Resonances related to the *A* wave are observable in the frequency range 20–35. Frequency shifts as a function of incidence angle have been described by the authors of Ref. 17. Closer to the characteristic incidence angles, resonances are shifted towards high frequencies. The continuum feature of this angular evolution is noted here for the S_0 and T_0 wave modes. The broken aspect in frequency is artificially introduced by the angular step (0.2 deg). In what follows, the T_0 wave is of main interest.

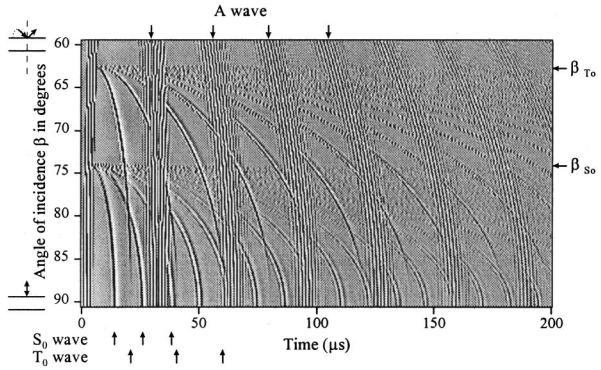


FIG. 2. Evolution of calculated echo arrival times from infinite cylindrical shell ($b/a = 0.97$, $a = 10.25$ mm) as a function of the incidence angle β. Amplitude increases linearly from black to white zone.

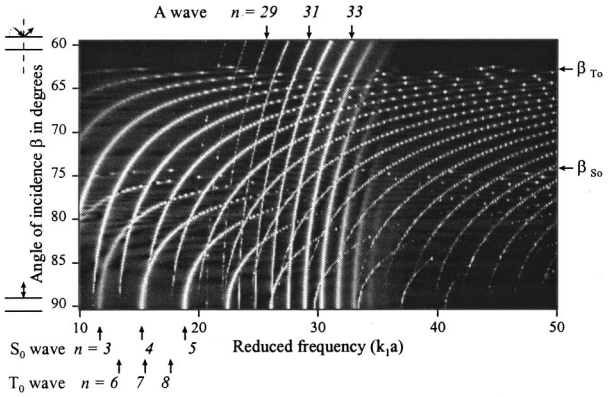


FIG. 3. Evolution of calculated resonance spectra from infinite cylindrical shell ($b/a = 0.97$) as a function of the incidence angle β. Amplitude increases linearly from black to white zone. The circumferential vibration mode is noted *n*.

III. THE TARGETS AND THE MEASUREMENT SETUP

The cylindrical shells are made of stainless steel of density $\rho = 7900\ \mathrm{kg\,m^{-3}}$, with longitudinal and transversal sound velocities of, respectively, $C_L = 5790\ \mathrm{m\,s^{-1}}$ and $C_T = 3200\ \mathrm{m\,s^{-1}}$. The outer radius is $a = 10.25$ mm and the shell thickness $e = 0.3$ mm. Three cylindrical shell lengths ($L = 200$, 100, and 50 mm) are successively studied. Thin rubber diaphragms (0.03 mm thick) ensure shell-end water tightness; they are glued at the cylindrical shell extremities making a 2-mm-wide ring. The air-filled cylinder is horizontally submerged in a circular water filled tank 3 m in diameter; it is suspended at the center of the circular path followed by the transducer rotating around it (Fig. 4). The angular displacement intervals are 1 deg each. The transducer is positioned 1 m away from the center of the target. Monostatic setup of the pulse MIIR[18,19] (Méthode d'Isolement et d'Identification des Résonances) is used. A short electric impulse (Dirac pulse) is converted into a volume wave by an emitter/receiver transducer. The transducer is a broadband Panametrics, model V392-SU (Fig. 5), with central frequency equal to 1 MHz ($k_1 a = 44$). The backscattered pressure is converted into an electric signal by the same transducer. A time-domain signature of the target is thus obtained at each incidence angle. Considering this particular setup in Fig. 4 (monostatic con-

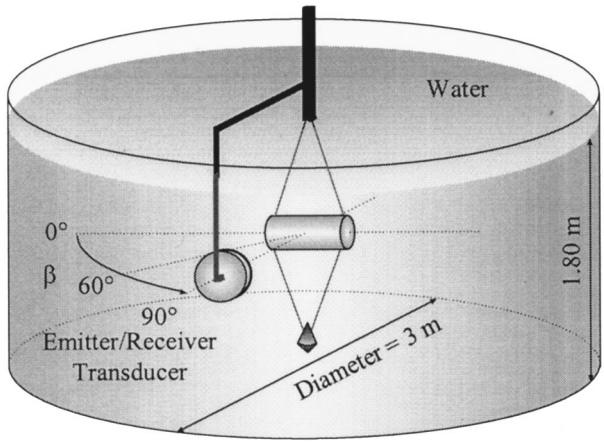


FIG. 4. Monostatic experimental setup.

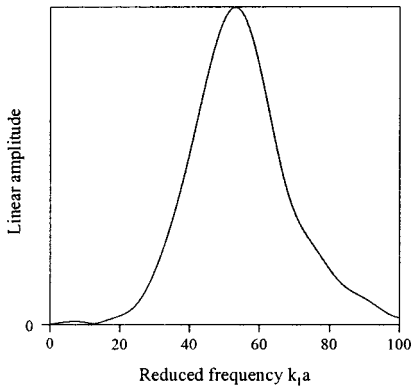


FIG. 5. Transducer passband with central frequency: $f_C = 1$ MHz ($k_1 a = 44$).

figuration), the received echoes include those representing travel paths of waves that have undergone odd number of reflections at the ends of the cylindrical shell.

The extremities of the finite cylindrical shells are considered as mirrors, from which an analogy between the experimental monostatic setup and the bistatic configuration considered in the theoretical scattering calculation can be drawn. In the following sections, time- and frequency-domain results are analyzed for the three finite cylindrical shells. Further comparison is made between experimental and theoretical results on infinite cylindrical shells.

IV. RESULTS AND DISCUSSION

A. Time-domain analysis

Time measurements from finite-length cylindrical shells ($L = 200$, 100, and 50 mm) are presented in Figs. 6(a)–8(a), successively. The time origin is arbitrarily fixed by the arrival time of the specular reflection echo at normal incidence ($\beta = 90°$). At the beginning of the recordings, echoes without periodic features come from direct reflections on the target extremities, the rubber diaphragms, and the anchoring mechanism. The S_0 wave echoes are observable close to normal incidence, within about $\pm 4°$, while A wave echoes appear with a weak relative amplitude throughout the angular range [first echoes of these two waves are indicated in Fig. 6(a)].

Results on the longer cylindrical shell are shown in Fig. 6. A series of nine echoes of the T_0 wave occupies the central part of the figure. The top view of Fig. 6(a) is the relevant time extract at $\beta = 80°$. Over the analyzed time window (200 μs), the T_0 wave has undergone one reflection at the cylindrical shell extremity ($r = 1$, r: number of shell end reflections). Through an example, we examine in detail the time interval between two successive echoes. The corresponding travel path is detailed in a 2D representation of the cylindrical shell [Fig. 9(a)]. The cylindrical shell is unrolled to obtain the topological equivalent of the cylinder. The considered T_0 surface wave is generated under the incidence angle β with a helical angle α (see the Appendix). At the shell extremity, it undergoes a reflection according to the Snell–Descartes law. It is continuously re-emitted lengthwise when propagating. However, the setup defines two particular lines (MM' and NN') where the surface wave is re-emitted in water in the direction of the transducer (for r odd). The points A, B, and C stand on one of these two lines. When the T_0 wave reaches B, the possibilities are: emission from B and reception at B', or one more helical lap (BC) and reception at C'. The time interval between two successive echoes is related to the distance: BC'-BB', i.e., one helical lap on the cylinder (from B up to C), minus a path in water (d_1). When β is fixed, this interval between any successive T_0 wave echoes is constant. This time interval is identical to the one in the case of an infinite length cylindrical shell. When deviating from normal incidence (decreasing β), the helical step increases. Calculated arrival times, taking into account paths on the shell and in water, provided by the geometrical approach, yield Fig. 6(b). The expressions used, derived in Ref. 14, are briefly depicted in the Appendix (note changes in notations). They are modified in order to take into account travel paths involving many shell end reflections. The times are calculated from phase and group velocities assumed constant. Approximated expressions of velocities on infinite cylindrical shell are used [relations (17) of Ref. 14]. For a surface wave at a given incidence angle, the phase velocity is calculated for each mode, and the group velocity

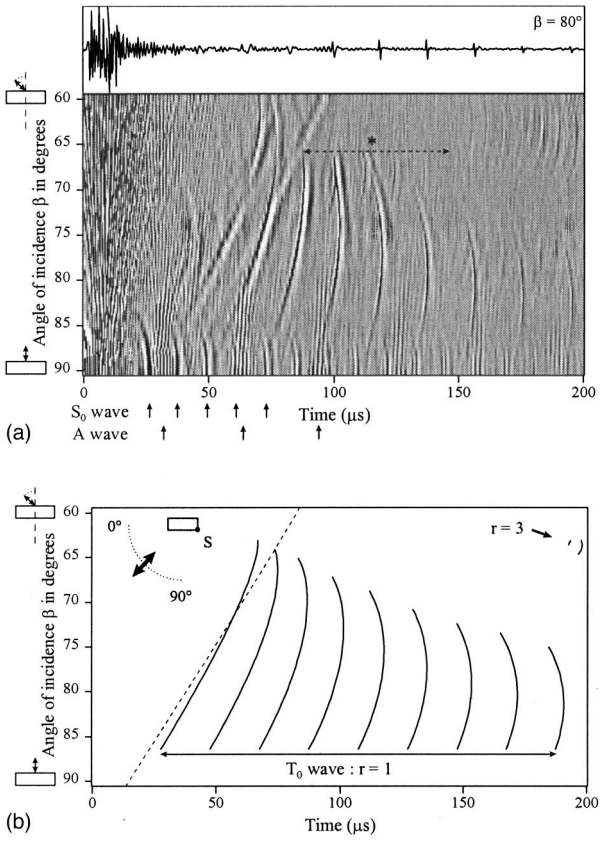


FIG. 6. (a) Evolution of experimental echo arrival times from finite cylinder ($L = 200$ mm) as a function of the incidence angle β. Amplitude increases linearly from black to white zone. The asterisk marks unexplained echoes. (b) Evolution of calculated echo arrival times of the T_0 wave from a finite cylindrical shell ($L = 200$ mm) as a function of incidence angle β (solid lines). Direct reflection from the corner S (dashed line).

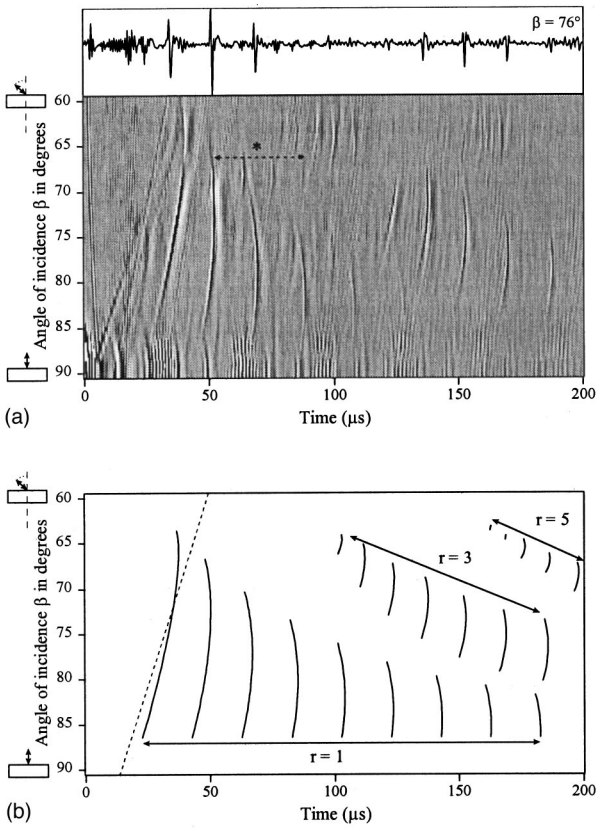


FIG. 7. (a) Figure 6(a) for $L=100$ mm. (b) Figure 6(b) for $L=100$ mm.

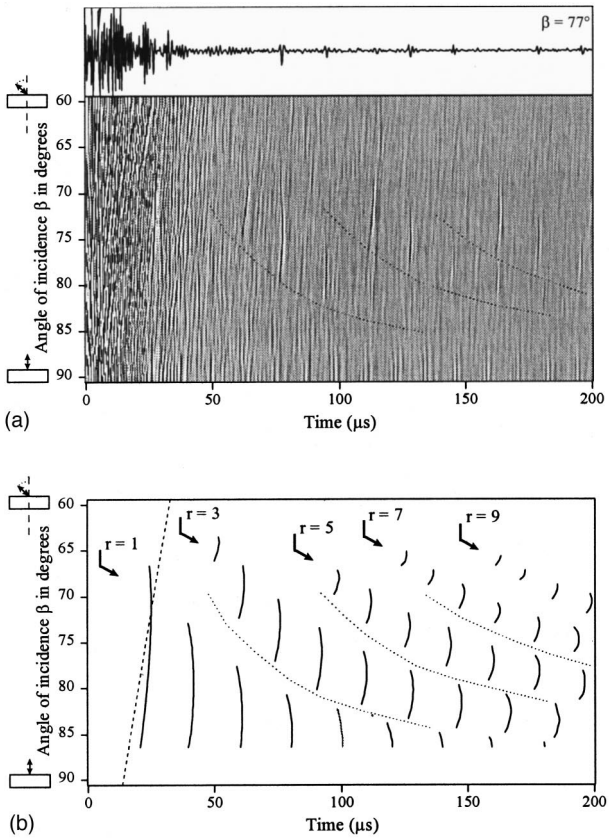


FIG. 8. (a) Figure 6(a) for $L=50$ mm. The patterns $r=3$, 5, and 7 are underlined (dotted line). (b) Figure 6(b) for $L=50$ mm. The patterns $r=3$, 5, and 7 are underlined (dotted line).

is estimated between two successive modes (Fig. 3). In the experiment frequency range, the wave phase and group velocity dispersion is weak enough to be neglected. The following averaged values are used in the computations:

$$C^{ph}_{T_0}=C^{g}_{T_0}=3200\ \mathrm{m\,s^{-1}}.$$

The resulting curve network represent the evolution of different echoes of the T_0 wave as a function of the incidence angle (solid line); echo arrival times are not shown for $\beta>86°$ since they were not observed beyond this angle. The dotted line locates direct reflection from the corner *S*. It shows good agreement with Fig. 6(a) in spite of velocity approximations used in calculations.

Figures 7(a) and (b) refer to the 100-mm cylindrical shell for, respectively, experimental results and calculated arrival times of T_0 wave echoes. Two patterns composed of many successive echoes of the T_0 wave are observable in the time window 0–200 μs. Echoes of the first pattern are from waves that have undergone one end reflection during propagation ($r=1$). The second pattern is constituted of echoes due to waves which have undergone three reflections at the cylindrical shell extremities ($r=3$). In the first step, the interval between two successive echoes of a given pattern has been described for the longer cylindrical shell. It remains valid here, since it only depends on the external radius of the cylinder. Certainly, the number of echoes of each pattern decreases with the length of the cylindrical shell. The difference with the case of the longer cylindrical shell is the appearance of several patterns of echoes whose relative positions evolve with the length of the target. The time interval between patterns is in agreement with the nondetection of echoes during return propagation of the wave (monostatic setup). Let us closely examine this interval between two successive patterns. At a fixed incidence angle β, the distance corresponding to the time interval between the last echo of a pattern and the first echo of the next pattern is illustrated in Fig. 9(b). Path follow-up depicted in Fig. 9(a) is considered. At point *C*, the surface wave is re-emitted in water and received at point C', or continues its travel on the cylinder up to *D*, after four helical laps and two end reflections. The surface wave is then re-emitted and received at D'. The difference in traveled length is given by an integer number of helical laps on the cylinder (four in our case), minus the traveled distance in water (d_2). For a given β, the gap between successive patterns is not constant. It depends on the ratio of the helical step projected on the shell main axis over the length of the shell. Generally, this ratio is not an integer; so the outlet point of the first echo of the group varies with *r*. Hence, the echo number may differ from one group to another. The top part of Fig. 7(a) is an extract from the bottom part at $\beta=76°$. The two most important echo series are characteristic of the peculiar phenomenon attracting our attention. In Fig. 7(b), the first few echoes of a third pattern related to five end reflections ($r=5$) are predictable, but not observed experimentally. The model yields the echo loci, but does not take their amplitudes into consideration. These amplitudes depend on the range on the line MM' (or NN') over

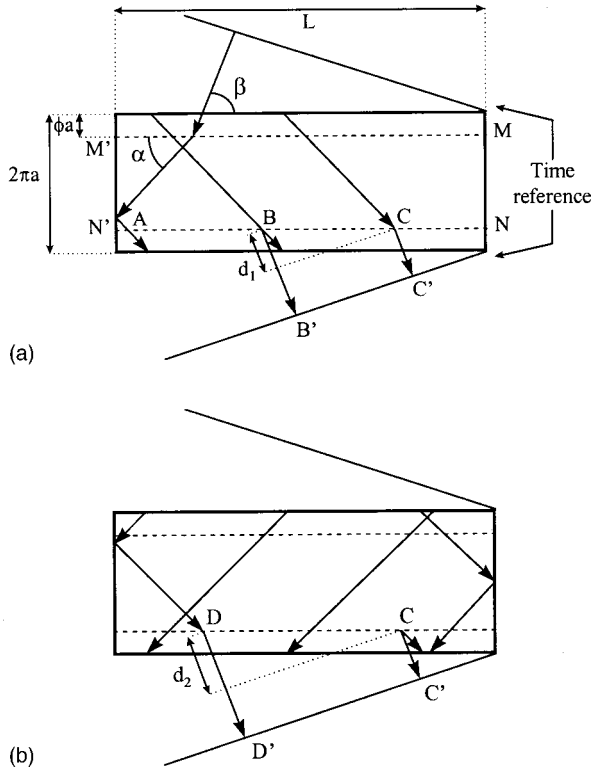


FIG. 9. (a) Topological equivalent of the cylinder: paths with one end reflection. (b) Topological equivalent of the cylinder: paths with one and three end reflections.

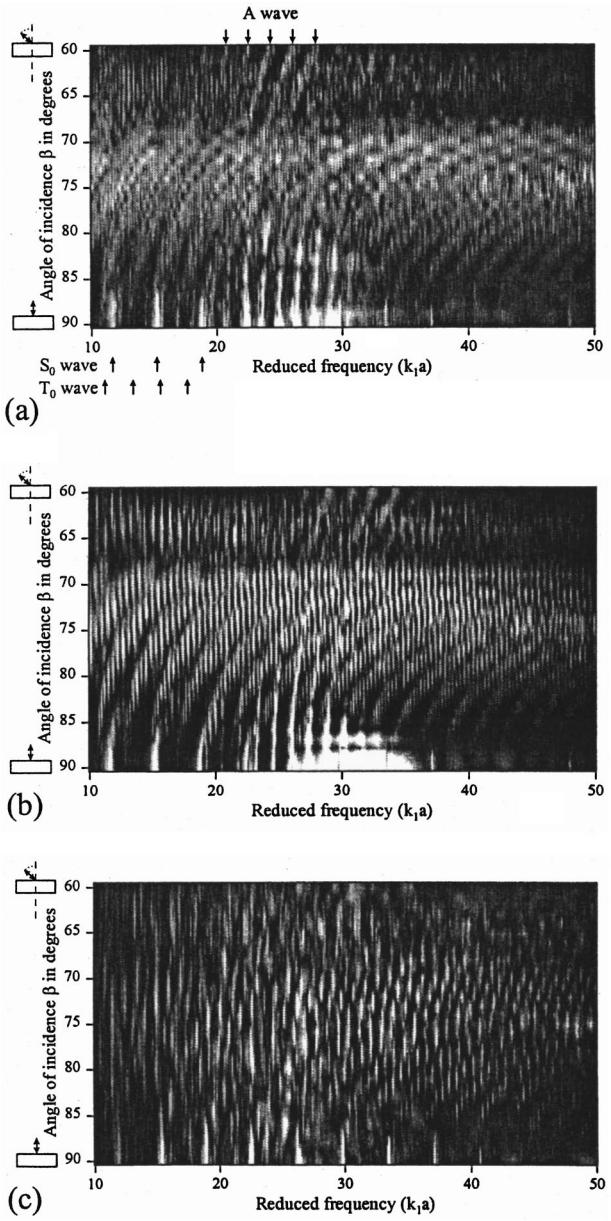


FIG. 10. Evolution of resonance spectra from finite cylindrical shells as a function of incidence angle β: (a) $L=200$ mm; (b) $L=100$ mm; (c) $L=50$ mm. Amplitude increases linearly from black to white zone.

which echoes can be reradiate in water (Fig. A1).

Figure 8(a) is obtained from the shorter cylindrical shell ($L=50$ mm). Three patterns are observable in the time window 200-μs wide. From left to right, the echo pulse series undergo an increasing odd number of reflections ($r=3,5,7$). This identification is provided in Fig. 8(b). The first pattern ($r=1$) is not easily distinguishable experimentally: the visualization is hidden by the presence of A wave echoes. Patterns are closer together than in the case of the longer cylindrical shells. For a given incidence, there are fewer echoes in each pattern since there are lesser helical steps on a shorter cylindrical shell. A recording at $\beta=77°$ shows the 3D representation utility.

In all three cylindrical shells, the presence of echo groupings around the characteristic angle ($\beta_{T0}=62.6°$) is linked to back-and-forth travels of the guided wave along a meridional ray (axial propagation).[8] Moreover, around $\beta=68°$, additional periodically spaced echoes are observable for $70\,\mu s \leqslant t \leqslant 150\,\mu s$ [Fig. 6(a)] and for $50\,\mu s \leqslant t \leqslant 100\,\mu s$ [Fig. 7(a)]. They are marked on the figures by asterisks. The anchoring mechanism is responsible for this.

With the reduction of the length of cylindrical shell, more echo groups appear in the same time window. The number of shell end reflections differentiates them. The geometrical approach clarifies the physical phenomena. In the next section, spectral analysis is carried out.

B. Spectral analysis

The link between helical surface wave propagation and resonance has been shown.[17] At oblique incidence, resonances are due to phase matching of circumnavigating helical surface waves. This observation is applied in the present analysis. When the length of the cylindrical shell decreases, changes in time domain can be used to explain the corresponding spectra modifications. Resonance spectra [Figs. 10(a)–(c)] are obtained from the experimental time values (for, respectively, $L=200$, 100, and 50 mm). The time-signal recordings are treated over a duration of 700 μs, after the specular zone. The observation frequency range extends over 230–1140 kHz ($k_1a=10-50$). The passband of the transducer is not normalized (Fig. 5).

The T_0 wave contributions are observable in two separated areas. In the lower angular zone, the backward and forward paths along a meridional ray are specific to the critical angle, in the range $60° \leqslant \beta \leqslant 65°$. The opening angle of the ultrasonic beam emitted by the transducer explains this

5-deg angular spread. The frequency gap between the resonances doubles when the length of the cylindrical shell is halved [Figs. 10(a)–(c)]. As yet, we do not explain the gap on the angular corridor ($65° \leqslant \beta \leqslant 68°$); however, this angular band coincides approximately with the surface wave paths that do not close in because of helical angles that are too large. For incidence angles beyond this limit, the helical behavior of the T_0 wave appears to predominate. In the case of the longest cylindrical shell [Fig. 10(a)], the shift of modes corresponding to the infinite cylindrical shell as a function of the angle is observable (Fig. 3). Circumferential modes are split up here in many resonances that are close together. Hence, the modes appear as streaks. In the source time range of 700 μs, numerous echoes belonging to four different patterns are considered. The splitting of modes is caused by the breaking of periodicity in echo arrangement. Figure 10(b) shows resonance spectra from the 100-mm-long cylindrical shell. The T_0 wave modes include two distinct features. First, the frequency shift of the infinite cylindrical shell modes with the angle of incidence stills visible: the modes are widened. Second, each mode is split up so that, for a given incidence angle, several separate resonances with the same mode n of circumferential vibration are observable. The relevant time plots contain echoes from eight patterns. Each pattern is made up of fewer echoes than in the previous case ($L = 200$ mm). So, the influence of surface wave helical propagation diminishes. The spectra modification continues when the length of the cylindrical shell decreases. In Fig. 10(c) ($L = 50$ mm), the organization of the modes is not as apparent as for the previous longer cylindrical shells. Many more irregularly spaced patterns with only few echoes each take place in the time recordings. The disorganized appearance in Fig. 10(c) comes essentially from the frequency interval between successive peaks of a circumferential split mode. At a given incidence angle, resonances of different circumferential modes can be mixed together.

The splitting phenomenon is illustrated in Fig. 11 through the comparison of the resonance spectra at the incidence angle $\beta = 70°$. Figure 11(a) is the calculated spectrum for the infinite cylindrical shell; it is extracted from Fig. 3. The three other spectra [Figs. 11(b)–(d)] are extracted from the 3D views [Figs. 10(a)–(c), respectively]. These spectra are shown in expanded form in the range $30 \leqslant k_1 a \leqslant 50$, on one hand, and beyond significant A wave contributions, on the other hand (many peaks still appear between $30 \leqslant k_1 a \leqslant 35$). In the case of the infinite cylindrical shell, each mode is associated with one peak only [Fig. 11(a)]. For finite cylindrical shells, many resonance peaks may be associated with the same circumferential vibration mode [Figs. 11(b)–(c)]. This is not indicated on the two spectra, since no experimental identification of modes n was carried out. Resonances due to mode splitting are fewer and more spaced out the shorter the cylindrical shell is. The last spectrum [Fig. 11(d)] has a new look: the envelope of infinite cylindrical shell modes is not recovered anymore. It becomes difficult to associate peaks to corresponding circumferential modes because of the continuum aspect of the resonance spectrum.

A significant frequency shift of the envelope is observed in Fig. 11(b). This is due to the experimental adjustment of the incidence angle on the shell. Away from normal incidence ($\beta = 70°$), the resonance frequency position is more sensitive to small variations of the incidence angle.

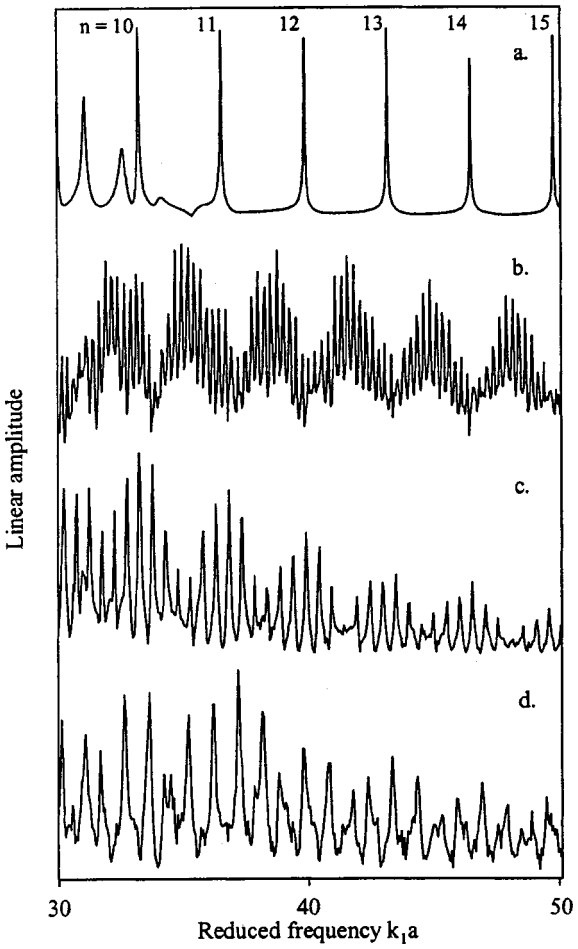


FIG. 11. A comparison of resonance spectra from the three finite cylindrical shells at incidence angle $\beta = 70°$; (a) calculated results for infinite cylindrical shell, (b)–(d) experimental results from the finite cylindrical shells (respectively, $L = 200$, 100, 50 mm).

V. CONCLUSION

This study, backed by a geometrical expansion, shows both time and frequency experimental responses, of scattering from immersed finite cylindrical shells excited at oblique incidences. The first guided wave T_0 is investigated. In the time domain, the reduction of the length of the cylindrical shell is noted through the presence of echo series succession. The T_0 wave follows helical paths on the shell. Reradiations are received when the path includes odd numbers of reflections at the extremities. The periodicity in any given pattern is associated with the helical way of wave travel; the time interval between two patterns is related to the cylindrical shell length. This makes clear the observed spectral feature of a finite cylindrical shell: a splitting of circumferential modes takes place. Thus, for a given angle of incidence, several resonances can be associated with the same circumferential vibration mode of the first guided wave. To observe such peak multiplication, related to a maintained propagation system along the length of the cylindrical shell, the processed time window has to contain several echo patterns.

ACKNOWLEDGMENT

The authors are grateful to Dr. Jeremiah Chiumia for his valuable assistance.

APPENDIX

The following geometrical relations are valid for each wave that adopts a guided behavior. They are written for the T_0 wave since it is exclusively dealt with in the experimental part of this paper. The T_0 guided wave is generated under oblique incidence. Then, $\mathbf{k}_1$ (wave vector of the incident acoustic beam in water) and $\mathbf{n}$ (external normal vector to the cylindrical shell) form the characteristic angle θ_{T_0} ($\sin\theta_{T_0} = C_1/C_{T_0}^{ph}$, C_1 is the sound velocity in water). Two generating lines MM' and NN' (Fig. A1), parallel to the cylinder axis, are thus described. ϕ is the opening half angle between the two generating lines, with regard to the cylinder axis. The incidence angle β and the helix angle α are measured from the cylinder axis. Two relations between these angles are as follow:

$$\cos\alpha = \cos\beta/\sin\theta_{T_0}, \tag{A1}$$

$$\cos\phi = \cos\theta_{T_0}/\sin\beta. \tag{A2}$$

The model validity range is over $\pi/2 - \theta_{T_0} < \beta < \pi/2$.

So, in (A1) we have $0 < \alpha < \pi/2$, whereas in (A2) we have $0 < \phi < \theta_{T_0}$.

The echo arrival times, t_q, are given by

$$t_q = S_{cq}/C_{T_0}^g + S_{wq}/C_1, \quad q = 1,2,..., \tag{A3}$$

S_{cq} is the distance traveled on the cylinder, S_{wq} refers to that in water

$$S_{cq} = 2(q\pi - \phi)a/\sin\alpha, \quad q = 1,2,..., \tag{A4}$$

q replaces index n. It still refers to the number of complete helical laps made by the propagating T_0 wave, independent of the number of end reflections r. More precisely, the first lap connecting the two generating lines is not complete.

$$S_{wq} = 2a(1-\cos\phi)\sin\beta + [L(r+1) - 2h_q]\cos\beta$$

$$q = 1,2,..., \quad r = 1,3,5,.... \tag{A5}$$

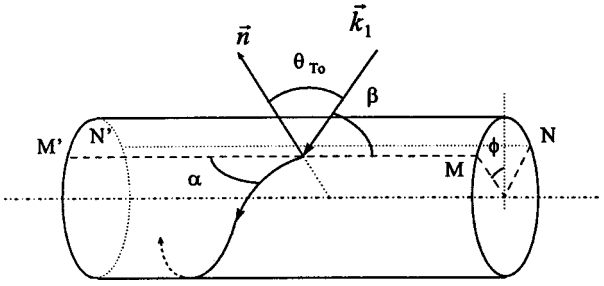


FIG. A1. Sketch indicating angles.

Relation (6) of the original issue is extended here to paths involving end reflections. Index r is odd because of the way echoes are received (monostatic setup). The projection on the cylinder's axis of the halfway path on the cylindrical shell Δh_n (Ref. 14) is noted h_q

$$h_q = (q\pi - \phi)a/\tan\alpha, \quad q = 1,2,.... \tag{A6}$$

The echoes are received when the surface wave covers a distance between an odd number of end reflections to an even number of end reflections. The number of helical outgoings (reradiation in the surrounding medium) is determined for each incidence angle by

$$h_1 \leqslant h_q \leqslant L, \quad r = 1, \quad q = 1,2,...,$$

$$(r-1)L \leqslant 2h_q \leqslant (r+1)L, \quad r = 3,5,..., \quad q = 1,2,....$$

The experimental setup leads to a change in reference point which no longer considers the edge of the cylindrical shell. Rather, the geometrical rotational center is considered in this case.

[1] H. Überall, L. R. Dragonette, and L. Flax, ''Relation between creeping waves and normal modes of vibration of a curved body,'' J. Acoust. Soc. Am. **61**, 711–715 (1977).
[2] L. Flax, V. K. Varadan, and V. V. Varadan, ''Scattering of an obliquely incident acoustic wave by an infinite cylinder,'' J. Acoust. Soc. Am. **68**, 1832–1835 (1980).
[3] X. L. Bao, H. Cao, and H. Überall, ''Resonances and surface waves in the scattering of an obliquely incident acoustic field by an infinite elastic cylinder,'' J. Acoust. Soc. Am. **87**, 106–110 (1990).
[4] N. D. Veksler, *Resonance Acoustic Spectroscopy*, Springer Series on Wave Phenomena **11**, edited by L. M. Brekhovskikh (Springer, Berlin, 1993).
[5] D. Brill and G. C. Gaunaurd, ''High-frequency monostatic echoes from finite-length cylinder,'' Proceedings of the 1st French Conference on Acoustics, Lyon, Vol. C-2, 383–386 (1990).
[6] A. Harari, B. E. Sandman, and J. A. Zaldonis, ''Analytical and experimental determination of the vibration and pressure radiation from a submerged, stiffened cylindrical shell with two end plates,'' J. Acoust. Soc. Am. **95**, 3360–3368 (1994).
[7] S. F. Morse, P. L. Marston, and G. Kaduchak, ''High-frequency backscattering enhancements by thick finite cylindrical shells in water at oblique incidence: Experiments, interpretation, and calculations,'' J. Acoust. Soc. Am. **103**, 785–794 (1998).
[8] S. F. Morse and P. L. Marston, ''Meridional ray contributions to scattering by tilted cylindrical shells above the coincidence frequency: Ray theory and computations,'' J. Acoust. Soc. Am. **106**, 2595–2600 (1999).
[9] G. Kaduchak, C. M. Wassmuth, and C. M. Loeffler, ''Elastic wave contributions in high resolution acoustic images of fluid-filled, finite cylindrical shells in water,'' J. Acoust. Soc. Am. **100**, 64–71 (1996).
[10] K. Gipson and P. L. Marston, ''Backscattering enhancements due to reflection of meridional leaky Rayleigh waves at the blunt truncation of a tilted solid cylinder in water: Observations and theory,'' J. Acoust. Soc. Am. **106**, 1673–1680 (1999).
[11] M. L. Rumerman, ''Contribution of membrane wave reradiation to scattering from finite cylindrical steel shells in water,'' J. Acoust. Soc. Am. **93**, 55–65 (1993).
[12] F. Lecroq, F. Léon, D. Décultot, and G. Maze, ''Diffusion d'une onde acoustique par un tube limité par des disques plans: études théoriques et expérimentales,'' Acustica **74**, 51–62 (1991).
[13] H. Überall, ''Helical surface waves on cylinders and cylindrical cavities,'' Trait. Signal **2**, 381–385 (1985).
[14] X.-L. Bao, ''Echoes and helical surface waves on finite elastic cylinder excited by sound pulses in water,'' J. Acoust. Soc. Am. **94**, 1461–1466 (1993).
[15] F. Léon, F. Lecroq, D. Décultot, and G. Maze, ''Scattering of an obliquely incident acoustic wave by an infinite hollow cylindrical shell,'' J. Acoust. Soc. Am. **91**, 1388–1397 (1992).

[16] J. L. Izbicki, G. Maze, and J. Ripoche, ''Influence of the free modes of vibration on the acoustic scattering of a circular cylindrical shell,'' J. Acoust. Soc. Am. **80**, 1215–1219 (1986).
[17] J. M. Conoir, P. Rembert, O. Lenoir, and J. L. Izbicki, ''Relation between surface helical waves and elastic cylinder resonances,'' J. Acoust. Soc. Am. **93**, 1300–1308 (1993).
[18] G. Maze and J. Ripoche, ''Méthode d'Isolement et d'Identification des Résonances (M.I.I.R) de cylindres et de tubes soumis à une onde acoustique plane dans l'eau,'' Rev. Phys. Appl. **18**, 319–326 (1983).
[19] P. Pareige, P. Rembert, J. L. Izbicki, G. Maze, and J. Ripoche, ''Méthode impulsionnelle numérisée (MIN) pour l'isolement et l'identification des résonances de tubes immergés,'' Phys. Lett. A **135**, 143–146 (1989).

Incorporation of loudness measures in active noise control

Scott D. Sommerfeldt[a] and Timothy O. Samuels
Department of Physics, Brigham Young University, Provo, Utah 84602

(Received 24 December 1999; revised 6 September 2000; accepted 13 November 2000)

An attempt has been made to use a modified version of a standard active noise control algorithm in order to take into account the unique response of the human auditory system. It has been shown in the past that decreasing the sound pressure level at a location does not guarantee a similar decrease in the perceived loudness at that location. Typically, active noise control is based on minimizing the ''error signal'' from a mechanical device such as a microphone, whose response is nominally flat across the frequency response range of the human ear. However, if the response of the ear can be approximated by digitally filtering the error signal before it reaches the adaptive controller, one can, in effect, minimize the more subjective loudness level, as opposed to the sound pressure level. The work reported here entails simulating active noise control based upon minimizing perceived loudness for a collection of input noise signals. A comparison of the loudness of the resulting error signal is made to the loudness of that resulting from standard sound pressure level minimization. It has been found that the effectiveness of this technique is largely dependent upon the nature of the input noise signal. Furthermore, this technique is judged to be worth considering for use with applications of active noise control where the uncontrolled noise more prominently constitutes low range audio frequencies (approximately 30 Hz–100 Hz) than medium range audio frequencies (approximately 300 Hz–600 Hz).
[DOI: 10.1121/1.1339824]

PACS numbers: 43.50.Ki, 43.66.Cb [MRS]

I. INTRODUCTION

As the field of active noise control has developed, numerous applications and issues associated with active control have been investigated. However, in all of this development, there is a potentially important area that has been largely ignored. This area involves the issue of how human beings respond to the controlled field.

In implementing an active noise control system, one must of necessity use some form of sensor to obtain information regarding the acoustic field. Originally, this consisted of nothing more than a simple microphone to detect the acoustic pressure. As it became apparent that simply using a microphone could lead to undesirable results, such as localized control, other techniques began to be developed in an attempt to achieve a ''better'' solution. Such techniques include using multiple microphones to obtain a more global response,[1] ''energy density sensors'' to try to avoid local minima,[2] modal sensors to control dominant modes in the field,[3] intensity sensors to minimize propagation in a certain direction,[4] and radiation mode sensors to minimize acoustic radiation with a minimal number of structural sensors.[5] While all of these methods have been shown to have certain advantages for various applications, it has also been noted that it is often possible to implement an active control system where the performance function is attenuated substantially, but in which human observers have noted little difference between the uncontrolled and controlled states. In other words, although the mechanical sensor used in the control system detects a substantial reduction, the human ear as a sensor detects little, if any, reduction. For applications involving only stealth and/or detection, such as military applications, the response of the human ear may be of little interest. However, there are many applications where the objective of the control system is to achieve a ''better'' acoustic environment for human beings. Such examples include interior aircraft and automobile noise, active control of transformer noise, active control of highway noise, and so forth. In such applications, the only really important criterion is the human perception of the control achieved.

Recently, there has been some interest shown in addressing the question of improving an active control system in terms of human response. Saunders and Vaudrey[6] showed that a signal with significant reduction in the controlled sound pressure level may only exhibit a very modest reduction in the perceived loudness of the signal, as determined using Zwicker's method.[7] Thus their work suggested that an alternative approach to active control of sound, which could effectively reduce the loudness perceived by listeners could be useful.

This work has focused on implementing a technique that approximates the minimization of loudness by an active control system, as a means of investigating the anticipated effectiveness of such an approach. The approach taken makes use of the filtered-E algorithm, as developed by Kuo and Tsai.[8,9] Kuo and Tsai present the filtered-E algorithm as a means of altering the spectral shape of the residual noise. While they indicate the approach could be implemented based on the frequency response of the human ear, they do not investigate this possibility further. The work reported here provides a practical implementation of the filtered-E algorithm based on the response of the human ear, and evaluates how effective the method is in minimizing the perceived loudness of the

[a]Electronic mail: s_sommerfeldt@byu.edu

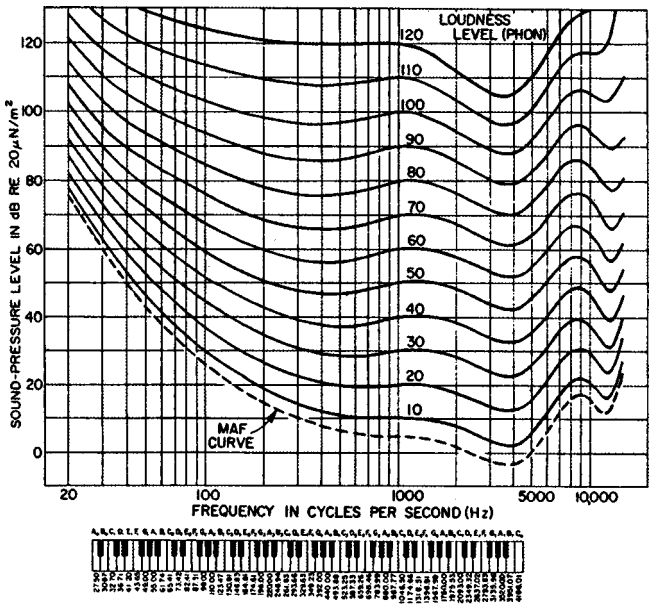


FIG. 1. Free-field equal loudness contours for pure tones, determined by Robinson and Dadson in 1956 at the National Physical Laboratory, Teddington, England (ISO/R226-1961). (Taken from Peterson, *Handbook of Noise Measurement*.)

signal. The work reported is numerical in nature, with the loudness of the controlled signals being compared for different minimization techniques as a means of judging the improvement that would be perceived by a listener.

II. DEVELOPMENT OF LOUDNESS

Loudness is a quantitative measure that is based on the subjective response of human listeners. As such, it appears that loudness could perhaps be an ideal performance function for implementing active noise control when human perception is involved. However, it is not always straightforward to calculate, and thus is difficult to implement directly in an active control system. As a result, an approximation to minimizing loudness was implemented in the active control system that was simulated for this work. However, the calculation of loudness for the various signals was still used as an analysis tool, to give an indication how well the ''loudness-based'' control could be expected to perform.

The concept of loudness was developed as a means of quantifying the manner in which the human ear responds to sound. It is well known that the human ear responds differently to the same sound pressure level presented to the listener at different frequencies. Research in this area has resulted in the development of Equal Loudness Contours, which represent the ear response over the range of human auditory sensitivity (20 Hz–20 kHz), and which can be seen in Fig. 1.

The loudness level is measured in *phons* and, like sound pressure level, is rated on a logarithmic scale. Each of the curves in Fig. 1 corresponds to a different loudness level, or phon level, chosen to be identical to the sound pressure level of the curve at 1000 Hz. Each single curve shows the sound pressure level at which various frequencies must occur in order to sound equally loud to a listener. For example, a 30-Hz sound must be at a level of about 80 dB *re*: 20 μPa to sound as loud as a 1000-Hz sound at a level of 40 dB. It is apparent that the human ear is less sensitive to low-frequency sound than to mid-frequency sound.

A distinction should be made at this point between loudness and loudness level. The loudness level is measured in phons on a logarithmic scale. Loudness is measured in *sones*, and corresponds to a linear scale. Because the scale of sones is linear, a noise signal that is measured to have one-half the number of sones as another is said to be half as loud.[10] It is possible to convert between loudness and loudness level, just as one may convert between sound pressure level (logarithmic scale) and sound pressure (linear scale).

The Stevens Mark VII procedure was used for calculating the loudness of the signals. This method utilizes all of the third-octave band levels of the signals. The corresponding perceived magnitude, S, in sones, for each band is found using a comprehensive table of values tabulated by Stevens. From the maximum, S_m, of these corresponding values a factor, F, is located in an additional table of values. The next step is to add all of the perceived magnitudes for the third-octave bands together (ΣS), and then subtract from that sum the maximum, S_m. This difference must then be multiplied by the factor F, and this product added to the maximum perceived magnitude, S_m.[11] After accomplishing these steps, the total perceived magnitude, S_t, in sones, has been determined. The preceding steps may be written mathematically as follows:

$$S_t=\left[\left(\sum S\right)-S_m\right]F+S_m. \tag{1}$$

After algebraic manipulation, this equation may be written as

$$S_t=(1-F)S_m+F\sum S, \tag{2}$$

which is the form of the equation used to calculate loudness, in sones, for this project. The value of the perceived magnitude in sones may be converted to a perceived level in decibels, if desired, by consulting an additional table developed by Stevens.

III. LOUDNESS-BASED ACTIVE CONTROL

Because there is no direct method of calculating loudness for all acoustic signals, there arises the question of how to minimize loudness with an ANC system. It is in the calculation of loudness that the Equal Loudness Contours (ELCs) become quite significant. Similar sound pressure levels at different frequencies correspond to different loudness levels (measured in phons). The difference in the loudness levels between two frequencies of similar sound pressure level is also dependent upon the sound pressure level itself; there is a trend (not without exception) such that the greater the sound pressure level, the smaller the difference in loudness levels between two frequencies of similar sound pressure level.

There already exist algorithms designed to minimize the signal received by an error microphone (namely the overall sound pressure level) in an ANC system. The filtered-x LMS

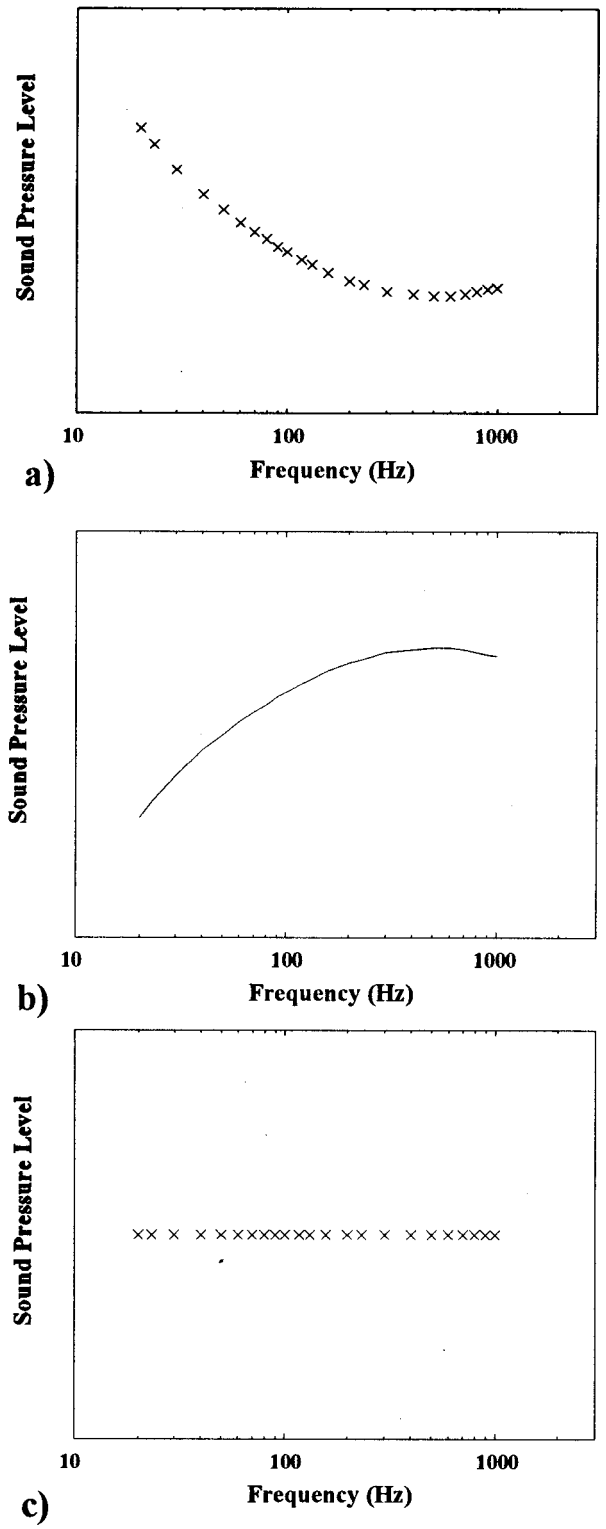


FIG. 2. Filtering operation for minimizing loudness. (a) Sample noise signal, showing relative third-octave band levels. (b) Loudness-based filter. (c) Filtered signal, normalized for loudness.

algorithm[12] is the most widely used of these algorithms. As suggested previously,[8] the error signal can be passed through a ''residual noise shaping filter,'' and then this altered signal may be used as the signal minimized by the ANC system. If this filter can be designed to have exactly the inverse shape of a given ELC, then the minimization that results will approximate the minimization of loudness associated with the

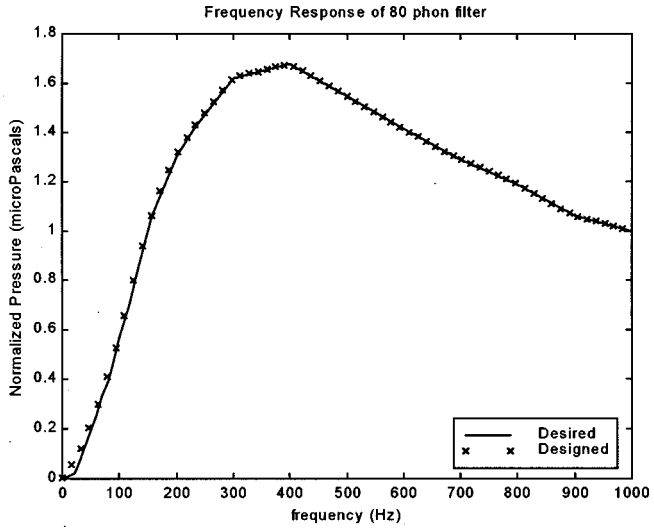


FIG. 3. 80-phon loudness-based filter. Solid line is desired filter shape. ×'s show designed filter shape.

signal. For example, since the filter is of a shape inverse to that of an ELC, then if the original signal happens to have the shape, in frequency space, of the noninverted ELC, then the filtered signal would look *flat* in frequency space. The difference between this signal and the original signal is that the levels of the frequencies in this filtered signal, which is flat in frequency space, correspond to loudness levels, as opposed to sound pressure levels. With this filtering, different frequencies are essentially normalized with respect to loudness level, and by this method, minimizing the new filtered signal by a standard method (i.e., the filtered-*x* LMS algorithm) has the effect of approaching loudness minimization of the original signal. (Note: The original signal need not have the shape in frequency space of the ELC filter used in order to attain the desired effect of filtering. This case was simply used as an example.) Figure 2 presents a simple example intended to make this concept clear. Figure 2(a) presents the third-octave band levels of a hypothetical noise signal, which has been chosen to match an ELC for the sake of illustration. Figure 2(b) indicates the loudness-based filter that can be used to properly normalize the noise spectrum. Figure 2(c) then shows the resulting filtered noise signal that could be used in an active control system designed to approximately minimize the loudness. Notice that in this example, the lower frequencies of high sound pressure level become much less significant after filtering, because the human ear is less sensitive to noise in this frequency range. For these plots sound pressure level is arbitrary.

The loudness-based filters used in the research were obtained in the following manner. The values associated with each ELC were determined, for the range of 10 phon to 120 phon, in 10-phon increments. These values, which were on a logarithmic scale, were multiplied by -1 to invert the contour, and then converted to a linear scale. The finite impulse response (FIR) loudness-based filter was then obtained using the MATLAB function ''yulewalk.'' This function uses the frequency and modulus values to obtain the FIR filter which most closely approximates the desired magnitude response. As an example, Fig. 3 shows the filter response obtained for

the 80-phon loudness-based filter. A similar fit was obtained for each of the other ELCs used.

While ideally all frequencies would be attenuated to an imperceptible level, an active noise control system is always limited in performance. Suppose the signal filtered by the loudness-based filter is in fact the error signal used to update the controller parameters. The controller will in this case not utilize as many resources attenuating the high level, low-frequency content of the original noise signal as it would have, had the signal remained unfiltered.

In order to implement this technique with the LMS filtered-x control algorithm, the error signal, as well as the reference signal, must be filtered by a loudness-based filter, A, before it is used to update the controller parameters. Figure 4 shows schematically the control system in which the error signal is filtered for loudness-based control. In this block diagram, P represents the transfer function of the system to be controlled, W represents the controller transfer function, H represents the control path transfer function, $\mathcal{H}$ represents the model of H, and A represents the transfer function of the loudness-based filter. Notice that the filtered-x signal, $r(t)$, as well as the error signal, $e(t)$, is filtered by the loudness-based filter.

It is not immediately obvious that the filtered-x signal, $r(t)$, needs to be filtered by A as well, in order for the controller update scheme to work properly. However, this requirement can be illustrated mathematically in a relatively straightforward manner. It can be seen that the error signal can be expressed as

$$e(t)=d(t)+y(t). \tag{3}$$

Also, the convolution of the filter coefficients with the input signal can be represented as

$$\sum_{m=0}^{L} w_m x(t-m)=\mathbf{W}^T\mathbf{X}, \tag{4}$$

where $\mathbf{W}$ is a vector of the filter coefficients, $\mathbf{X}$ is a vector of the input data samples, and $L+1$ is the number of coefficients in the control filter. Examining the schematic in Fig. 4, it can be seen that

$$\begin{aligned} e_f(t)&=\mathbf{E}^T\mathbf{A}=\mathbf{A}^T\mathbf{E}\\ &=\sum_{i=0}^{I} a_i e(t-i)\\ &=\sum_{i=0}^{I} a_i[d(t-i)+y(t-i)]\\ &=d_f(t)+\sum_{i=0}^{I} a_i\sum_{j=0}^{J} h_j u(t-i-j)\\ &=d_f(t)+\sum_{i=0}^{I} a_i\sum_{j=0}^{J} h_j\sum_{l=0}^{L} w_l x(t-i-j-l)\\ &=d_f(t)+\sum_{i=0}^{I} a_i\sum_{l=0}^{L} w_l\sum_{j=0}^{J} h_j x(t-i-j-l). \end{aligned} \tag{5}$$

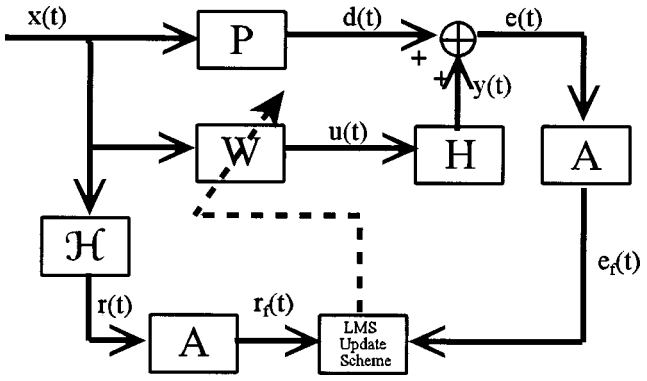


FIG. 4. Filtered-x scheme with loudness-based filters for $e(t)$ and $r(t)$.

Notice here that $\Sigma h_j x(t-i-j-l)$ is simply the filtered-x signal, $r(t-i-l)$, or in other words the noise signal, $x(t)$, after being filtered by the actuator/propagation path transfer function, H. Therefore, it follows that

$$\begin{aligned} e_f(t)&=d_f(t)+\sum_{i=0}^{I} a_i\sum_{l=0}^{L} w_l r(t-i-l)\\ &=d_f(t)+\sum_{l=0}^{L} w_l\sum_{i=0}^{I} a_i r(t-i-l). \end{aligned} \tag{6}$$

Because $\Sigma a_i r(t-i-l)=r_f(t-l)$, where $\mathbf{R}_f=\mathbf{R}^T\mathbf{A}=\mathbf{A}^T\mathbf{R}$, it follows that

$$\begin{aligned} e_f(t)&=d_f(t)+\sum_{l=0}^{L} w_l r_f(t-l)\\ &=d_f(t)+\mathbf{W}^T(t)\mathbf{R}_f(t). \end{aligned} \tag{7}$$

From this representation of $e_f(t)$, the new controller update equation can be formulated using standard minimization techniques as

$$\mathbf{W}(t+1)=\mathbf{W}(t)-\mu e_f(t)\mathbf{R}_f(t). \tag{8}$$

This is the update equation used in the simulations to investigate the effectiveness of using loudness-based active noise control.

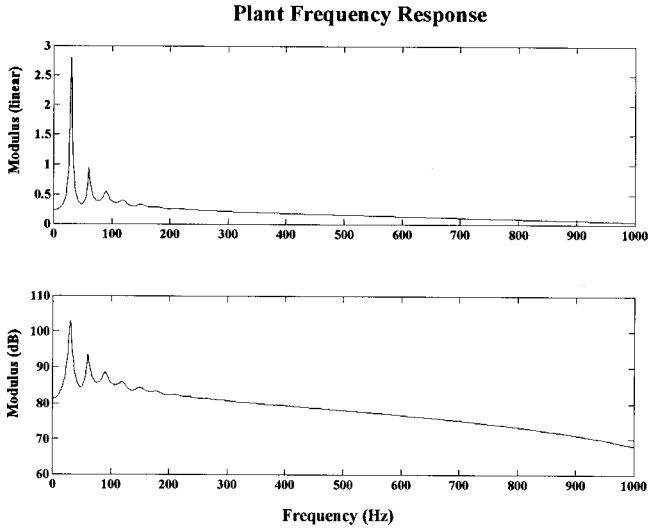


FIG. 5. Response in frequency space of the plant, P.

TABLE I. Best case results for SIM1, SIM2, and SIM3.

	Random noise input					
	SIM1—10 coefficients		SIM2—30 coefficients		SIM3—60 coefficients	
	L_p (dB)	Loudness (sones)	L_p (dB)	Loudness (sones)	L_p (dB)	Loudness (sones)
$d(t)$	76.9	17.0	77.4	17.1	77.1	16.9
$e(t)$	76.1	14.8	76.3	15.0	77.0	15.4
$e2(t)$	73.8	15.2	74.1	15.6	73.5	15.8

To investigate the effectiveness of the new approach, simulations were run using both loudness-based active control, and standard minimization of the squared error (pressure) signal. The plant frequency response, representing the system to be controlled, used in this work can be seen in Fig. 5. Following the simulation, the plots of the third-octave band levels superposed upon the ELCs are examined to visualize the effect of loudness control versus overall sound pressure control. Even more importantly, the actual calculated values of sound pressure level and loudness are compared for the three signals of interest, namely the two error signals based upon both methods of control, and the uncontrolled noise signal, $d(t)$. After the plots and quantitative values are examined, an assessment is made as to the apparent effectiveness of loudness control. This step may at times be somewhat difficult, due to the subjective nature of the results, although it seems clear in many cases that the difference in loudness between signals is quite significant. The simulations and assessments are repeated for a variety of input noise signals, as well as for different loudness-based filters, and for a varied number of controller coefficients. It should be noted that while the ELCs are utilized for loudness control, they do not explicitly take part in the loudness calculations. Furthermore, only the loudness calculations are used to evaluate the effectiveness of this technique.

IV. RESULTS

Throughout these comments, $d(t)$ signifies the uncontrolled noise signal as theoretically detected by the error microphone, $e(t)$ signifies the steady state error signal resulting from loudness-based control, and $e2(t)$ signifies the steady state error signal resulting from standard sound pressure level control. Furthermore, W represents the controller transfer function after loudness-based control, and $W2$ represents the controller transfer function after sound pressure level control. The transfer function used for the plant, P, nominally corresponds to a duct, and can be seen in Fig. 5. The focus of this research was on comparing minimization using the standard filtered-x algorithm with loudness-based control using the filtered-E algorithm. As a result, a simple model for the secondary path transfer function was chosen, consisting of a simple delay and gain factor $(0.99\,z^{-1})$. The convergence parameter, μ, was kept fixed at a value of 0.001 for these simulations.

For each of the simulations run, the information obtained consists of: plots of the two error signals, namely $e(t)$ and $e2(t)$, a plot of $d(t)$, FFTs of these three signals as well as of the input signal, $x(t)$, plots of the two controller transfer functions (based on the values of the controller coefficients), namely W and $W2$, and finally a single plot of the ELCs with the third-octave band levels of $d(t)$, $e(t)$, and $e2(t)$ superposed upon them. This final plot provides a visual conception of the difference between the two types of control, namely loudness-based control and overall sound pressure level control.

The first simulation that was run utilized purely random noise, generated by the ''rand'' function in MATLAB. The amplitude of random input was chosen such that $d(t)$, the

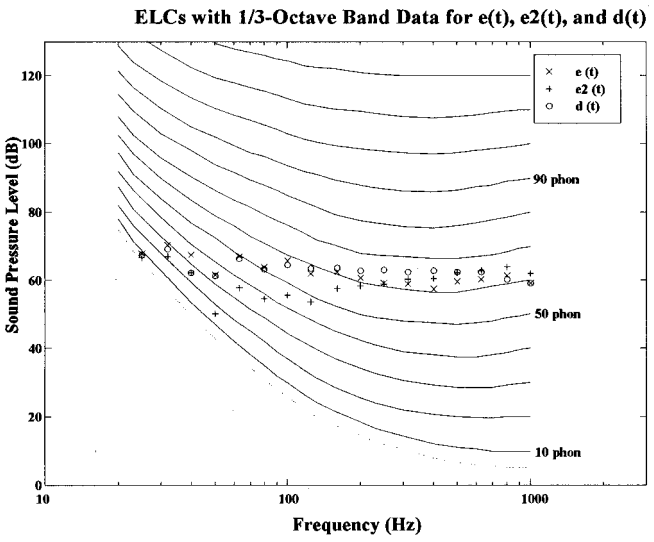


FIG. 6. Third-octave band levels of $d(t)$, $e(t)$, and $e2(t)$ for SIM1.

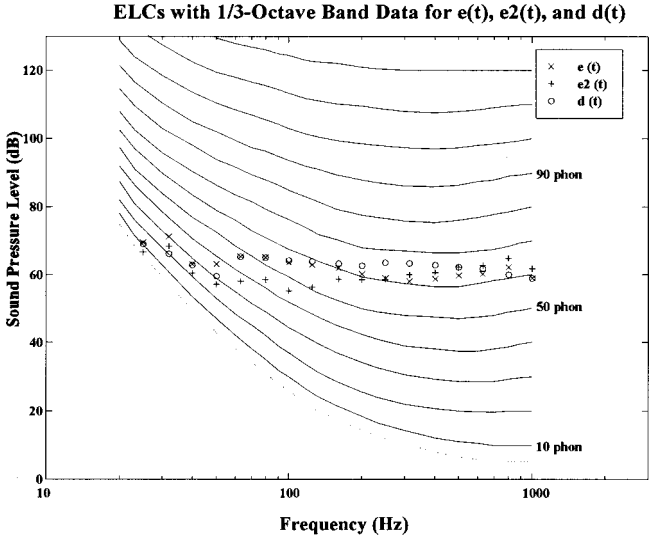


FIG. 7. Third-octave band levels of $d(t)$, $e(t)$, and $e2(t)$ for SIM2.

signal reaching the error microphone before control, constituted an overall sound pressure level of between 70 and 80 dB. Utilizing 10 controller coefficients (SIM1), this input was minimized with respect to overall sound pressure level and to loudness, using each of the 12 different phon level filters. Similar trials were run utilizing 30 controller coefficients (SIM2), and 60 controller coefficients (SIM3). The best results obtained for each of the three cases, including overall sound pressure levels and loudness for all signals, are found in Table I. These results correspond to using the 80-, 90-, and 70-phon curves respectively, which correspond closely with the level of the noise being controlled. For these results, the best case was defined as the case where the largest difference exists between the loudness of $e(t)$ and that of $e2(t)$.

Upon examination of these numerical results, little difference is found for any of the cases run with purely random input, between either the overall sound pressure level or the loudness of the two error signals. It is interesting to note, however, that in all of these cases, the overall sound pressure level of $e(t)$ is *higher* than that of $e2(t)$, while the loudness of $e(t)$ is *lower* than that of $e2(t)$. (This trend will greatly magnify in upcoming simulations, where the input signal is no longer purely random noise.) For the simulation data, the most informative plots were those showing the third-octave band levels for the uncontrolled signal, $d(t)$, as well as the levels for the loudness-based and sound pressure level control. These plots for these three random noise cases can be seen in Figs. 6–8. For this signal, it can be seen that the loudness is primarily affected by the response above 200 Hz. The loudness-based control generally provides slightly better attenuation in this frequency region, and the result is a slight improvement in the loudness of the controlled signal.

The next set of simulations (SIM4) consists simply of four sinusoids as input. It was hoped that some input signal could be created which would provide significant differences between the two methods of control, irregardless of how realistic the input signal might be. Two sinusoids were chosen at low frequencies, specifically 30 Hz and 36 Hz, where the human ear is fairly insensitive. The other two sinusoids were at frequencies of much more significant response with respect to the human auditory system, specifically 350 Hz and 400 Hz. The lower-frequency sinusoids were also chosen to be of higher amplitudes than those at the higher frequencies. This input signal was minimized utilizing 6 controller coefficients, and each of the 12 loudness-based filters. Table II presents the results for the single best case for this input signal.

TABLE II. Best case results for SIM4.

	Four sinusoids input SIM4—6 coefficients	
	L_p (dB)	Loudness (sones)
$d(t)$	71.5	2.53
$e(t)$	66.1	0.92
$e2(t)$	65.0	2.43

The results of these simulations are dramatic, especially in the difference between the loudness of the two error signals. In every case, the loudness of $e(t)$ was *significantly* lower than that of $e2(t)$. In fact, the loudness of $e(t)$ is almost always less than one-half the loudness of $e2(t)$. Furthermore, in every trial, the overall sound pressure level of $e(t)$ is *higher* than that of $e2(t)$, although not always significantly. It is interesting to note that while the overall sound pressure level was reduced in $e2(t)$ by 6.5 dB, the loudness was reduced in $e2(t)$ by a mere 4%. While one may expect to perceive a reduction of 6.5 dB, examining loudness indicates that this sound pressure level reduction would most likely be imperceptible. However, while the overall sound pressure level was reduced in $e(t)$ by 5.4 dB (for the best case), it seems very likely that, after examining the change in loudness, the difference between $d(t)$ and $e(t)$ would be very perceptible. The best case here has the loudness reduced by 62% in $e(t)$. This best case has been chosen for graphical display in Fig. 9.

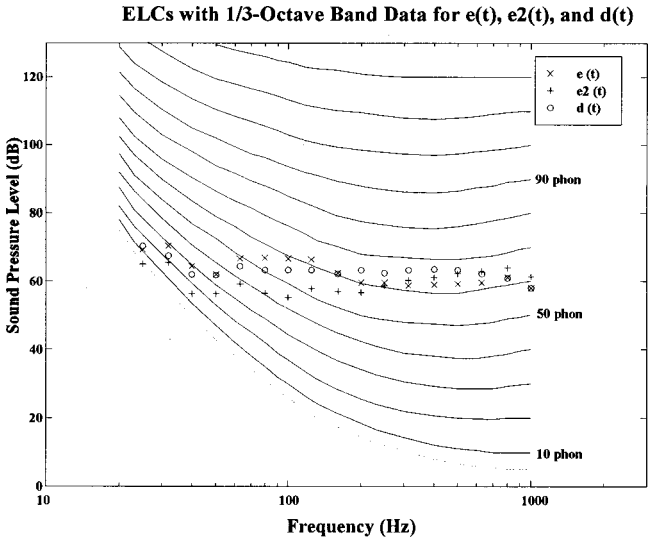


FIG. 8. Third-octave band levels of $d(t)$, $e(t)$, and $e2(t)$ for SIM3.

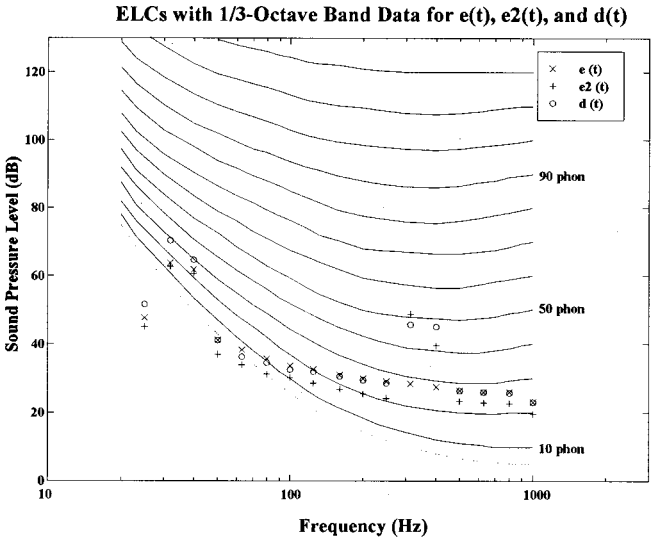


FIG. 9. Third-octave band levels of $d(t)$, $e(t)$, and $e2(t)$ for SIM4.

TABLE III. Input data for SIM7.

Frequencies included							
Freq (Hz)	Mag (Pa)	Freq (Hz)	Mag (Pa)	Freq (Hz)	Mag (Pa)	Freq (Hz)	Mag (pa)
30	0.600	210	0.255	390	0.165	570	0.0825
90	0.525	270	0.225	450	0.135	630	0.060
150	0.240	330	0.195	510	0.105	rand	±0.5

The next set of simulations combines both random and sinusoidal input. The exact frequencies and amplitudes associated with the input signal are shown in Table III. An attempt was made at generating a noise signal similar to that which might arise in a real-life situation. This signal consisted of random noise and a fundamental pure tone (30 Hz) along with odd harmonics of that fundamental tone. Each consecutive harmonic was input at a lower level than the one before (see Fig. 10), which is typical of noise signals in many different situations. Simulations were run with this input signal utilizing both 20 (SIM7) and 100 (SIM8) controller coefficients, as well as with each of the 12 different loudness-based filters. The sound pressure level of the signal $d(t)$ was between 88.5 and 89.1 dB, and the loudness was always between 15.4 and 16.3 sones. With 20 coefficients, $e2(t)$ resulted in a controlled sound pressure level of approximately 10 dB *less* than that of $d(t)$. However, the loudness of $d(t)$ *increased* in $e2(t)$ to between 18.7 and 19.8 sones, which again illustrates that while the sound pressure level has been significantly attenuated, a human observer would very likely perceive an increase in noise level. With 20 coefficients, loudness control resulted in a sound pressure level for $e(t)$ between 1 and 4 dB below that of $d(t)$, depending upon the loudness-based filter used. While these differences may seem insignificant at first, a closer look at loudness reveals some interesting results. In the best case, shown in Table IV, the loudness of $e(t)$ was lower than that of $d(t)$ by 4.2 sones. This loudness difference amounts to a decrease of almost 27%. The results for this specific result can be seen in Fig. 11.

Utilizing 100 coefficients with the same input signal finds that the loudness of $e2(t)$ is significantly lower than it was with 20 coefficients. In fact, the loudness control is slightly better in $e2(t)$ than in $e(t)$, although the difference seems insignificant. The best result here has the sound pressure level decreased in $e2(t)$ by 23 dB, and decreased in $e(t)$ by only 5.2 dB. The loudness, however, is decreased in $e2(t)$ by 6.6 sones (41%), and in $e(t)$ by 5.9 sones (37%). The difference in sound pressure level between $e2(t)$ and $e(t)$ is extreme. However, the difference in loudness is slight, further illustrating the trend that as the number of controller coefficients is increased, the difference between the two methods of control tends to decrease, specifically in regard to loudness. This last result is shown in more detail in Fig. 12.

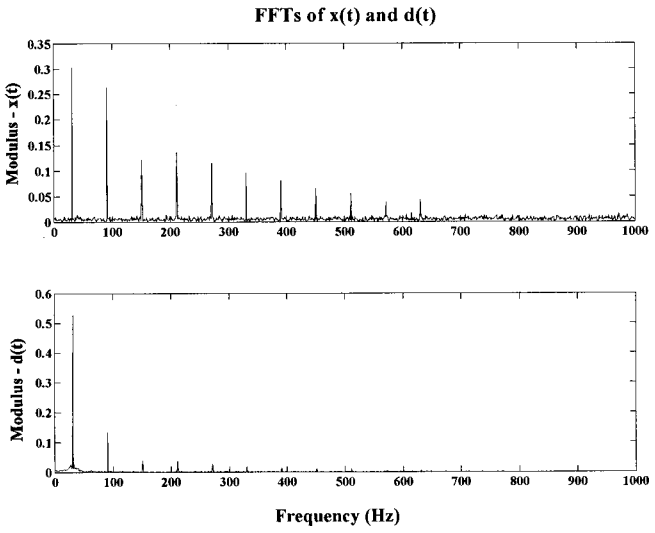


FIG. 10. FFT of $x(t)$ and $d(t)$ for SIM7.

These results are consistent with results that have been found in psychoacoustics regarding the response of the human ear.[13,14] In particular, Hellman and Zwicker[13] have shown that the loudness associated with a 1-kHz tone combined with broadband noise is not correlated with the overall sound pressure level. Some of their results indicated that one can reduce the sound pressure level by 6 dB, while simultaneously doubling the loudness. As can be seen here, similar results can be obtained when trying to minimize the loudness versus the sound pressure level.

The final set of simulations investigated were designed to examine the effect of increasing the number of coefficients made available to the controller. The trend discovered previously was such that as the number of controller coefficients increased, the difference in loudness between the two controlled signals, namely $e(t)$ and $e2(t)$, decreased. The difference in overall sound pressure level, however, tended to remain significant, and therefore the two resulting error signals were not identical, only similar in loudness. The simulations here incorporate the same input signal as in the last set (see SIM7 or SIM8), but this time the 80-phon loudness-based filter was consistently used. The number of controller coefficients was changed from 10 to 100 in increments of 10. Also included were simulations utilizing 150 and 200 coefficients. As expected, as the number of controller coefficients

TABLE IV. Best case results for SIM7 and SIM8.

	Many sines (30 Hz Fund.) plus random input			
	SIM7—20 coefficients		SIM8—100 coefficients	
	L_p (dB)	Loudness (sones)	L_p (dB)	Loudness (sones)
$d(t)$	88.9	15.8	88.7	16.0
$e(t)$	87.5	11.6	83.5	10.1
$e2(t)$	78.2	19.0	65.7	9.40

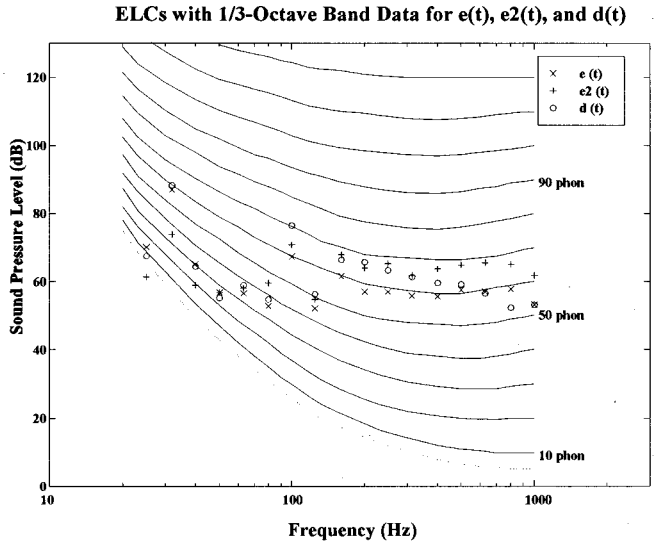


FIG. 11. Third-octave band levels of $d(t)$, $e(t)$, and $e2(t)$ for SIM7.

increased, the difference in loudness between the two error signals decreased significantly. Table V presents the results of these simulations.

The most noticeable difference occurred when increasing the number of coefficients from 30 to 40. With 30 coefficients, the difference in loudness between the two methods of control was significant. However, with 40 controller coefficients, the traditional method of sound pressure level control resulted in a loudness very similar to that resulting from loudness control. Indeed, the trend in these simulations is clear. Increasing the number of controller coefficients decreases the difference in loudness between the steady state error signals resulting from the two types of noise control. However, it is interesting to note that even when the resulting loudness is similar for the two types of control, there is a significant difference in the sound pressure level associated with the two error signals.

One possible explanation for the significant change in performance of the sound pressure level-based controller, when increasing the number of controller coefficients from 30 to 40, is that the controller has acquired enough coefficients to attempt attenuation of every sinusoid in the input noise signal. In general, the controller requires a minimum of two coefficients for each frequency it attempts to attenuate in the uncontrolled signal. These two coefficients contain the amplitude and phase information necessary for the signal matching and attenuation. Because the input noise signal for these simulations contains low-level random noise plus 11 sinusoids, the controller will not be able to attempt attenuation of all 11 sinusoids until it has at least 22 coefficients available. It appears that for this configuration, about 30 coefficients is near the minimum number of coefficients for the controller to begin to effectively attenuate all of the sinusoidal components in the signal. The loudness controller attenuates those frequencies which contribute most to the loudness of the sound, and is therefore able to achieve better loudness control with fewer coefficients.

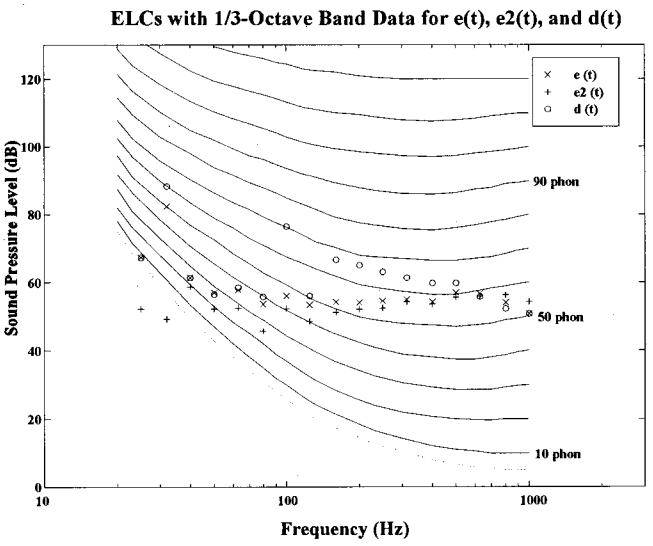


FIG. 12. Third-octave band levels of $d(t)$, $e(t)$, and $e2(t)$ for SIM8.

V. CONCLUSIONS

Since the development of the digital signal processor in the 1980's, active noise control has gained a considerable amount of attention among noise control engineers. A distinct focus of active noise control has been on applications of noise attenuation which affect the human listener. Therefore, the response of the human auditory system should at least be considered, if not explicitly incorporated, when designing the electronic controller in an ANC system.

Loudness, which is related to the human ear response, was chosen as a minimization criterion for simulations of noise control. The simulations indicate that utilizing this more subjective criterion than those traditionally used, such as sound pressure, does allow for noise control which in many cases, would likely be more pleasing to the human observer.

The apparent effectiveness of the technique employed for this project is certainly dependent upon the nature of the input signal. Several input noise signals were studied, and those that contained higher levels of low-frequency than high-frequency noise were most prominently affected by loudness control. Furthermore, because many naturally occurring acoustic signals do resemble some of the noise signals simulated during this research, it seems likely that this method would be effective in a selection of real ANC applications.

It was further noted that the number of coefficients made available to the electronic controller also affects how well this method performs. The more coefficients available, the closer the results of this technique of loudness control resemble those of traditional control, specifically in regard to loudness. Furthermore, utilizing many coefficients resulted in a significant difference in overall sound pressure level when comparing the two methods. These results are significant in that some applications of noise control have a restriction on the number of controller parameters that may be incorporated into the control system. This restriction may occur in applications involving many controllers and/or actuators, where processing time limitations exist. The results obtained here indicate that often the number of controller

TABLE V. Results which show the effect of increasing the number of controller coefficients.

Number of coefficients	Many sines (30 Hz Fund.) plus random input					
	Sound pressure level (dB)			Loudness (sones)		
	$d(t)$	$e(t)$	$e2(t)$	$d(t)$	$e(t)$	$e2(t)$
10	89.0	88.1	80.3	15.7	12.4	19.8
20	88.8	87.8	78.4	15.9	12.2	19.1
30	88.8	88.2	74.9	16.2	11.7	18.0
40	88.9	85.7	70.9	15.8	10.1	10.8
50	89.1	85.3	68.7	16.1	10.2	10.5
60	88.8	85.7	68.7	16.1	10.2	10.5
70	88.8	84.4	66.9	15.8	9.91	9.76
80	88.9	83.6	66.3	16.0	9.88	9.63
90	88.9	84.0	65.7	15.5	9.61	9.49
100	88.9	83.4	65.4	15.4	9.92	9.25
150	88.8	80.9	64.4	15.8	9.73	9.03
200	88.9	79.2	64.2	15.9	10.44	8.98

coefficients could be substantially reduced without sacrificing the perceived loudness attenuation that is achieved.

One would not necessarily need to physically incorporate this procedure in an actual control system in order to determine the likelihood of successful implementation. By examining the frequency content of a given noise signal, as well as the constraints of the electronic controller and the physical system, one would likely be able to determine before actual experiment if a technique such as digital filtering to minimize loudness would indeed be profitable. Along with the decision to utilize this filtering technique comes the choice as to which phon level loudness-based filter should be used in conjunction with a given noise signal. Many of the results seem to indicate that there is a correlation between the third-octave band levels of the uncontrolled input noise and the phon level filter which produced the ''best case,'' although these indications are not without exception. After examination of the frequency content of the input signal, it is likely that a reasonable choice could be made as to which phon level filter would produce favorable results.

Overall, this project was found to be worthwhile, and the results practical. The real implementation of this specific procedure would not be burdensome, and may feasibly be incorporated into many ANC systems. Furthermore, because all of the results mentioned in this text are numerical in nature, true subjective effectiveness of the described loudness minimization technique may not be absolutely qualified. Research to include a panel of human observers may constitute a useful extension of this project. The subjective qualification techniques of magnitude estimation and semantic differential are two methods by which human evaluations of the effectiveness of noise control may be analyzed.[15] These methods are recommended for possible future research endeavors regarding the incorporation of subjective measures in active noise control.

[1] J. Guo and J. Pan, ''Actively created quiet zones for broadband noise using multiple control sources and error microphones,'' J. Acoust. Soc. Am. **105**, 2294–2303 (1999).

[2] J. W. Parkins, S. D. Sommerfeldt, and J. Tichy, ''Narrowband and broadband active control in an enclosure using the acoustic energy density,'' J. Acoust. Soc. Am. **108**, 192–203 (2000).

[3] N. Tanaka, Y. Kikushima, and N. J. Fergusson, ''One-dimensional distributed modal sensors and the active modal control for planar structures,'' J. Acoust. Soc. Am. **104**, 217–225 (1998).

[4] R. Koehler and S. D. Snyder, ''Intensity error sensing in the active control of sound,'' J. Acoust. Soc. Am. **105**, 1155 (1999).

[5] S. J. Elliott and M. E. Johnson, ''Radiation modes and the active control of sound power,'' J. Acoust. Soc. Am. **94**, 2194–2204 (1993).

[6] W. R. Saunders and M. A. Vaudrey, ''Active noise control systems: Designing for the human auditory system,'' Proc. Noise-Con 96, pp. 385–390.

[7] E. Zwicker, H. Fastl, and C. Dallmayr, ''BASIC—Program for calculating the loudness of sounds from the 1/3-oct band spectra according to ISO532 B,'' Acustica **55**, 63–67 (1984).

[8] S. M. Kuo and J. Tsai, ''Residual noise shaping technique for active noise control systems,'' J. Acoust. Soc. Am. **95**, 1665–1666 (1994).

[9] S. M. Kuo and D. R. Morgan, *Active Noise Control Systems: Algorithms and DSP Implementations* (Wiley, New York, 1996), pp. 98–99.

[10] A. P. G. Peterson, *Handbook of Noise Measurement* (GenRad, Inc., Concord, MA, 1985), p. 33.

[11] A. P. G. Peterson, *Handbook of Noise Measurement* (GenRad, Inc., Concord, MA, 1985), pp. 37–38.

[12] B. Widrow and S. D. Stearns, *Adaptive Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1985), pp. 288–294.

[13] R. Hellman and E. Zwicker, ''Why can a decrease in dB(A) produce an increase in loudness?'' J. Acoust. Soc. Am. **82**, 1700–1705 (1987).

[14] R. Hellman, ''Loudness, annoyance, and noisiness produced by single-tone-noise-complexes,'' J. Acoust. Soc. Am. **72**, 62–73 (1982).

[15] A. Shoji and H. Hamada, ''Evaluation of the subjective effects for ANC,'' J. Acoust. Soc. Am. **100**, 2620 (1996).

Computation of edge diffraction for more accurate room acoustics auralization

Rendell R. Torres,[a] U. Peter Svensson,[b] and Mendel Kleiner
Chalmers Room Acoustics Group, Department of Applied Acoustics, Chalmers University of Technology, SE-412 96 Göteborg, Sweden

(Received 7 September 2000; revised 14 November 2000; accepted 15 November 2000)

Inaccuracies in computation and auralization of room impulse responses are related in part to inadequate modeling of edge diffraction, i.e., the scattering from edges of finite surfaces. A validated time-domain model (based on analytical extensions to the Biot–Tolstoy–Medwin technique) is thus employed here to compute early room impulse responses with edge diffraction. Furthermore, the computations are extended to include combinations of specular and diffracted paths in the example problem of a stage-house. These combinations constitute a significant component of the total nonspecular scattering and also help to identify edge diffraction in measured impulse responses. The computed impulse responses are then convolved with anechoic signals with a variety of time-frequency characteristics. Initial listening tests with varying orders and combinations of diffraction suggest that (1) depending on the input signal, the diffraction contributions can be clearly audible even in nonshadow zones for this conservative open geometry and (2) second-order diffraction to nonshadowed receivers can often be neglected. Finally, a practical implementation for binaural simulation is proposed, based on the singular behavior of edge diffraction along the least-time path for a given source–edge–receiver orientation. This study thus provides a first major step toward computing edge diffraction for more accurate room acoustics auralization. [DOI: 10.1121/1.1340647]

PACS numbers: 43.55.Ka, 43.55.Br, 43.55.Hy [JDQ]

I. INTRODUCTION

Edge diffraction is a fundamental component of the sound field around any finite reflecting surface. In room acoustics, for example, it describes not only the commonly noted phenomenon of sound propagating around corners (or emitting from orchestra pits) but also the scattering in *all* directions from wedges of any angle, e.g., the boundaries of a stage-house proscenium or the ''knife-edges'' of thin overhead reflectors. For frequencies low enough that small-scale surface roughness becomes negligible with respect to wavelength, one may even consider much of a room's interior as a simple assemblage of various wedges that reflect and diffract.

Conventional methods for computing room impulse responses (RIRs) for auralization[1,2] typically neglect edge diffraction and rely on geometrical acoustics, e.g., the image-source model,[3] ray tracing,[4] or some variation of these.[5–12] (Our definition of auralization, the acoustical analog of visualization, was introduced in Ref. 1: ''Auralization is the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space.'') It is known, however, that a pure geometrical-acoustics model erroneously predicts a discontinuous field. Figure 1 illustrates a simple case (see also Refs. 13, 14) where one can delineate three ''zones'' around a rigid right-angle wedge: zone I (listener receives direct and reflected waves), zone II (listener receives direct wave but not reflected wave), zone III (listener shadowed from direct and reflected waves). In addition, sound is diffracted from the edge *into all three zones*. The geometrical-acoustics interpretation (depicted by spikes in the schematic impulse responses) is discontinuous at the zone boundaries; edge diffraction, portrayed by quasi-exponential ''tails'' in the figure, exactly compensates for this (with delay, magnitude, and discontinuous sign change) and thus maintains a continuous pressure field around the edge. One can, moreover, define edge diffraction as the difference between the total scattered field and that given by geometrical acoustics in the presence of a wedge.

A second related problem with geometrical-acoustics is that inaccuracies in computing reflection strength increase with wavelength relative to surface dimensions. Moreover, in computing the RIR for a particular field position, edge diffraction is necessary to correct the spectral content of reflected impulses from finite surfaces, e.g., balcony soffits, overhead reflectors, or wall edges. In a first ''round-robin'' comparison by Vorländer[15] of room acoustics computations with hall measurements at several room positions, certain parameters such as early decay time (EDT) and clarity (C) were more difficult to predict than others, e.g., reverberation time (RT). (For parameter definitions see Ref. 16 or 17.) The fact that such parameters as EDT, C, *Deutlichkeit* (D), and Lateral Energy Fraction (LEF) depend greatly on the early structure of the RIR further necessitates a model of edge diffraction that computes at least early-order room reflections accurately (after which the reverberation generally

[a]Electronic mail: rendell@ta.chalmers.se
[b]Current address: Acoustics Group, Department of Telecommunications, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway; electronic mail: svensson@tele.ntnu.no

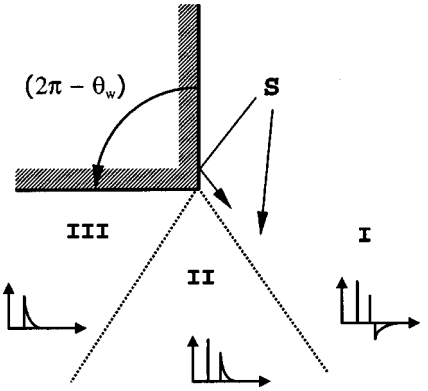


FIG. 1. The area near an edge can be defined by three ''zones,'' each with a different combination of waves (direct, specularly reflected, diffracted), depicted by schematic impulse responses. A purely geometrical-acoustics interpretation (corresponding to the impulsive ''spikes'' above) is discontinuous at the zone boundaries; edge diffraction, present in all zones, compensates for this. The open wedge angle is given by θ_w.

becomes more statistical). Finally, as the ear's sensitivity exceeds the frequency- and time-resolution of standard parameters, auralization requires even greater detail to recreate the wide-band ''filtering'' (i.e., the natural coloration) that a real room endows upon an anechoic input signal. Although somewhat ''realistic sounding'' auralization is achievable even for simplified, ''virtual-reality'' rendering,[18,19] inclusion of edge diffraction could increase ''auralization *accuracy*'' (i.e., the replication of an actual room), particularly at lower frequencies, and thus enable auralization to complement numerical parameters with a more complete impression of the room's acoustics.

Our primary aim, therefore, is to include edge diffraction in computed room impulse responses, which thus entails modeling of diffracted specular reflections and specularly reflected diffractions. Such ''specular/diffractive'' combinations are shown here to comprise a significant part of the nonspecular scattering in measured scale-model impulse responses. We then present an example auralization based on the early RIR (i.e., from the stage-house ''sending end'' of a concert hall) and perform initial listening tests to assess the audibility of diffraction for various input signals. Two aspects of auralization, however, are left to subsequent studies on subjective aspects. First, since the amplitude of each diffraction order decreases by one or more orders of magnitude, we focus on modeling edge-diffraction effects in the early-order RIR and exclude computation of reverberation, which can be calculated sufficiently well using statistically based methods.[6] Modeling a nonreverberant geometry also enables clearer identification of diffraction when comparing measured and computed RIRs and can be directly applied to simulating outdoor amphitheaters and other ''open'' venues. One can fairly argue, however, that reverberation may partially mask diffraction effects during auralization; nevertheless, the ''early signature'' of the auralized RIR still warrants close study, as it imparts information on the room's initial filtering of the input signal. For example, in auralization listening tests,[20] changes in early coloration (e.g., perceived shifts in ''bass'' or ''treble'') due to varying surface ''diffusion'' were clearly audible even with 2-s reverberation and using a continuous input signal (pink noise, 5 s). The second aspect is that our initial example auralizations are monaural, i.e., exclude head-related transfer functions (HRTFs). However, this still simulates basic coloration changes (due to including edge diffraction) in the early RIR for source–receiver orientations that involve predominantly frontal HRTF angles and are situated near the centerline, as in our case. Also, since a brute-force inclusion of a large number of HRTFs along each wedge is impractical and presumably unnecessary subjectively, we propose below a practical implementation for binaural simulation of edge diffraction, based on its singular behavior along the least-time path for a given source–wedge–receiver orientation. Moreover, despite these limitations, this initial study provides the first major step toward computing edge diffraction for room acoustics auralization and shows how these two aspects should be included in future work.

Section I discusses previous work and the theory behind the time-domain edge-diffraction model used here, based on analytical extensions by Svensson *et al.*[21] to the Biot–Tolstoy–Medwin (BTM) technique.[22,23] In addition, the computations are extended to include combinations of specular and diffracted components, which improve agreement with measured RIRs from a model stage-house, described in Sec. II. Computed RIRs with various orders and components of diffraction are then compared in frequency responses and ABX listening tests in Sec. III, followed by proposals for binaural simulation and future work.

II. COMPUTATIONAL APPROACH

A. Previous work

The literature on diffraction describes many models, formulated in both time and frequency domains. The latter encompasses several approaches such as the well-known Geometrical Theory of Diffraction (GTD)[24] and work by Kouyoumijian and Pathak,[25] Pierce,[26,14] Salomons,[27] and Menounou *et al.*[28] (see these for more comprehensive reviews), although the model in Ref. 28 can also be used in the time domain. Berkhout *et al.*[29] calculated a two-dimensional room field with a frequency-domain model, from seismic imaging, which was derived from the Kirchhoff–Helmholtz integral. In room acoustics, however, time-domain formulations distinguishing the different (specular and diffracted) transient components are attractive as they enable clearer interpretation of measured data and insight into predicted phenomena [Refs. 13, 22, and 30 (Sec. IV)]. Time-domain models are also appropriate for room acoustics analysis and auralization because one typically requires impulse responses (instead of frequency data) at specific positions. As noted by Keiffer and Novarini,[30] their use of the time-domain BTM solution (discussed below) was much faster in computing impulse responses than more traditional harmonic synthesis, which can involve relatively small frequency increments and high cutoff frequencies for ''full-bandwidth'' auralization. Although we focus the following discussion on previous work in room acoustics, we underscore below that

our method is an extension of work developed and validated in other fields, including diffraction from barriers and surface scattering.

The most approximate treatments of edge diffraction in room acoustics prediction use Lambert ''diffusion'' methods.[31,32] Dalenbäck,[33] for example (followed by Farina[34]), employed an ''edge-diffusion'' technique that placed Lambert-radiating sources on a strip (of quarter-wavelength width) along the perimeter of a finite reflector. Specular reflections near the edges would then decrease in amplitude, approximately simulating the interference of edge diffraction in the specular direction. Such energy-based methods, however, neglect phase information and inherently cannot replicate the correct interference effects of edge diffraction components from nonspecular directions to the receiver.[20] Furthermore, even if some phase is assumed (e.g., minimum phase), a Lambert edge source directs its lobe maximum in the normal direction from the plane and does not reproduce the actual, more complex (diffraction) directivity, which varies with incidence angle and position along the edge. Finally, the selection of the appropriate Lambert-diffusion coefficient is somewhat arbitrary or, at best, only roughly determinable.

[Regarding terminology, one may note that the term ''diffusion'' is typically related to the spreading of incident specular *energy* into nonspecular directions (in its broad usage in room acoustics and auralization). ''Diffusion'' is sometimes equated with ''diffuse reflection''[32] and perhaps offers a more heuristic (than rigorous) picture of scattering phenomena. ''Scattering,'' as defined, e.g., by Medwin and Clay,[35] refers to the redirection of sound when it interacts with a body and, thus encompasses reflected, diffracted, and transmitted waves. When referring specifically to the nonspecular components of this redirection of sound, the term ''scattering'' may still be preferable to ''diffusion,'' as the latter can be easily confused with *diffusivity of the sound field*, which is related but *not* equivalent to diffuse reflection.[31] See Ref. 20 for further discussion on ''diffusion'' versus ''scattering'' versus ''diffraction'' within room acoustics.]

Other diffraction studies in room acoustics invoke the Kirchhoff Approximation (KA).[36–38] Although KA diffraction models can be useful near the specular scattering angle[36,39] and sometimes at higher frequencies,[40] they *completely fail* in predicting the diffracted component for *all* frequencies at certain incident and scattering angles, as discussed and well illustrated by Jebsen and Medwin[41] and Norton *et al.*[42] For example, the oblique edge scattering from overhead reflectors would have incorrect amplitudes with these modeling methods. Fresnel and Fraunhofer diffraction methods have similar failings, as shown by Cox and Lam.[40]

B. Biot–Tolstoy (BT) time-domain formulations

More accurate time-domain models of edge diffraction have been based on some form of the exact Biot–Tolstoy (BT) solution[43] for a point source irradiating an infinite wedge of arbitrary angle θ_w (see Fig. 1). In Refs. 22 and 23, Medwin *et al.* adapted the BT solution to include finite-length wedges and second-order (double) diffraction by using a ''discrete Huygens interpretation'' that utilized secondary sources along edges. This method, often called the BTM technique, was extended by Svensson *et al.*,[21] who derived analytical directivity functions for the edge sources. These derivations justified Medwin's assumptions[22] for modeling first-order diffraction and refined the BTM modeling of second-order diffraction, which becomes more important for grazing angles and decreasing surface dimensions.[44] The model by Svensson *et al.* can compute first- and second-order diffraction for curved edges as well. Although the BT solution requires that wedges have Neumann (rigid) or Dirichlet (pressure-release) boundary conditions, this is not a greatly disappointing restriction, as many (although not all) concert-hall surfaces of interest for edge diffraction are essentially acoustically hard. (This would not apply, for example, to the audience area, absorptive decor, or light paneling.) The analytical expressions and calculation method that we employ (and extend below) are extensively illustrated in Svensson *et al.*[21] [Eqs. (12), (19), (28)–(29), and Sec. II] and are excluded here for brevity. Figure 2 schematically depicts the angle-dependent interference of edge sources with the specular field (i.e., the image-source's wave front) reflected from a finite plane. The image-source model is an exact ''equivalent source'' for an infinite rigid plane (or ''before'' the wave front reaches the edge); the change of impedance (at the finite reflector edges) is modeled by adding edge sources of the resulting scattering, i.e., edge diffraction. (It should be noted that the representation of the total scattered field by the image source plus edge sources does not rigorously apply to the modeling of impedance boundaries. Potential methods for handling impedance wedges are given in Refs. 45 and 46.)

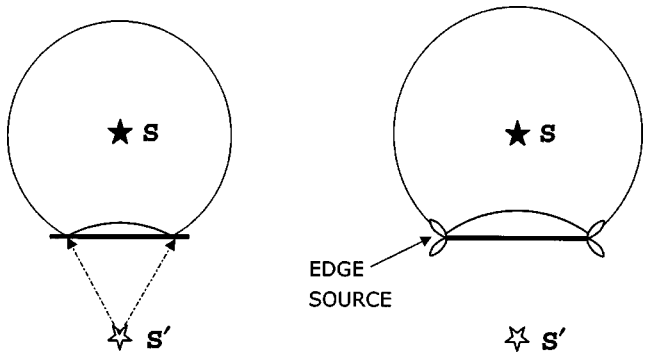


FIG. 2. The image-source model is an exact ''equivalent source'' for reflection from an infinite rigid plane. A change of impedance on the plane (i.e., at the edges of a finite reflector) results in scattering (modeled by edge sources schematically drawn above) which interferes with the specular reflection (modeled by the image source S').

Previous use of the BT solution in room acoustics is sparse. Ouis,[47] for example, employed the BT solution to calculate diffraction from a thin ''infinite'' strip in a simple rectangular auditorium. He also compared the BT solution with other frequency-domain formulations by examining diffraction from an infinite strip on a plane.[48] In another study Kovitz[49] adapted the BT solution to an image-source room model and applied the same truncation factor (1/2) as Medwin for first-order diffraction from finite (truncated) wedges. Kovitz did not continue, however, with double diffraction or

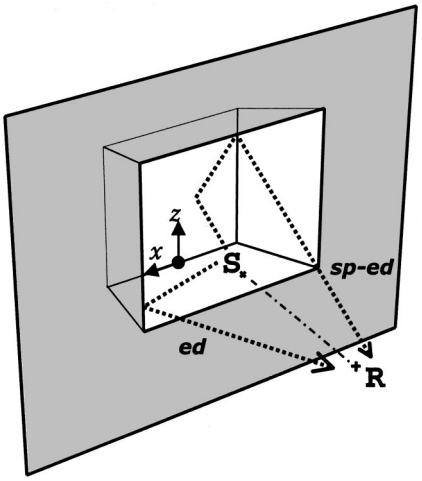


FIG. 3. The test geometry is a stage-house in a finite baffle. One must not only compute ''direct'' edge-diffraction paths but also combinations of specular/diffracted components. Direct edge-diffraction paths are denoted by *ed* or *ed-ed* (second-order edge diffraction, by two successive edges). Path *sp-ed* represents specular reflection (*sp*) to edge diffraction (*ed*).

reflected diffraction components. The work here advances these studies by examining a room geometry with several wedge types (i.e., wedge angles) of finite length, and by incorporating second-order diffraction and three combinations of specular/diffractive components, discussed further below. This study is also a first step toward accurate inclusion of edge diffraction in auralization, compared to current methods that use Lambert approximations or neglect diffraction entirely.

The BTM model has been validated for several different geometries, such as a circular disk,[44] a right-angle step discontinuity,[13] thick barriers,[23,50] double-diffraction from knife-edges,[50] and ''wedge assemblages'' of spheres[51] and rough surfaces.[30] Svensson *et al.* also validated their extensions to the BTM model in the test problems of a circular disk and a thin rectangular plate.[21] The individual wedges that compose the stage-house example below fall within these cases, so we concentrate primarily on determining the necessary combinations and orders of diffraction paths to adapt the computations to room acoustics.

C. Extension: Specular/diffractive combinations

Svensson *et al.* describe in Ref. 21 how multiple-order edge diffraction could be computed with the model, which was numerically implemented up to second-order in example computations. These benchmark comparisons only required ''direct'' diffraction paths, i.e., source to edge(s) to receiver. Room acoustics problems, however, typically involve intermediate reflections, e.g., edge-diffracted components (*ed*) that reflect specularly, and specular reflections (*sp*) that diffract (see Fig. 3). We refer to these as combinations of specular/diffracted components. To improve modeling of the early RIR, we begin by extending the computations to include the following combinations: *sp-ed*, *ed-sp*, and *sp-ed-sp*. For a single wedge (e.g., a thin barrier) on a reflecting plane, these three paths are modeled with straightforward use of an image source and image receiver (e.g., Sec. 4 in Ref. 52). For a room, in contrast, with a multitude of reflection paths among concave and convex wedge geometries, one must construct physically sensible ''visibility checks'' of whether a given edge contributes diffraction for a selected source–receiver orientation.

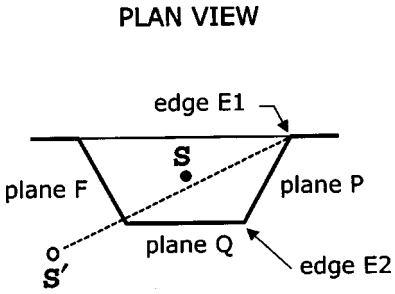


FIG. 4. The edge-visibility check should determine that edge $E1$ is visible to the source and that $E2$ is not. Furthermore, all edges of the reflecting plane F must be excluded from the diffraction calculations for image source S'.

1. Modeling of paths sp-ed, ed-sp, sp-ed-sp

The paths *sp-ed* are modeled by replacing each reflecting plane with a (first-order) image source and then computing diffraction from edges visible to both the receiver and each image source. (Image-source models and plane-visibility checks are discussed in Refs. 3, 49, and 53.) As depicted in Fig. 4, for example, the edge-visibility check should determine that edge $E1$ (perpendicular to the page) is visible to the image source S' and that edge $E2$ is not. Thus the following steps are performed for each image source and edge: (1) Check that the path from the image source to the first edge endpoint passes through the reflecting plane and is not blocked (e.g., by other planes). (2) Repeat for the second edge point. (3) If at least one of the two edge points is visible, assume that the edge is visible. Furthermore, all edges of the reflecting plane that created the image source must be excluded from the diffraction calculations for each *sp-ed* path. For example, although plane F is not ''visible'' (having spawned the image source S'), plane Q and its edges may be visible via plane F for a different source position. If the edge that plane Q shares with plane F is then considered visible, additional edge diffraction would be erroneously calculated.

Thus, the components *sp-ed* are given by the following:

$$h_{sp\text{-}ed}(t;S|R)=\sum_i \sum_j h_{\mathrm{diffr1}}(t;S_i'|E_j|R), \tag{1}$$

where $h_{\mathrm{diffr1}}(t;S_i'|E_j|R)$ is the first-order diffraction impulse response from the image source S_i' via the edge E_j to the receiver R, t is time, and the edge summations (i.e., diffraction calculations) are done over the indices j remaining after the visibility checks. Note that the free-field impulse response is defined here (and in Ref. 21) as $r^{-1}\delta(t-r/c)$, where r is propagation distance in meters and c is the speed of sound in air.

This initial extension of the program does not divide edges into visible and hidden segments. Thus step (3) above is a choice between two approximations: to consider partially visible edges (1) as totally visible or (2) as totally hidden.

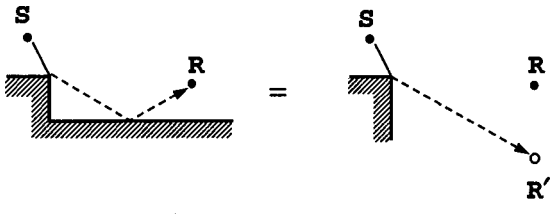


FIG. 5. An *ed-sp* combination is modeled using an image receiver R', with the assumptions of a point source S, a point receiver R, and rigid boundaries.

The first approximation, discussed below, is selected because the visible part of the edge *still* diffracts, although with shortened edge length, whereas the second approximation entirely neglects this diffraction.

One can analyze the first approximation in detail. The main changes to a partially obscured edge are (a) that its effective length decreases and (b) that the ''least-time point'' on the edge (on the shortest path from S' to E to R) could *possibly* shift. The first change results in a small change in amplitude at low frequencies, as the diffraction contribution is maximum at its onset from the least-time point but decreases slowly from farther points along the edge {cf. Sec. IV, and Refs. 21 [Eqs. (1), (9), (19)], 22, and 30}. Moreover, since the transient onset mainly affects higher frequencies while the edge ''outskirts'' mainly affect lower frequencies, the effective ''truncation'' of the edge has less significance for higher frequencies. Regarding the second change (*if* the least-time point shifts), there are one or possibly two additional effects. One effect is a small error in the angle and time delay of the edge diffraction relative to the receiver (the error may increase at very close distances and extreme S'-E-R orientations). Another effect is a possible drop in the amplitude, which occurs if the least-time point on a given *finite* wedge coincides with the least-time point for the corresponding (''virtual'') *infinite* wedge (Fig. 2 in Refs. 21 and 30). If it does (condition 1), its diffraction contribution is theoretically singular (but computationally finite), and if this ''infinite edge'' least-time point is then obscured (condition 2), the edge's peak diffraction amplitude should decrease significantly. Whether both conditions occur for a given path depends on the room geometry and source–receiver orientation. Although the measurements below indicate that we have chosen the better approximation (or that the choice makes little difference for the case below), future calculations should divide edges into visible and non-visible segments.

As depicted in Fig. 5, the paths *ed-sp* (edge diffraction that subsequently reflects) are modeled as diffraction from the edges E_j visible to both the source and to each image *receiver* R_i':

$$h_{ed\text{-}sp}(t;S|R)=\sum_i \sum_j h_{\mathrm{diffr1}}(t;S|E_j|R_i'). \qquad (2)$$

Finally, the combinations *sp-ed-sp* (specular reflection, diffraction at each visible edge, and subsequent reflection to the receiver) correspond to diffraction from edges that are visible to all possible pairs of image sources and image receivers:

$$h_{sp\text{-}ed\text{-}sp}(t;S|R)=\sum_i \sum_j \sum_k h_{\mathrm{diffr1}}(t;S_i'|E_j|R_k'). \qquad (3)$$

One may note that the ''allowable'' (visible) combinations of S_i', E_j, and R_k' for paths *sp-ed-sp* are given simply by the intersections of corresponding visibility checks already performed for Eqs. (1) and (2). Thus no additional visibility checks are required for Eq. (3).

2. Comments

One can similarly include additional combinations of first-order diffraction and higher-order image-sources/receivers (e.g., *sp-sp-ed*, *ed-sp-sp*), although the latter become successively weaker by spherical spreading and from absorption and scattering, e.g., by the musicians and audience. Such effects should be included in future studies applying the above method. Although the calculations are restricted to point sources and receivers, one could practically construct larger sources (e.g., orchestras, choirs) as an ensemble of monopoles, although their relative amplitudes must be determined. Such modeling of directive sound sources for auralization has been investigated by Giron.[54]

A similar extension can also be applied to combinations including double diffraction. However, as second-order edge diffraction is typically orders-of-magnitude below first-order diffraction (and even weaker compared to the total specular field), such computationally intensive combinations may practically be neglected, as also suggested by Wadsworth and Chamber's[50] measurements of *ed-sp-ed* paths. [In fact, in modeling rough-surface scattering, Keiffer and Novarini[30] only used first-order scattering to successfully implement the wedge-assemblage method (WAM), i.e., use of the BTM method to compute scattering from a surface or object modeled as a construction of wedges.] Second-order edge diffraction is included here for direct paths (*ed-ed*), although its relative importance is still under study and depends on various factors (e.g., wedge angle, wavelength relative to projected surface extent, grazing incident/scattering angles). Third- and higher-order diffraction appears unnecessary for most validated cases[21,23,44] and in computations below. In any case, specular/diffractive combinations of first-order diffraction with higher-order specular reflection may offer a suitable compromise between speed of calculations and computation of audible diffraction contributions in the early RIR.

III. TEST GEOMETRY

A stage-house form (Fig. 3) is selected as an ideal, conservative geometry to test for the significance of edge diffraction in the early RIR. First, it is a common concert-hall component (sometimes called the ''sending end'') and involves several different wedge angles and reflection/diffraction paths. It is also a ''conservative'' scale, such that geometrical acoustics should hold except at very low frequencies, relative to the large dimensions of the smooth reflecting surfaces. (The stage-house, described below, is proportioned similarly to that in Boston Symphony Hall, although replication of any particular stage is not intended.) The nonshadowed source–receiver orientation is also conser-

vative, compared to receiver positions hidden from the source, and can be appropriate for future subjective tests judging the audibility of edge diffraction to such positions. For computations below, the numerical procedure in Sec. II is used to calculate the edge diffraction and its specular/diffractive combinations; the specular reflections are calculated with a commercially available room acoustics program.[55] As discussed above, the computations here assume rigid boundary conditions.

A. Measurements

Although the numerical model has been validated for relatively simple geometries and first-order rough-surface scattering, one must still determine the necessary edge-diffraction *paths* (e.g., various diffraction components) to sufficiently model the early RIR for this concave roomlike geometry. Following Refs. 13, 28, and 50, we measure impulse data from a scale model and verify that the measured edge-diffraction contributions are also observed in the computations.

A 1:20 scale-model is constructed of 8-mm-thick Perspex acrylic (1190 kg/m^3) and reinforced for greater rigidity. The following dimensions are given in ''full-scale'' (i.e., 20 times the model's dimensions): The proscenium height and width are 14.5 m and 18.6 m. The back wall is 12.5 m high and 14 m wide. The stage depth is 8.3 m with a horizontal floor. The stage-house sits eccentrically ($\Delta x = -0.8$ m, $\Delta z = 1.25$) in the finite baffle, with height 30 m and width 40 m. With the origin bisecting the rear-floor edge, the source's and receiver's coordinates are (-2, 7, 1.18) and (-2, 20.9, -0.4) [m]. (Other nonshadowed source–receiver combinations are also measured, of which this is a representative combination.)

A calibrated condenser microphone (6.35 mm in diameter) is used to measure impulse data from a spark source composed of two electrodes inserted through the scale-model floor. The directivity of the spark source (without floor) is measured separately, however, in 15-deg increments in three orthogonal planes. Up to 80 kHz (4-kHz full scale), the measured spark-source directivity is essentially omnidirectional (deviations within $\pm$ 1 dB). The sampling frequency for the stage-house measurements is 80 kHz, although the measurements are passively low-pass filtered by the microphone amplifier up to about 25 kHz (about 1.3-kHz full scale), approximately where air absorption is dominant. For these large reflecting surfaces, the dominant edge-diffraction contributions lie well below this cutoff frequency. Each measurement is averaged over approximately 200 captures.

B. Results

Measured and computed data in Figs. 6–7 show how travel paths of nonspecular components in the early RIR are well represented when edge diffraction is modeled. (Measured data are given in ''full-scale'' unless noted otherwise.) In Fig. 6, the upper plot shows the measured impulse response, while the lower plot shows the computed impulse response, convolved with the (spark) source signal for clearer comparison with the measured data. Additionally, the computed impulse response is (low-pass) Butterworth-

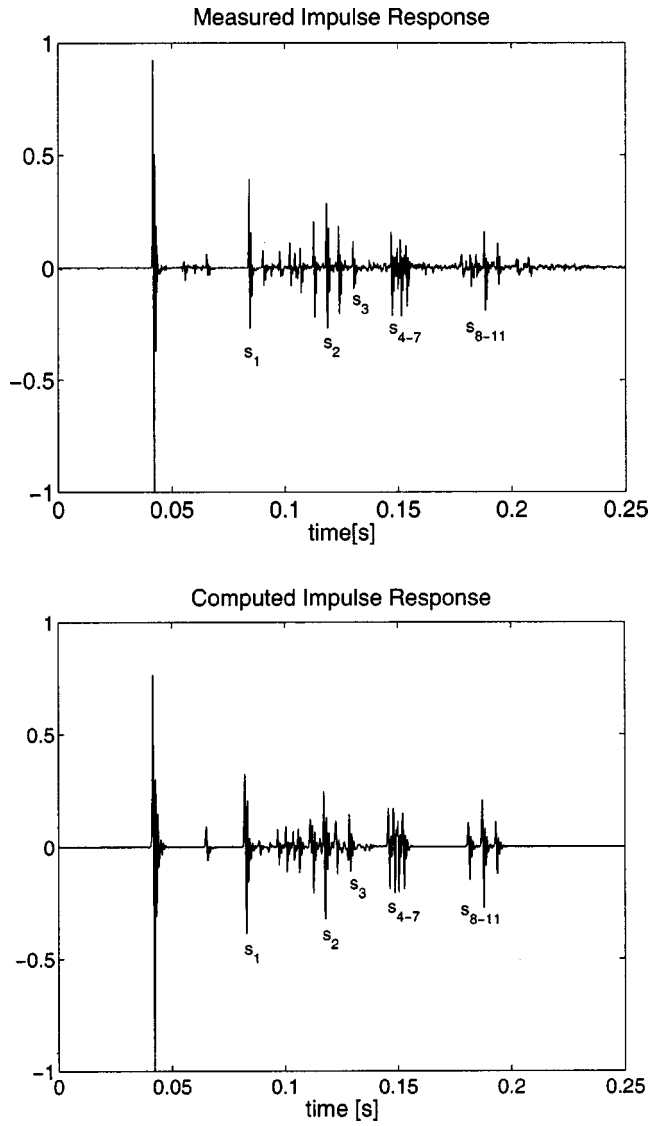


FIG. 6. The upper figure plots the measured impulse response in front of the stage-house. The vertical axis is linear pressure, normalized to the direct sound. The lower figure plots the computed impulse response convolved with the (spark) source signal. The arrival times of the specular reflections are denoted by ''*s*'' (s_9 and s_{10} come nearly simultaneously); the nonspecular arrivals, particularly between s_1 and s_3, are edge-diffraction components, shown more clearly in Fig. 7.

filtered to the same effective frequency range of the measurements, and the decay due to air-attenuation at high (scale-model) frequencies is also modeled, as described further below. The label s_n denotes the 11 specular reflections (of which s_9 and s_{10} arrive nearly simultaneously). Peak-by-peak comparison, particularly between s_1 and s_3, shows that the computed impulse response includes nearly all of the major nonspecular arrivals (edge-diffraction components) in the measured RIR. (The measured scattering from the spark-source's arms and bracing is not included in computations. The second-order edge diffraction from the edges of the finite baffle, however, is included.) This time-domain numerical model also allows straightforward dissection into the individually computed transient components (i.e., specular reflections, direct diffraction, *sp-ed* combinations, etc.), which were separately plotted (not shown here) to identify

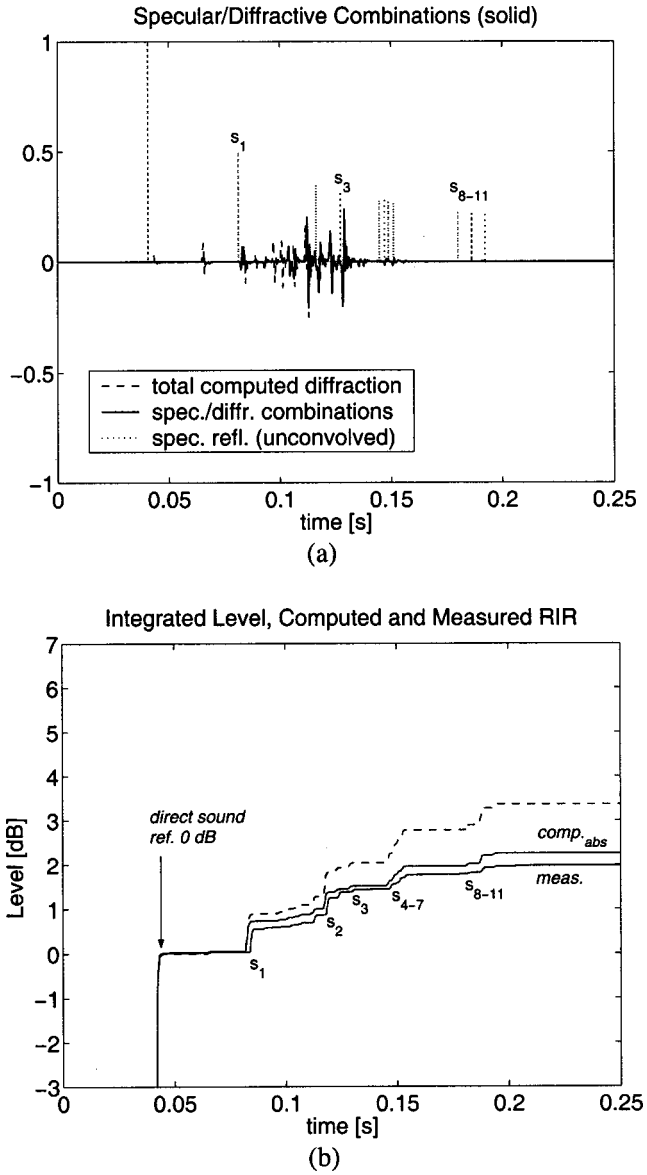


FIG. 7. (a) The upper figure shows the total computed edge diffraction, plotted in dark dashed lines, with specular/diffractive combinations in solid lines. The positive dotted lines represent the (unconvolved) direct sound and specular reflections, plotted for comparison. One sees from the arrival times (e.g., at the third specular reflection) that specular reflections and edge diffraction can interfere closely. (b) The lower figure shows the integrated level (over time) of the computed and measured impulse responses. The lowest curve is the measured data, which lies 0.25 dB below the computed results that are approximately compensated for air absorption. The dashed line shows the computed results without the modeled attenuation, which represents the levels at more typical audio frequencies.

various components of the measured response. (Note that such dissection is not easily performed with, e.g., BEM/FEM techniques.) These components are superposed to obtain the total computed RIR.

In Fig. 7, the upper plot shows the total computed diffraction plotted in dark dashed lines, with specular/diffractive combinations emphasized in solid lines. The vertical dotted lines represent the (unconvolved) direct sound and specular reflections, plotted for comparison. One sees from the arrival times, e.g., at the third specular reflection, that specular reflections and edge diffraction can interfere destructively, evident in the reduced measured and computed amplitude of s_3 in Fig. 6. Without the computed edge diffraction, the amplitude of s_3 would be much greater in this (lower) frequency range. (However, one can easily show that if the low-pass filter on the computed RIR is extended higher in frequency, the destructive interference of the edge diffraction would have less effect on the total amplitude of peak s_3, which would essentially retain its specular amplitude.) Surprisingly, the specular/diffractive combinations constitute a major part of the total nonspecular scattering, as one has a potentially large number of edge-diffraction paths when the corresponding image-sources and image-receivers are generated. Moreover, without these extra combinations, the total edge diffraction would be incompletely predicted. Even the inclusion of just three path-combinations (*sp-ed*, *ed-sp*,*sp-ed-sp*) significantly increases agreement with the measured early RIR by resolving the major ''unknown'' nonspecular peaks. Computing combinations with higher-order image-sources/receivers (e.g., *sp-sp-ed*, etc.) may further improve agreement.

The lower plot in Fig. 7 shows the integrated level of the impulse responses over time. The lower solid curve corresponds to the measured response, while the upper dashed line corresponds to the computed response, not accounting for the air absorption at high, scale-model frequencies (here about 10–25 kHz). The frequency-dependent air attenuation could be included in computed results by, for example, constructing a low-pass filter with time-dependent filter coefficients, or by analyzing separate frequency bands (see also Refs. 56 and 57). To obtain a rough estimate, however, of how much the cumulative difference from the measured response depends on air absorption, one could apply an exponential decay corresponding to the air absorption at 25 kHz to the computed impulse response. The attenuation coefficient at this frequency is chosen because the spark signal's spectrum increases by about 4–6 dB/octave, peaks at approximately 25 kHz, and then falls sharply above this. The measured temperature, relative humidity, and ambient atmospheric pressure in the room are 22 °C, 38%, and 761 mm Hg (although these values can have some time-variance). Using Eqs. (3)–(5) in ISO 9613-1:1993(E), the computed attenuation by air absorption at 25 kHz is 0.79 dB/m, corresponding to a decay of $\exp[(-0.79/10\log e)r]$. Applying this exponential decay to the computed impulse response (resulting in the upper solid line) and comparing with the measured data, one sees excellent agreement in the progressive rise in level and that the total cumulative difference is about 0.25 dB. Although this is an ''over''-approximation of the attenuation, the resulting agreement suggests that the overall differences in level are indeed primarily due to air absorption. The ''true'' compensated curve would presumably lie between the dashed and the upper solid line.

One may also note in Fig. 7(b) that diffraction is not the main cause of discrepancies in the integrated levels (as the computed curve follows the measurement well, e.g., between s_1 and s_3). Rather, the specular reflections are, in general, too strong. This may be due to assuming ideally rigid reflecting surfaces, which actually have finite surface impedance in the scale model. [The curve for the purely specular computation (neglecting diffraction) is omitted for clarity; it is sim-

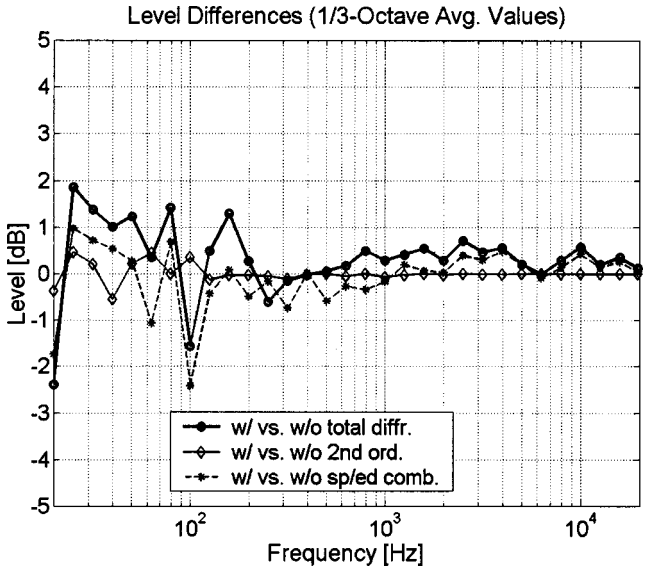


FIG. 8. Level differences, for third-octave bands, when various orders and combinations of diffraction are included in the computed room impulse response. The curves are linearly fitted through the average values (e.g., circles). The circles (with heavy solid line) correspond to including the total computed edge diffraction. Neglecting second-order diffraction corresponds to level differences (diamonds, solid line) up to about 0.5 dB. Neglecting specular/diffractive combinations can create level differences (asterisks, dashed line) up to 2.5 dB in certain frequency ranges.

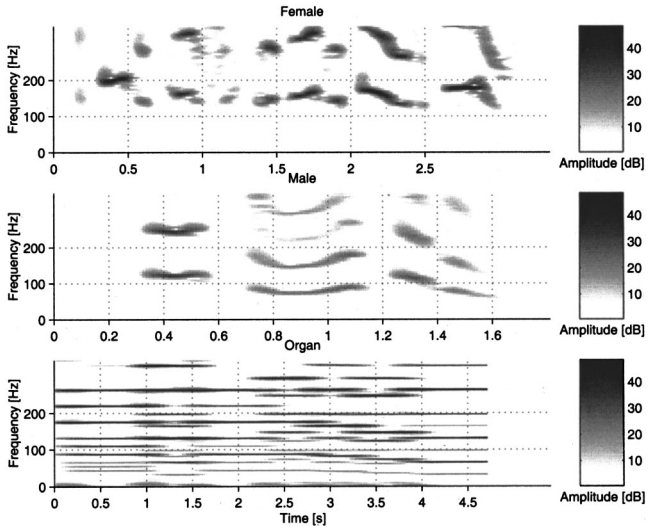


FIG. 9. Spectrograms of anechoic signals in the lower-frequency region where edge diffraction has greatest effect. From top to bottom: female speech, male speech, synthesized organ music. The vertical axes range from 0 to 350 Hz (left side, linear scale) and 0 to 50 dB (right side).

ply constant (horizontal) between specular reflections (s) and deviates most from the other curves between reflections s_1 and s_3, where the diffraction contributions are greatest.] Finally, the slight temporal deviations from measured data are related to small differences in the source–receiver positions.

IV. ASPECTS OF AUDIBILITY

A complete psychoacoustical investigation of edge diffraction in auralization extends beyond the scope of this paper, but a short initial study can help map interesting areas for future work. We first examine basic coloration changes in frequency-response functions for different orders/combinations of diffraction. To assess the audibility of these changes, we use ABX listening tests to compare a variety of anechoic signals (e.g., music, speech) convolved with computed RIRs having various orders/combinations of diffraction. (''Coloration'' here refers simply to frequency-dependent level differences due to edge diffraction; other definitions are discussed, e.g., by Meynial and Vuichard.[58] Additional subjective effects may also be present, of course.)

A. Frequency response comparisons

Figure 8 shows spectral level differences (averaged over third-octave bands) that occur when various orders and combinations of edge diffraction are included in the computed RIR. For ease of comparison, curves are linearly fitted between the average values. The circles (with heavy solid line) correspond to average level differences between the total computed RIR and the purely specular RIR (i.e., with-or-without the total computed edge diffraction). The diamonds mark level differences with-and-without second-order diffraction in the total computed RIR, and the asterisks (with dashed line) correspond to level differences with-and-without specular/diffractive combinations in the total computed RIR.

For this geometry, level differences occur at relatively low frequencies, where geometrical acoustics becomes invalid for the large bare surfaces. When the total edge diffraction is computed, third-octave-averaged level differences (circles) of about 1–2 dB occur below about 160 Hz. For input signals with significant low-frequency content, these small coloration changes are still clearly audible, demonstrated by listening tests below. Note, however, that realistic cases could have more diffractive wedges of different scales, which would result in greater diffraction effects, extending higher in frequency.

Computing second-order diffraction corresponds to level differences of about 0.5 dB at very low frequencies (here below 30 Hz). Thus second-order computations appear sufficient and perhaps even unnecessary in nonshadowed positions here, although it can become more important, e.g., for smaller reflecting surfaces (relative to wavelength), and grazing incidence/scattering angles. Finally, including specular/diffractive combinations results in level differences of up to 2.5 dB within limited frequency ranges.

B. Initial subjective tests

In Fig. 8, level differences of only about 1–2 dB appeared at low frequencies, for this conservative geometry and nonshadowed source–receiver orientation. Nevertheless, initial ABX listening tests below show that such small spectral changes are clearly audible for various input signals. Impulse responses with different diffraction orders and components (i.e., the cases in Fig. 8) are convolved with the following anechoic signals: pink noise, (synthesized) organ music, male and female speech, and a unit impulse (i.e., the bare impulse response, which one may note, emphasizes effects in higher-frequency bands). The spectrograms for the speech and organ music are shown in Fig. 9. The test taker

TABLE I. ABX listening-test results (0.05 significance level) from convolving anechoic signals with stage-house RIRs with various orders and components of edge diffraction. Note that comparisons using signals with greater low-frequency content (e.g., impulse, organ, pink noise) were deemed more audible than others. The questions are reformulations of the cases discussed in the text. The ratios in parentheses correspond to correct identifications of *X*. (''YES'' corresponds to a significant audible difference, whereas ''NO'' actually only means the null hypothesis $A=B$ was not disproven.)

Input signal	Is the total diffraction audible?	Is second-order diffraction audible?	Are specular/diffractive combinations audible?
Impulse	YES (18/18)	NO (6/18)	POSSIBLY (11/18)
Organ	YES (13/18)	NO (10/18)	NO (10/18)
Pink noise	YES (13/18)	···	···
Female voice	NO (9/18)	···	···
Male voice	NO (7/18)	···	···

specifies whether a given sound ''*X*'' is the same as sound ''*A*'' or ''*B*'' where, for example, ''*A*'' is a sound sample with diffraction and ''*B*'' is one without diffraction. Moreover, *A*-*B* pairs with-or-without the total computed diffraction address the question ''Is the total diffraction audible?'' for each input signal (Table I). (Such questions, of course, are not posed to the test takers.) A significance level of 0.05 is utilized; thus for the 18 subjects, 13 correct answers of 18 (probability value $p=0.03$, binomial distribution) correspond to the threshold of a significant audible difference. Cases where 11–12 correct answers are given ($p=0.07-0.12$) are only marginally significant and called ''possibly audible'' differences in Table I. Lower ratios are interpreted in Table I to mean ''no'' proven audible difference. The double-blind tests are conducted in a sound-isolated listening room and completed by adults about 20–50 years old, intended to represent a broad sampling of listeners with no known hearing problems. The subjects are trained with an example pair at the beginning and are allowed to listen to each pair in the test as many times as they need before evaluating it. To avoid listener fatigue, the number of test cases is limited.

Results are shown in Table I, where parentheses contain ratios of correct answers for different cases. It is somewhat surprising that although the total computed diffraction in this conservative case may have seemingly small effect on numerical parameters (i.e., low-frequency changes of 1–2 dB), it is nevertheless significant for auralization, for certain input signals. This suggests that conventional numerical parameters are not sufficient to describe the perceived field and that important frequency ranges below the 125-Hz octave band are neglected by room-acoustics quality criteria. Thus although conventional Lambert-diffusion models may be useful for better prediction of room acoustics *parameters* (compared[15] to purely specular modeling), they may not suffice for accurate *auralization*. The tests also show that diffraction is audible even in illuminated positions (i.e., not only in shadow zones) in a conservative geometry for which one might otherwise suspect that edge diffraction is negligible. (The test listeners are not given audiograms, but an expert ''qualified'' jury would presumably hear better, which underscores the significance of the audible cases.)

Audibility depends, of course, on the characteristics of the input signal; e.g., the low-frequency edge-diffraction effects are essentially inaudible for the speech signals, with frequencies concentrated in higher regions, but clearer for the impulse, pink noise, and organ, all with richer spectral content at low frequencies. One may also recall that perception of level differences increases at lower frequencies, where equal loudness contours are more compressed; thus, a given low-frequency increase in decibels corresponds to a proportionally *greater* increase in *loudness level* (phon), which emphasizes the importance of modeling edge-diffraction components, which increase with wavelength. Such low-frequency components are also important for the perceived spaciousness and bass response of the hall.[59] One would expect even greater edge-diffraction effects, extending higher in frequency, from smaller-scale surfaces distributed throughout a realistic concert hall (e.g., with overhead reflectors or irregular wall and ceiling profiles), compared with the stage-house's relatively large, sterile walls.

Neglecting second-order diffraction for this source–receiver orientation has apparently no audible effect, as indicated by the frequency-response curves and listening tests. Nevertheless, second-order diffraction may be audible for shadowed receiver positions, long wavelengths (relative to edge separation), and grazing angles. Although specular/diffractive combinations seem significant in the time- and frequency-domain plots, they appear only marginally significant in listening tests for the tested geometry and input signals. Their frequency effects are considerable but band-limited, and the results also depend on the source–receiver orientation. Further investigation is clearly necessary.

As the frequency responses correspond to a point receiver, the listening tests are monaural. Of course, a full binaural implementation with head-related transfer functions (HRTFs) is necessary for future subjective tests and is outlined in the following section. Nevertheless, the listening tests here still reveal basic coloration effects for source–receiver orientations that involve predominantly frontal HRTF angles and are situated near the centerline, as in our case. Moreover, this study provides some initial insight into the perception of edge diffraction and how one could proceed with future subjective studies, as discussed with the conclusions.

V. PROPOSED BINAURAL MODELING

In addition to computing the RIR, binaural simulation requires the filtering of each transient component (e.g., whether reflection or edge diffraction) with the appropriate

HRTF, in order to model the influence of the head-and-torso scattering on the signals to the left and right ear.[60] However, a brute-force assignment of HRTFs for all of the edge sources along each wedge would be computationally inefficient and presumably subjectively unnecessary, as HRTF magnitude functions (as a rule of thumb) vary slowly over 5–10 deg increments.[61] We thus propose below a practical implementation for binaural simulation of edge diffraction, based on its singular behavior along the least-time path for a given source–wedge–receiver orientation. The method has reasonable mathematical and psychoacoustical basis, and is described to inspire further research in auralization and to encourage the method's practical application in at least lower-resolution (e.g., fast ''virtual-environment'') binaural simulations.

As shown in Fig. 2 of Ref. 21 (see also Fig. 2 in Ref. 30), the edge's ''least-time'' point (or ''apex'' point) lies on the shortest path from the source S to the receiver R via the (infinite) wedge. If this point lies on the actual finite edge, its contributed diffraction amplitude is theoretically singular. This corresponds to the onset time t such that $\sinh[\eta(t)]=0$ in the denominator of Eq. (1) in Ref. 21 (the diffraction IR), after which the amplitude drops and then decays with time (or distance along the edge). The time dependence is initially about $t^{-1/2}$ [see Eqs. (4)–(6) in Ref. 22]. Thus the outer parts of the edge contribute significantly less to the diffraction (except possibly at very low frequencies), as also observed in measurements. Furthermore, the ''law of the first wave front'' (or the precedence effect) also dictates that the initial (onset) arrival dominates the perceived localization. With these considerations in mind, one could use the least-time point on the wedge as a coordinate representing the entire edge relative to the listener, and with this approximation each wedge then requires only one HRTF, whose angle corresponds to the least-time point. If the least-time path for the actual finite wedge does not correspond to that for its (''virtual'') infinite counterpart, the ''initial-incidence point'' (i.e., the endpoint that diffracts first from the finite wedge) could still be used to determine the equivalent HRTF angle.

VI. CONCLUSIONS AND FUTURE WORK

This paper primarily explores the computation of transient edge-diffraction components, including specular/diffractive combinations, in the early room impulse response. Impulse measurements show that specular/diffractive combinations constitute significant components of the total edge diffraction. Moreover, although edge diffraction is often described as the ''bending'' of sound around corners into shadow zones, this study demonstrates its importance as a specific type of *scattering* into *all* regions (i.e., including nonshadow zones) surrounding a wedge. The accurate, time-domain formulation of the BTM-based model allows direct identification of edge diffraction in measured transient data and is suitable for computing impulse responses for auralization. (Also, compared with frequency-domain models, the method is more directly adaptable to programs based on geometrical-acoustics.) For computing binaural impulse responses, one could use the least-time point on a wedge's apex as a coordinate representing the entire edge relative to the listener. With this practical approximation, only one HRTF per wedge is required.

Initial listening tests indicate that, although edge diffraction seems only weakly present (up to 2 dB at low frequencies) for this conservative case, its inclusion is clearly audible even in nonshadow zones, depending on the input signal. It could become even more significant for geometries with smaller surfaces and/or more wedges, and for modeling the perceived spaciousness and bass response of the hall. Thus although further subjective studies are necessary, this indicates that edge diffraction should be modeled in room simulations, although few (if any) auralization programs currently do this. Second-order diffraction appears negligible for nonshadowed positions here but could become more important for long wavelengths (relative to edge separation) and grazing angles.

Future work should include complementary geometries with smaller-scale features (e.g., stage reflectors, profiled walls/ceilings) and shadowed source–receiver combinations (e.g., opera pits). One could also judge the mathematical model's practical range of validity for finite-impedance wedges and investigate how such surfaces could be modeled. Future psychoacoustical studies should determine the subjective effects of reverberation and binaural implementation. One could vary parameters such as reverberation time, wedge-scale/distribution (in room)/packing-density, receiver position (e.g., across different zones), input signal, among other factors. From such studies, one may determine when various edge-diffraction components are audible in auralization and perhaps discern where simplified computations may be employed.

ACKNOWLEDGMENTS

Support was provided by the Axel & Margaret Axson Johnson Foundation, Sweden. Sincere appreciation for helpful discussions and assistance with the listening tests is extended to Bengt-Inge Dalenbäck, Daniel Västfjäll, Vincent Rioux, Portia and John Maisano-Torres, Prances Torres, and all participants in the listening tests. The authors are particularly grateful to Herman Medwin and Asbjørn Krokstad for constructive critique and encouragement. The program CATT-Acoustic was used to compute specular reflections and to post-process signals for auralization. Preliminary results were presented in conference proceedings (Refs. 62, 63) of the ICA/ASA in 1998 and of the EAA in 2000.

[1] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, ''Auralization-an overview,'' J. Audio Eng. Soc. **41**, 861–875 (1993).
[2] H. Kuttruff, ''Auralization of impulse responses modeled on the basis of ray-tracing results,'' J. Audio Eng. Soc. **41**, 876–880 (1993).
[3] J. Borish, ''Extension of the image model to arbitrary polyhedra,'' J. Acoust. Soc. Am. **75**, 1827–1836 (1984).
[4] A. Krokstad, S. Strøm, and S. Sørsdal, ''Calculating the acoustical room response by the use of a ray-tracing technique,'' J. Sound Vib. **8**, 118–125 (1968).
[5] B.-I. Dalenbäck, ''Verification of prediction based on Randomized Tail-Corrected cone-tracing and array modeling,'' Proc. of 137th Meeting of ASA, 2nd Convention of EAA (Forum Acusticum 99) and the 25th DAGA conference, Berlin (1999).
[6] B.-I. Dalenbäck, ''Room acoustic prediction based on a unified treatment

of diffuse and specular reflection,'' J. Acoust. Soc. Am. **100**, 899–909 (1996).
[7] M. Vorländer, ''Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm,'' J. Acoust. Soc. Am. **86**, 172–178 (1989).
[8] J. J. Embrechts, ''Broad spectrum diffusion model for room acoustics ray-tracing algorithms,'' J. Acoust. Soc. Am. **107**, 2068–2081 (2000).
[9] I. A. Drumm and Y. W. Lam, ''The adaptive beam-tracing algorithm,'' J. Acoust. Soc. Am. **107**, 1405–1412 (2000).
[10] G. M. Naylor, ''ODEON—Another hybrid room acoustical model'' Appl. Acoust. **38**, 131–143 (1993).
[11] D. Van Maercke and J. Martin, ''The prediction of echograms and impulse responses within the Epidaure software,'' Appl. Acoust. **38**, 93–114 (1993).
[12] T. Lewers, ''A combined beam tracing and radiant exchange computer model of room acoustics,'' Appl. Acoust. **38**, 161–178 (1993).
[13] J. P. Chambers and Y. H. Berthelot, ''Time-domain experiments on the diffraction of sound by a step discontinuity,'' J. Acoust. Soc. Am. **96**, 1887–1892 (1994).
[14] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, NY, 1989), pp. 488–501.
[15] M. Vorländer, ''International round robin on room acoustical computer simulations,'' Intl. Congress on Acoustics (ICA), Trondheim, Norway, Proceedings Vol. II, pp. 689–692 (1995).
[16] ISO/DIS 3382 (1995).
[17] L. Beranek, *Concert and Opera Halls: How They Sound* (Acoustical Society of America, New York, 1996), pp. 567–575.
[18] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, ''Creating interactive virtual acoustic environments,'' J. Audio Eng. Soc. **47**, 675–705 (1999).
[19] J. Blauert, H. Lehnert, J. Sahrhage, and H. Strauss, ''An interactive virtual-environment generator for psychoacoustic research. I: Architecture and implementation,'' Acust. Acta Acust. **86**, 94–102 (2000).
[20] R. R. Torres, M. Kleiner, and B.-I. Dalenbäck, ''Audibility of 'diffusion' in room acoustics auralization: An initial investigation,'' Acust. Acta Acust. **86**, 919–927 (2000).
[21] U. P. Svensson, R. I. Fred, and J. Vanderkooy, ''An analytic secondary source model of edge diffraction impulse responses,'' J. Acoust. Soc. Am. **106**, 2331–2344 (1999).
[22] H. Medwin, ''Shadowing by finite noise barriers,'' J. Acoust. Soc. Am. **69**, 1060–1064 (1981).
[23] H. Medwin, E. Childs, and G. M. Jebsen, ''Impulse studies of double diffraction: A discrete Huygens interpretation,'' J. Acoust. Soc. Am. **72**, 1005–1013 (1982).
[24] J. B. Keller, ''Geometrical theory of diffraction,'' J. Opt. Soc. Am. **52**, 116–130 (1962).
[25] R. G. Kouyoumijian and P. H. Pathak, ''A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface,'' Proc. IEEE **62**, 1448–1461 (1974).
[26] A. D. Pierce, ''Diffraction of sound around corners and over wide barriers,'' J. Acoust. Soc. Am. **55**, 941–955 (1974).
[27] E. M. Salomons, ''Sound propagation in complex outdoor situations with a nonrefracting atmosphere: model based on analytical solutions for diffraction and reflection,'' Acustica **83**, 436–454 (1997).
[28] 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**, 2973–2986 (2000).
[29] A. J. Berkhout, D. de Vries, J. Baan, and B. W. van den Oetelaar, ''A wave field extrapolation approach to acoustical modeling in enclosed spaces,'' J. Acoust. Soc. Am. **105**, 1725–1733 (1999).
[30] R. S. Keiffer and J. C. Novarini, ''A time domain rough surface scattering model based on wedge diffraction: Application to low-frequency backscattering from two-dimensional sea surfaces,'' J. Acoust. Soc. Am. **107**, 27–39 (2000).
[31] H. Kuttruff, *Room Acoustics*, 3rd ed. (Elsevier Applied Science, London and New York, 1991), pp. 84–85, 110.
[32] B.-I. Dalenbäck, M. Kleiner, and P. Svensson, ''A macroscopic view of diffuse reflection,'' J. Audio Eng. Soc. **42**, 793–807 (1994).
[33] B.-I. Dalenbäck, Manual for *CATT-Acoustic* Version 7.2, Section 2.5, ''Automatic edge diffusion'' (1999) (''Edge diffusion,'' however, was implemented by Dalenbäck as early as 1993 in CATT-Acoustic version 4.0).
[34] A. Farina, ''Validation of the numerical simulation of the scattered sound field with a geometrical pyramid tracing approach,'' Proc. Inst. Acoust. **22**, 315–322 (2000).
[35] H. Medwin and C. S. Clay, *Fundamentals of Acoustical Oceanography* (Academic, New York, 1998), p. 24.
[36] C. S. Clay, D. Chu, and S. Li, ''Specular reflections of transient pressures from finite width plane facets,'' J. Acoust. Soc. Am. **94**, 2279–2286 (1993).
[37] J. H. Rindel, ''Attenuation of sound reflections due to diffraction,'' Proc. of Nordic Acoustical Meeting (NAM) 1986, pp. 257–260 (1986).
[38] Y. Sakurai and K. Nagata, ''Sound reflections of a rigid plane panel and of the 'live end' composed by those panels,'' J. Acoust. Soc. Jpn. (E) **2**, 5–14 (1981).
[39] C. S. Clay and W. A. Kinney, ''Numerical computations of time-domain diffractions from wedges and reflections from facets,'' J. Acoust. Soc. Am. **83**, 2126–2133 (1988).
[40] T. J. Cox and Y. W. Lam, ''Evaluation of methods for predicting the scattering from simple rigid panels,'' Appl. Acoust. **40**, 123–140 (1993).
[41] G. M. Jebsen and H. Medwin, ''On the failure of the Kirchhoff assumption in backscatter,'' J. Acoust. Soc. Am. **72**, 1607–1611 (1982).
[42] G. V. Norton, J. C. Novarini, and R. S. Keiffer, ''An evaluation of the Kirchhoff approximation in predicting the axial impulse response of hard and soft disks,'' J. Acoust. Soc. Am. **93**, 3049–3056 (1993).
[43] M. A. Biot and I. Tolstoy, ''Formulation of wave propagation in infinite media by normal coordinates with an application to diffraction,'' J. Acoust. Soc. Am. **29**, 381–391 (1957).
[44] R. S. Keiffer, J. C. Novarini, and G. V. Norton, ''The impulse response of an aperture: Numerical calculations within the framework of the wedge assemblage method,'' J. Acoust. Soc. Am. **95**, 3–12 (1994).
[45] J. Novarini and R. S. Keiffer, ''Impulse response of a density contrast wedge: Practical implementation and some aspects of its diffracted component,'' Appl. Acoust. **58**, 195–210 (1999).
[46] A. M. J. Davis, ''The complete extension of the Biot–Tolstoy solution to the density contrast wedge with sample calculations,'' J. Acoust. Soc. Am. **101**, 1821–1835 (1997).
[47] D. Ouis, ''Scattering by a barrier in a room,'' Appl. Acoust. **56**, 1–24 (1999).
[48] D. Ouis, ''Scattering of a spherical wave by a thin hard barrier on a reflecting plane,'' Appl. Acoust. **59**, 19–66 (2000).
[49] P. S. Kovitz, ''Extensions to the image method model of sound propagation in a room,'' Ph.D. dissertation, Penn State, State College, PA, 1994, pp. 1–203.
[50] G. Wadsworth and J. P. Chambers, ''Scale model experiments on the insertion loss of wide and double barriers,'' J. Acoust. Soc. Am. **107**, 2344–2350 (2000).
[51] R. S. Keiffer, J. C. Novarini, G. V. Norton, and J. R. Dubberley, ''Benchmarking the wedge-assemblage method,'' Math. Modell. Sci. Comput. **4**, 414–419 (1994).
[52] H. G. Jonasson, ''Sound reduction by barriers on the ground,'' J. Sound Vib. **22**, 113–126 (1972).
[53] H. Lee and B.-H. Lee, ''An efficient algorithm for the image model technique,'' Appl. Acoust. **24**, 87–105 (1988).
[54] F. Giron, ''Investigations about the directivity of sound sources,'' Ph.D. dissertation, Ruhr-Universität, Bochum, Shaker Verlag, Aachen, Germany, 1996.
[55] B.-I. Dalenbäck, *CATT-Acoustic*, CATT, Mariagatan 16A, SE-41471 Gothenburg, Sweden (http://www.catt.se).
[56] J.-D. Polack, X. Meynial, and V. Grillon, ''Auralization in scale models: Processing of impulse response,'' J. Audio Eng. Soc. **41**, 939–945 (1993).
[57] N. Xiang and J. Blauert, ''Binaural scale modeling for auralisation and prediction of acoustics in auditoria,'' Appl. Acoust. **38**, 267–290 (1993).
[58] X. Meynial and O. Vuichard, ''Objective measure of sound colouration in rooms,'' Acustica **85**, 101–107 (1999).
[59] J. S. Bradley, R. D. Reich, and S. G. Norcross, ''On the combined effects of early- and late-arriving sound on spatial impression in concert halls,'' J. Acoust. Soc. Am. **108**, 651–661 (2000).
[60] B.-I. Dalenbäck, ''A new model for room acoustic prediction and auralization,'' Ph.D. dissertation, Dept. of Applied Acoustics, Chalmers Univ. of Tech., Gothenburg, Sweden, 1995 (ISBN 91-7197-200-5).
[61] D. Begault and E. Wenzel, personal communication.
[62] R. R. Torres and M. Kleiner, ''Audibility of edge diffraction in auralization of a stage house,'' Proc. ICA **98**, 371–372 (1998).
[63] R. R. Torres, U. P. Svensson, M. Kleiner, ''Edge diffraction in room acoustics computations,'' Proc. of EAA Symposium on Architectural Acoustics, Madrid (October 2000).

Determination of the complex Young and shear dynamic moduli of viscoelastic materials

R. Lance Willis,[a] Lei Wu, and Yves H. Berthelot[b]
Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

(Received 10 August 2000; accepted for publication 28 November 2000)

The Young and shear dynamic moduli of viscoelastic materials are determined from laser vibrometric measurements of the surface motion of a three-dimensional sample excited by a piezoelectric actuator inside a chamber with controllable temperature and static pressure. The moduli are estimated from an inversion code that minimizes the difference between the data and the predictions from a finite element model in which the elastic moduli are the adjustable parameters. The technique is first used to measure the dynamic properties of homogeneous samples and the results are compared with those obtained by the standard rod resonance technique. Results are then obtained with microvoided samples in the 0.5–3 kHz frequency range, at temperatures ranging from 7 to 40 °C, and static pressures ranging from ambient to 34 atm (3.45 MPa or 500 psi). The limitations of the technique are discussed. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1342003]

PACS numbers: 43.58.Dj, 43.35.Mr, 43.35.Cg [SLE]

I. INTRODUCTION

Several methods and instruments have been devised for the measurement of the dynamic mechanical properties of viscoelastic materials as summarized in a comprehensive report by Capps.[1] Some of the most common measurement techniques are the rod resonance apparatus of Madigosky and Lee[2,3] which was adopted as an ANSI standard,[4] the mass loaded method,[5] and the cantilever beam technique[6] or Dynamic Mechanical Thermal Analyzer (DMTA), which is commercially available.[7] In the resonant apparatus of Madigosky and Lee, a longitudinal wave is transmitted down a rod of material with accelerometers attached at both ends. The complex Young's modulus is calculated at specific frequencies from the phase interference of the extensional waves in the rod. The technique works well throughout the audio-range, but it is limited to one-dimensional sample geometry, a discrete set of frequencies, and it measures only the Young's modulus. In the mass loaded method,[5] the test specimen is loaded with a mass and excited by a shaker. An impedance head measures the force and acceleration from which the complex Young's modulus can be obtained. In the DMTA apparatus,[7] the samples are very small (less than a few square centimeters), two-dimensional, clamped elements excited at very low frequencies (usually at a few hertz) inside a small temperature-controlled chamber. This indirect method relies on the time–temperature superposition principle[8] and the subsequent shift factors required to construct a master curve over a broad frequency range from measurements taken at multiple temperatures. Some inverse techniques have been developed for the case of samples submerged in large water tanks. For instance, transmission and reflection of sound incident at oblique incidence on flat panels can be used to deduce material properties although at low frequencies diffraction by the edges of the sample can cause serious errors. Piquette[9] has used perfectly spherical elastomeric targets to compute the bulk modulus by minimizing the difference between the measured scattering coefficients and those predicted analytically. Although it is a truly noncontact technique, it is restricted to the case where the bulk modulus is greater than the shear modulus, which may not be the case for some materials.

The techniques just mentioned have some limitations and there is a need to develop an independent method, capable of measuring two moduli on a three-dimensional sample. In an earlier paper, Willis *et al.*[10] proposed a new technique based on a combination of laser-based imaging of the surface vibrations of a three-dimensional sample and an optimization scheme involving finite element modeling of the sample. The method consists in measuring by laser vibrometry the dynamic response of a sample excited by a shaker, and matching the data with predictions from a finite element code in which the complex elastic moduli are the adjustable parameters. In principle, this inverse technique enables the determination of two complex moduli on the same sample from one set of measurements. Results were obtained[10] with viscoelastic samples containing microinclusions, in the low kilohertz frequency range, at ambient pressure and temperature. The presence of microinclusions in the sample enables the conversion of compressional strain energy into shear energy, which significantly increases losses in the longitudinal mode through dissipation into heat.[11] There is a very large body of literature concerning models predicting the behavior of viscoelastic materials with various types of inclusions that are much smaller than a wavelength. One of the earliest models is Kerner's quasistatic, effective medium theory.[12] Many researchers have proposed self-consistent, effective medium theories,[13–21] including some multiple scattering theories.[22,23] However, the laser-based finite element model (FEM) method introduced in Ref. 10, which is expanded in this paper, does not attempt to model

[a]Current address: Motorola, Inc., 8000 West Sunrise Blvd., Plantation, FL 33322.
[b]Electronic mail: yves.berthelot@me.gatech.edu

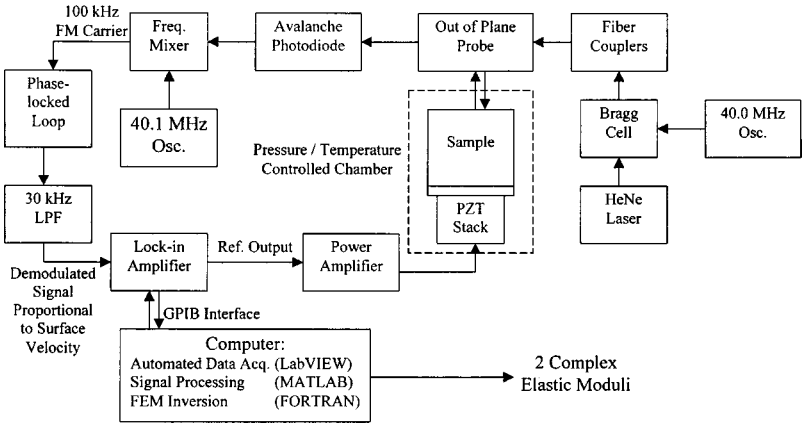


FIG. 1. Block diagram of the experiment.

the micromechanics of the scattering problem inside the voided material. Instead, the sample is modeled in a functional manner as an effective medium that is globally homogeneous and isotropic, with effective material properties. The objective of this paper is first to test the validity of the laser-based FEM technique using well-characterized homogeneous samples, and second to present results obtained with microvoided samples in the 0.5–3 kHz frequency range, at temperatures ranging from 7 to 40 °C, and static pressures ranging from ambient to 34 atm (3.45 MPa or 500 psi). The procedure is described in Sec. II. The calibrations are presented in Sec. III. The results obtained with microvoided samples are shown and discussed in Sec. IV.

II. PROCEDURE

A. Experimental arrangement

A block diagram of the system, shown in Fig. 1, illustrates the sample excitation system, the pressure or temperature controlled chamber, the laser interferometric measurement instrumentation, the data acquisition, and the numerical steps necessary to obtain the elastic moduli of the material. Figure 2 shows the excitation system inside the pressure chamber. A sample, generally on the order of $25\times50\times75$ mm in size, is bonded on one side to an aluminum base driven harmonically by a $25.4\times25.4\times25.4$ mm PZT stack (custom made by Sensor Technology). The stack is bonded to a massive steel base inside the chamber with a thin layer of uncured butyl rubber. An isolation pad is inserted between the steel base and the lower end-cap of the chamber. The chamber is constructed from a section of stainless steel pipe with inner diameter of 69.9 mm and a wall thickness of 20.3 mm. The inside of the chamber is 203 mm in height. The bottom of the chamber is a 15.9-mm-thick stainless steel end-cap that is welded to the pipe. The top of the chamber is fitted with a flange. At the top of the chamber, a 19.1-mm-thick glass end-cap with an O ring and a rubber gasket is inserted between the flange and the outer stainless steel ring bolted to the flange. The sample is bonded to the baseplate with either a cyanoacrylate adhesive or a thin layer (0.6 mm) of uncured butyl rubber. The cyanoacrylate adhesive is used for the samples whose volume does not change during the experiment. The butyl rubber bond is used to reduce lateral strains at the base of samples that shrink substantially at high static pressures. In that case, the sample is allowed to creep for several hours until it reaches its final dimensions. The temperature of the chamber is controlled within 0.5 °C by partially submerging it in a regulated water bath. The temperature of the experiment is monitored with a thermistor placed inside the chamber. Experiments have been conducted in the 7–40 °C temperature range, 0–500 psi (0–34 atm) pressure range, and 300 Hz–5 kHz frequency range. At higher frequencies, the motion of the base plate is not that of a uniform piston as modeled in the finite element code.

The laser vibrometers used for measuring the surface velocity of the sample utilize 10 mW He–Ne lasers in a heterodyne setup similar to those described in a previous paper.[10] The actual probe head is located outside the chamber and the laser beams pass through the transparent glass at the top of the chamber. Although only an out-of-plane probe

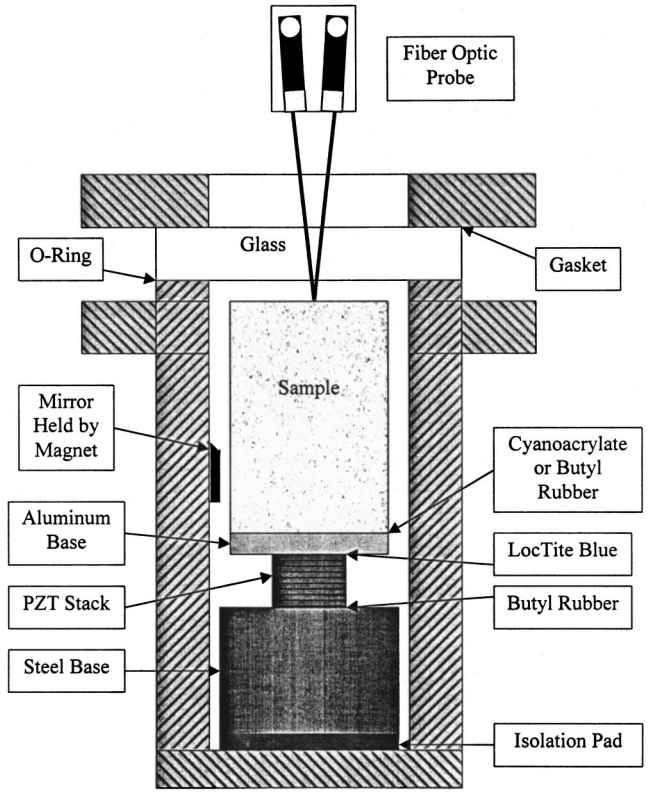


FIG. 2. Sample and base excitation inside the chamber.

is used for measurements inside the chamber, an in-plane probe has been constructed as well for use with the ambient measurements. In order to obtain measurements on the side of the sample inside the chamber, a small mirror consisting of a piece of mylar is attached to the side of the chamber using a magnet. The laser beam passes through the glass window at the top of the chamber and propagates downward between the wall of the chamber and the sample until it hits the small mirror. It is deflected toward the sample to measure the normal component of the surface velocity on the side of the sample. The experiment is automated via a general purpose interface bus (GPIB) to a computer running LABVIEW. A lock-in amplifier (Stanford Research Systems SR830) with an internal signal generator is controlled by LABVIEW to step through a range of frequencies with a fixed increment. The signal is fed through a power amplifier (Crown D-150A Series II) which drives the PZT stack. The typical peak amplitude of the pistonlike motion of the base is on the order of 50 nm. The demodulated signal from the vibrometer, which is the instantaneous surface velocity of the sample, is analyzed by a lock-in amplifier to resolve its real and imaginary parts using the internal signal generator as a phase reference. The data are then exported through the GPIB interface to the computer. After the experiment, the data are processed using MATLAB to reference the amplitude and phase of the measured signals to that of the base motion. Finally, the data are imported into a FORTRAN inversion code to determine the complex elastic moduli of the material sample.

B. Finite element model and numerical inversion

At each frequency, the elastic moduli are determined from an inversion code that minimizes the difference between the data (real and imaginary parts of the surface velocity) and the predictions obtained from a finite element model (FEM) in which the elastic moduli are the adjustable parameters. The inversion code searches for the values of E and G that minimize the sum of the residuals at all N measurement points, each residual being the squared differences between the measured and predicted surface velocities (real part and imaginary part) referenced to the base motion. Best results were obtained with absolute differences. Indeed, if the residuals of the real and imaginary parts are expressed as relative residuals of the form $[(X_{\mathrm{FEM}}-X_{\mathrm{data}})/X_{\mathrm{FEM}}]^2+[(Y_{\mathrm{FEM}}-Y_{\mathrm{data}})/Y_{\mathrm{FEM}}]^2$, where X stands for the real part and Y for the imaginary part, errors can be greatly magnified if the imaginary part is very small. In many situations of practical interest, the losses in the material are sufficiently high that it is best to use absolute differences in the residuals. Two standard optimization techniques are used: the downhill simplex method or Powell's method.[24] An option in the optimization is to add the constraint that, in the small frequency range of interest, the moduli must increase linearly with frequency. This constraint artificially smoothes the frequency dependence of the results. Each evaluation requires only function evaluations, not derivatives. At a given frequency, the optimization typically takes 5 min with the 450 MHz processor. In most of the results presented in this paper, the FEM code defines 275 nodes and 36 quadratic elements for the quarter symmetry of the sample: 6 elements along the height, 3 along the half-length, and 2 along the half-width. The base motion is assumed to be uniform and all other surfaces of the sample are modeled as free surfaces. The material is assumed to be globally homogeneous and isotropic, so that its dynamic response is fully described by only two complex moduli (e.g., the Lamé constants λ and μ, or the complex Young and shear moduli, E and G). Departure from these conditions in the experiment may cause the inversion code not to converge or to converge toward erroneous values of the moduli. A single FEM evaluation of the surface dynamics at all the nodes takes about 1 s on a 450 MHz Pentium processor. For our given application, no substantial computational savings were gained by using higher order finite element techniques. Finite difference techniques were found to be computationally expensive compared with the FEM approach. Also, evaluation of the surface dynamics based on the analytical model of a rectangular brick[25] was found to be numerically less efficient than the FEM technique. In addition, it restricts the sample shape to a rectangular geometry whereas the FEM does not.

The optimization requires initial guesses of the moduli to start the search. It is recommended to start at a frequency that corresponds to the first or second mode of the sample, defined as the first or second peak in velocity amplitude (normalized to that of the base) measured at the top of the sample. At that frequency, the result of the inversion is not sensitive to the choice of the initial guesses as long as they are within the correct order of magnitude. The result of the inversion at a given frequency f is then used as the initial guess for the next frequency $f+\Delta f$. Typically, a frequency increment $\Delta f=25\,\mathrm{Hz}$ is used. If the material properties are known (properties of the polymer, void fraction), initial guesses can also be obtained from Kerner's model.[12] If the material is completely unknown, static tests could be performed to find the order of magnitude of the moduli.

C. Selection of the moduli

In principle, the optimization can be implemented to search for any combination of two complex material parameters: Young, shear, bulk, plane-wave moduli (or corresponding sound speeds), or Poisson's ratio. However, the surface topography[26] of the objective function in the parameter space is critical to achieving robust results in the inversion code. For instance, it is in general difficult to get a reliable value of Poisson's ratio. Indeed, the topography of objective function using Poisson's ratio as one of the parameters is flat and sensitive to noise in the data. Similarly, in most cases it is difficult to get robust inversion of the bulk modulus. It turns out that the objective function is such that the best results are obtained with the complex Young and shear moduli, E and G.

As discussed in Ref. 10, it is reasonable to assume that, in the frequency range of interest, the Poisson ratio is real and independent of frequency. This assumption has been validated with some experiments[1,27] that revealed that the imaginary part of the Poisson ratio of elastomers is less than 1% of its real part. Physically, it means that, in the voided samples under consideration, all the losses are due to the conversion of dilatational energy to shear energy at the

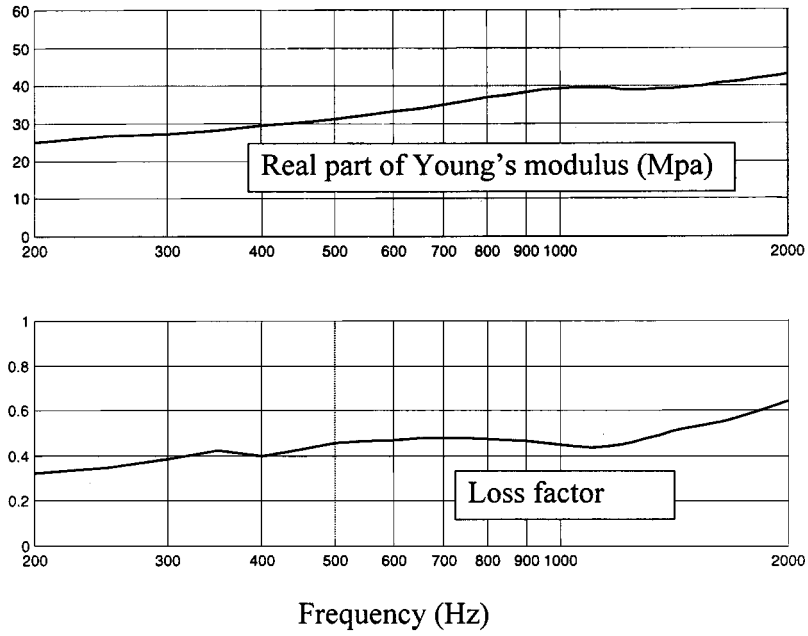


FIG. 3. Young's modulus and loss factor of a homogeneous polyurethane measured by the laser-based/numerical method [$T=23\,^{\circ}\mathrm{C}$ and $P=0$ psi (gauge)].

boundaries of the inclusions, with a corresponding dissipation into heat in the polymer.[11] This assumption also implies that the loss factors, η, of E, G, K, λ, and μ are the same since the complex moduli are related by the well-known expressions $E=G/2(1+\nu)$, $K=E/3(1-2\nu)$, $G=\mu$, $\lambda=3K\nu/(1+\nu)$, $K=2/3G(1+\nu)/(1-2\nu)$. The assumption of a real-valued Poisson ratio is also consistent with Kerner's model since, in this model, the Poisson ratio can be directly expressed as a real-valued function of the void fraction of the material, as indicated by Eqs. (7) and (9) of Ref. 10. Based on the above-mentioned arguments, it follows that any differences in the loss factors of the elastic moduli can be used to determine indirectly the imaginary part of Poisson's ratio. It should be noted that a small imaginary part of the Poisson ratio may have a very significant effect on the ratio of the shear to bulk loss factors. However, in this paper, the Poisson ratio is assumed to be real so that the optimization routine is reduced to a three-dimensional search for the global minimum, in E', G', and η, where $E=E'(1+i\eta)$ and $G=G'(1+i\eta)$.

If the proposed method enables the determination of E and G for globally homogeneous and isotropic samples, it is not, in general, capable of giving reliable estimates of the bulk modulus, K, or the plane wave modulus M, from E and G, using the standard equations:

$$K=\frac{EG}{3(3G-E)}=\frac{E}{3(1-2\nu)},$$
$$M=G\left(\frac{4G-E}{3G-E}\right)=\frac{E(1-\nu)}{(1+\nu)(1-2\nu)} \tag{1}$$

unless the Poisson ratio, ν, is known or determined precisely. For most materials of interest, the Poisson ratio is between 0.4 and 0.5, so that the quantity $(1-2\nu)$ in the denominators of Eq. (1) ranges from 0 to 0.2. Since $\nu=(E/2G)-1$, small errors in E and G are considerably magnified in the calculated values of K and M based on Eq. (1). Therefore, the proposed inversion technique is in general better suited for the determination of the Young and shear moduli.

III. CALIBRATIONS

The technique was validated with homogeneous samples machined from a translucent block of material. A rectangular block of dimensions $30.40\times51.64\times81.13$ mm (density 1077 kg/m^3) was machined for the laser-based/FEM experiment. A rectangular rod (7.2 cm long and 0.6 cm by 0.6 cm in cross section) from the same block of material was used to provide comparative data from the standard resonant rod technique of Madigosky and Lee.[3] In the laser-based/FEM experiment, the surface velocities (amplitude and phase) were measured with the vibrometer at four points—three points at the top of the sample and one point on its side. The base velocity (amplitude and phase) was also measured for reference. The results of the inversion code are shown in Fig. 3. For brevity, only the Young modulus is shown. The frequency dependence is relatively smooth, with a slight positive slope in frequency for both the real part and the loss factor, which is to be expected in this frequency range. As expected for these homogeneous polyurethanes, the Poisson ratio is between 0.49 and 0.5. The inversion code converges at all frequencies (taken in 25 Hz steps) throughout the frequency range.

For comparison, the Young's modulus of the sample was determined from the resonant technique[3] with a 7.2-cm-long rod excited at one end by a shaker. Instead of using accelerometers at both ends, the complex velocities at both ends of the samples were measured directly with a vibrometer (Polytec PSV-200). The absence of mass loading considerably simplifies the equations used to obtain the complex Young's modulus and makes the experiment less susceptible to spurious flexural waves in the sample. If U_0 denotes the complex velocity of the excitation and U denotes the complex velocity of the free end, the ratio of U_0/U is $\cos(kL)$, where $k=(\rho\omega^2/E)^{1/2}$ is the complex extensional wave num-

TABLE I. Comparison of Young's modulus and loss factor of a homogeneous sample by the resonant rod technique [one-dimensional (1D)] and the block [three-dimensional (3D)] laser-based/FEM technique.

	Resonant rod (1D) Newton–Raphson	Resonant rod (1D) FEM+inversion code	Block (3D) FEM+inversion code
Young's modulus E (MPa)	39.9	38.9	34.1
Loss factor η	0.45	0.47	0.48

ber. Once the complex ratio U_0/U is determined from amplitude and phase measurements at both ends of the rod, the complex Young's modulus E is found using a two-dimensional Newton–Raphson technique. At 23 °C, the first resonance was observed at 663 Hz. At that frequency, the real part of Young's modulus was 40 MPa and the loss factor was 0.45. To test the validity of the FEM code and the optimization routine discussed in Sec. II B, the data obtained with the rod were also used to compute the complex Young's modulus with the FEM code and the inversion routine. The results are shown in Table I. The computations were made with 12 quadratic elements, i.e., 152 nodes. As expected, there is good agreement between the FEM approach and the Newton–Raphson inversion. The results can then be compared with those obtained with the three-dimensional block of the same material. Several factors may explain the differences seen in Table I. First, the experiment with the resonant rod was not performed in a temperature-controlled chamber. It was performed in an open environment where the ambient temperature was maintained at 23 °C within about 1 °C by a portable heater. (The ambient temperature without the heater was 17 °C). It turns out that the dynamic Young's modulus for this material is extremely sensitive to temperature in the frequency range under investigation. (For instance, at 17 °C, the Young's modulus was about 60 MPa). It should also be pointed out that because the rod was made from the block of material, its length was limited to about 7.2 cm. Due to manufacturing difficulties encountered with this soft material, the cross section was only 0.6 cm by 0.6 cm, and not perfectly constant along the sample length, so that the assumption of a thin rod is questionable in the experiment. Another possible explanation for the discrepancy is that the experiment with the block preceded the experiment with the rod by six months and it is possible that some aging may have occurred in the sample. Within experimental errors, it appears that the Young's moduli are comparable when measured by the laser-based/FEM technique or the resonant rod method.

Another way to analyze the results obtained with the laser-based/FEM method is to plot the real part of Young's modulus as a function of the real part of the shear modulus. Figure 4 shows such a plot for the case of ambient conditions. The slope E/G is 3 which, as expected, corresponds to a Poisson ratio of 0.5. The linearity of the slope E/G is a good indicator of the quality of the data and it confirms that the Poisson ratio is independent of frequency in the range under consideration. The experiment was repeated at static pressures ranging from 0 to 300 psi and the results of the inversion code are shown in Fig. 5.[28] The dependence on static pressure, albeit very small, shows the right trend in the data, with a small increase in stiffness at higher pressure which can be expected in these samples without inclusions. The sudden jump in Young's modulus at 950 Hz is unphysical and represents a bad data point.

The method was also tested on a block of Lucite™ with dimensions 305.3 mm×204.5 mm×24.8 mm. The density

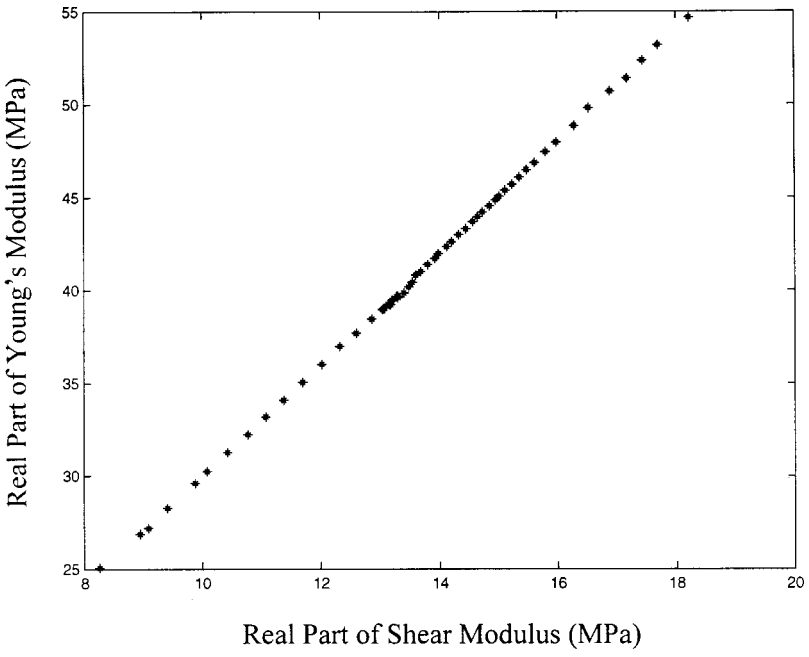


FIG. 4. Young's modulus (real part) versus shear modulus (real part), at $P=0$ psi (gauge), and $T=23$ °C, for the homogeneous sample The slope of E/G is 3, which corresponds to a Poisson ratio of 0.5.

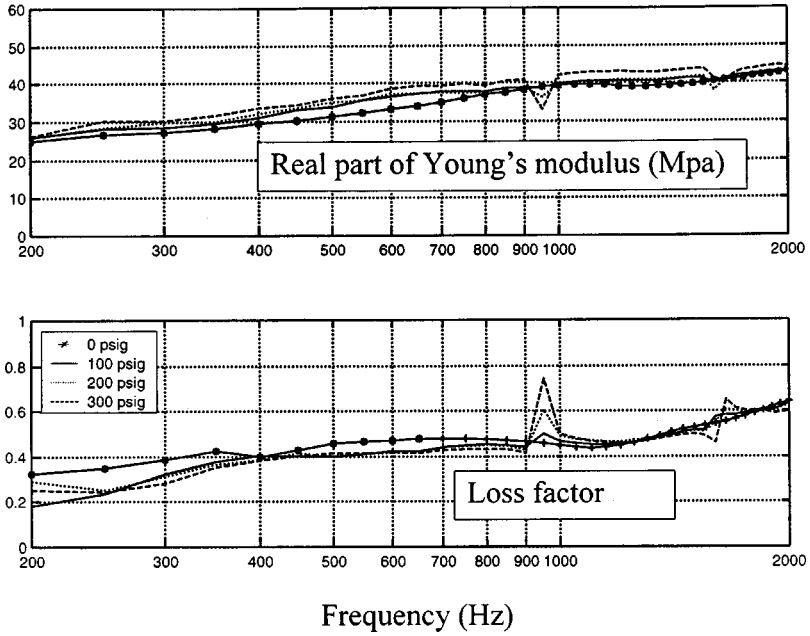


FIG. 5. Young's modulus and loss factor of a homogeneous polyurethane measured by the laser-based/numerical method ($T=23\,^{\circ}\mathrm{C}$) at $P=0$, 100, 200, and 300 psi (gauge).

was 1188 kg/m^3. The height requirement was dictated by the high value of the elastic moduli and the corresponding sound speed in Lucite, which results in long wavelengths. Due to the size of the sample, this particular experiment was conducted with the five-probe vibrometer system that incorporates a Ling Dynamics Systems 556 shaker, as described in Ref. 10. The surface velocity was measured at several locations on the sample between 1000 and 2500 Hz. Below 1000 Hz, the wavelengths are so large that the block starts to behave like a rigid body and, as expected, the inversion code has poor convergence. The results of the inversion code indicate that the real part of Young's modulus is $E'=4.8\,\mathrm{GPa}$, the loss factor is $\eta=0.05$, and the Poisson ratio is $\nu=0.33$, ($G'=1.8\,\mathrm{GPa}$), with no discernable frequency dependence in that frequency range. By comparison, results obtained by Lagakos *et al.*[29] with a resonant technique using accelerometers at both ends of a cylindrical rod loaded with a mass on the top indicate that the Young's modulus of Lucite can be approximated by $E'=0.509\log_{10} f+3.790$, which corresponds to $E'=5.3\,\mathrm{GPa}$ at 1 kHz and $E'=5.5\,\mathrm{GPa}$ at 2.5 kHz, with very small losses in that frequency range. These values of E' are within 11% of the results obtained with the laser-based method.

IV. RESULTS WITH SAMPLES CONTAINING MICROINCLUSIONS

A. Sample shape

First, a rectangular block 76.2 mm×51 mm×31 mm (referred to as sample 1) was used. The measured density was 570 kg/m^3. The result of the inversion code is shown in Fig. 6. For brevity, results obtained with the shear modulus are

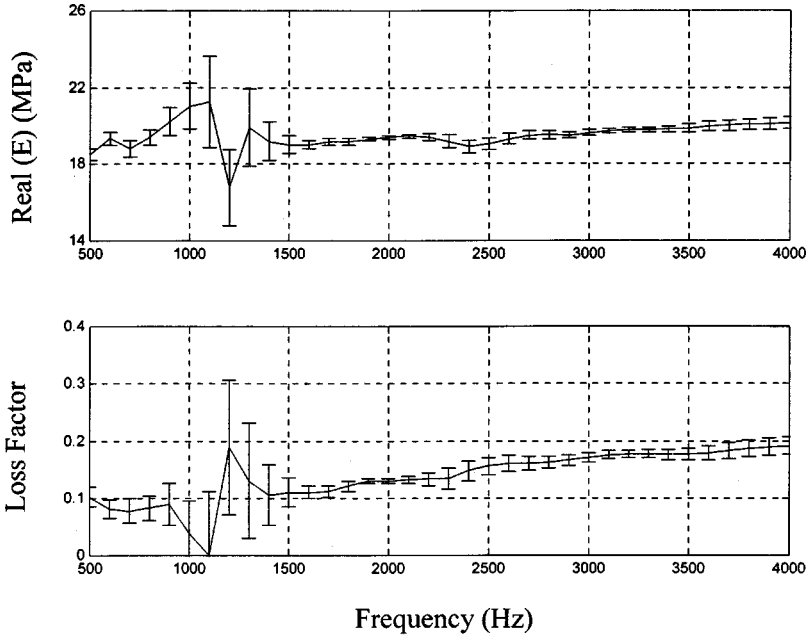


FIG. 6. Young's modulus and loss factor of a voided sample (sample 1) of rectangular shape, measured at ambient conditions ($P=1$ atm, $T=23\,^{\circ}\mathrm{C}$). The error bars indicate the variance of the estimates. The convergence is poor between the first and second modes.

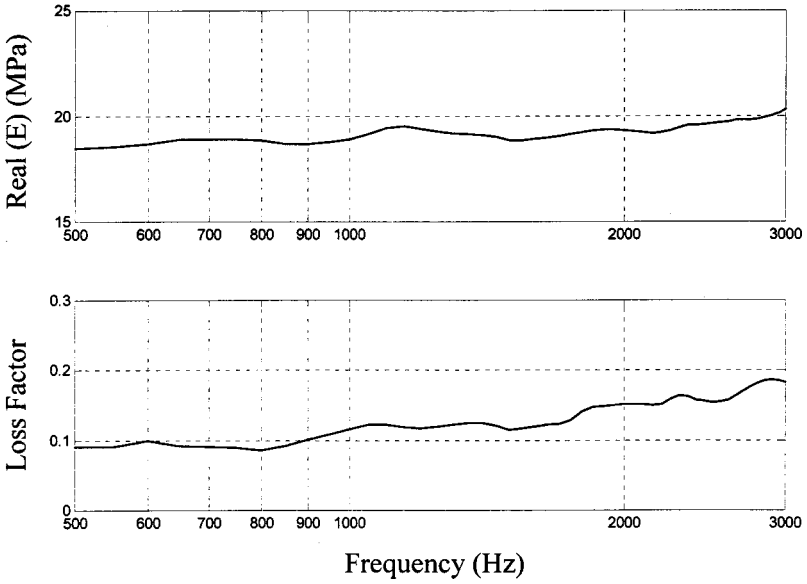


FIG. 7. Young's modulus and loss factor of a voided sample (sample 2) of U shape, measured at ambient conditions ($P=1$ atm, $T=23\,^{\circ}\mathrm{C}$).

not included. The error bars were obtained[30] by computing the variance–covariance matrix of the parameter estimates[31] as explained in the Appendix. Large error bars indicate that the inversion routine is insensitive to material properties. Clearly, the inversion code does not converge well in the 900–1500 Hz frequency range. A careful analysis of the measured complex velocities reveals that, in that frequency range, all the measured surface velocities are approximately either in phase or out of phase with the base velocity because this frequency band falls just between the first two longitudinal modes of the block sample. This frequency band is referred to as the ''dead zone.'' Despite reliable surface velocity data at each point, the inversion code leads to unreliable estimates between modes. Buchanan[32] has investigated this limitation in detail for the case of a one-dimensional sample and concluded that little could be done about it. If data are needed in that particular frequency range, it is necessary to repeat the experiment with a sample of different height. For lossy materials, such as the polyurethane used for the calibrations (Sec. III), the frequency dependence of the phase is such that the ''dead zone'' is rather small. Indeed, good convergence was observed with the calibrated samples.

In order to reduce the frequency range where the inversion code does not converge well, a second experiment was done with a ''U-shaped'' sample (referred to as sample 2). The sample was milled from a block the same size as sample 1, with a notch 1/3 of the width and 1/3 of the height cut out of the top through the thickness, giving the appearance of a ''U.'' The result was to increase the modal density between the first and second longitudinal modes so as to improve the convergence of the inversion code. The results of the inversion code (Young's modulus and loss factor) are plotted in Fig. 7. Again, for brevity, results obtained with the shear modulus are not included. Reasonably good convergence is observed throughout the frequency range with the U-shaped sample (sample 2) for both the Young and shear moduli.

To give a qualitative representation of the combined measurement repeatability and sample to sample variability, experiments were conducted with two other rectangular blocks of the same material and same cross section as with samples 1 and 2 but with different lengths. The lengths were 62.6 mm (sample 3) and 70.0 mm (sample 4), respectively. The dynamic Young's moduli and loss factors of all four samples are shown in Fig. 8. As discussed earlier, poor convergence is observed with the rectangular samples (samples 1, 3, and 4) between the first and second modes. Outside this frequency range, the measurement repeatability is acceptable considering the inherent sample to sample variability of these voided materials.

B. Temperature dependence

The U-shaped sample (sample 2) was used to obtain data at 7, 15, 21, and 40 °C. The results of the inversion code are shown in Fig. 9. Clearly, the material is lossier and stiffer at lower temperature. This condition occurs when the material is above the glass transition temperature.[8] Therefore, this material is on the rubbery side of the glass transition. Results for the shear modulus show similar trends. Note the rather dramatic increase in the loss factor (by a factor of 3) when the temperature drops from 40 to 7 °C. As expected with the U shape, there is not significant ''dead zone'' and the code converges well at all frequencies. The small oscillations that can be observed for the frequency dependence of E and η are not physical and they are mostly due to a combination of numerical and experimental errors.

C. Static pressure dependence

Due to the large amount of voided space inside the material, the dimensions of the samples change considerably when the static pressure is applied. The sample volume decreases more than 35% at 34 atm [500 psi (gauge)]. The volumetric data are obtained by measuring the change in thickness of the sample by placing a laser spot on one edge of the top face and moving it across the thickness with a micrometer lead screw to the other edge. This measurement is accurate to about 125 μm. The sample volume is found by assuming that the dimensions of the macroscopically homo-

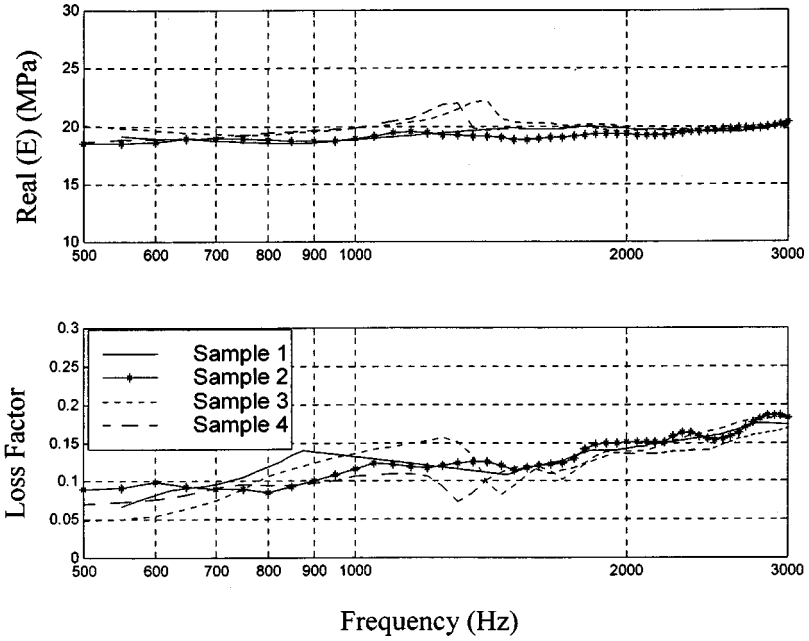


FIG. 8. Measurement repeatability and sample to sample variablity. Young's modulus and loss factors of several voided samples of different shape and aspect ratios.

geneous and isotropic sample change proportionately. These changes in volume and density are quite significant and must be included in the finite element model used in the inversion process. In light of the large change in the dimensions of the sample under pressure, it is clear that a rigid bond between the sample and baseplate would result in large lateral residual stresses in the bottom section of the sample. It is therefore necessary to replace the cyanoacrylate bond with one that will allow some degree of creep thereby relieving the residual stresses over time. The best results were obtained with a thin layer (0.66 mm) of uncured butyl rubber inserted between the sample and the aluminum base. The experiments were performed at room temperature by increasing the pressure in 100 psi increments (about 6.8 atm or 0.68 MPa increments) starting at 0 psi (gauge) and ending at 500 psi (gauge) (34 atm). The sample was allowed to stabilize over 8 h or more at each new pressure before taking measurements. A pressure cycle is defined as the increase from the lower to the higher pressure, maintaining the higher pressure for several hours, and the decompression to ambient pressure. No measurements were taken during the decompression phase of the cycle because air bubbles tended to form within the butyl rubber layer creating the possibility of introducing some unpredictable asymmetry in the motion of the sample.

The results obtained with the U-shape sample (sample 2) showed very poor convergence of the moduli. Under static pressure, the sample was seen to deform nonuniformly, especially at the top, thus indicating that the finite element model was inadequate for this experiment. Therefore, the rectangular block (sample 1) was used instead and the inversion did not converge well at all frequencies (''dead zone''). Data points are not plotted in the region of poor convergence. The inversion results obtained with data taken during

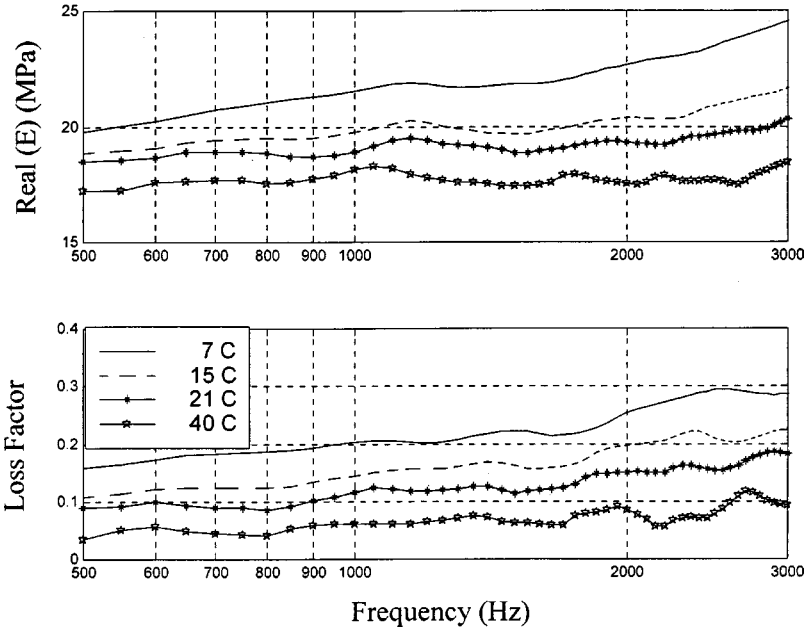


FIG. 9. Temperature dependence of the voided material. Sample 2 (U shape). Young's modulus and loss factor at 7, 15, 21, and 40 °C.

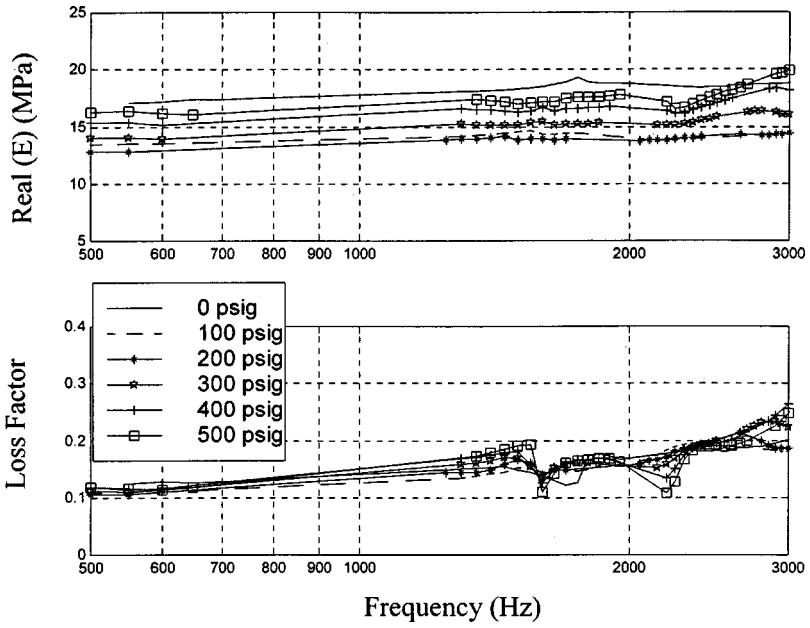


FIG. 10. Pressure dependence of the voided material. Sample 1 (rectangular shape). Young's modulus and loss factor at several static pressures: $P=0$, 100, 200, 300, 400, 500 psi (gauge).

the second pressure cycle are shown in Fig. 10. The frequency dependence of the moduli is not smooth indicating a limitation of the proposed technique. Nevertheless, the results show two interesting trends. First, within experimental errors, the loss factor does not change significantly with static pressure. Second, the Young's modulus initially decreases as the pressure is increased. After reaching a minimum, the Young modulus increases again as the static pressure increases. Figure 11 displays the Young's modulus at three frequencies, 550, 1700, and 2350 Hz, as a function of pressure to show this trend more clearly. The stiffness first decreases, reaching a minimum somewhere between 100 and 200 psi (gauge), and then starts to increase, approaching its original value at 500 psi (gauge). The most probable explanation is the collapse of the glass shells encasing the voids. Using a simple model of the material ignoring the glass shells, such as in Gaunaurd,[33] one would expect the Young's modulus to increase gradually with pressure as the voids are compressed and the void content decreases. However, the glass shells appear to introduce another mechanism into the behavior of the material as it is compressed. At low pressures the shells prevent the voids from collapsing until a critical pressure is reached. Here the void content may be increasing due to compression of the rubber matrix. It is not until after this critical pressure is exceeded that the void content of the material starts to decrease and the properties of the matrix material begin to take on more dominance.

The frequency dependence of the Young and shear moduli as a function of pressure is not smooth. Several factors may contribute to this limitation of the technique. A probable explanation is that the samples are not as isotropic and homogeneous as desired. Indeed, at high static pressures, the sample was seen to deform nonuniformly, especially at the top, clearly revealing the presence of anisotropy. These

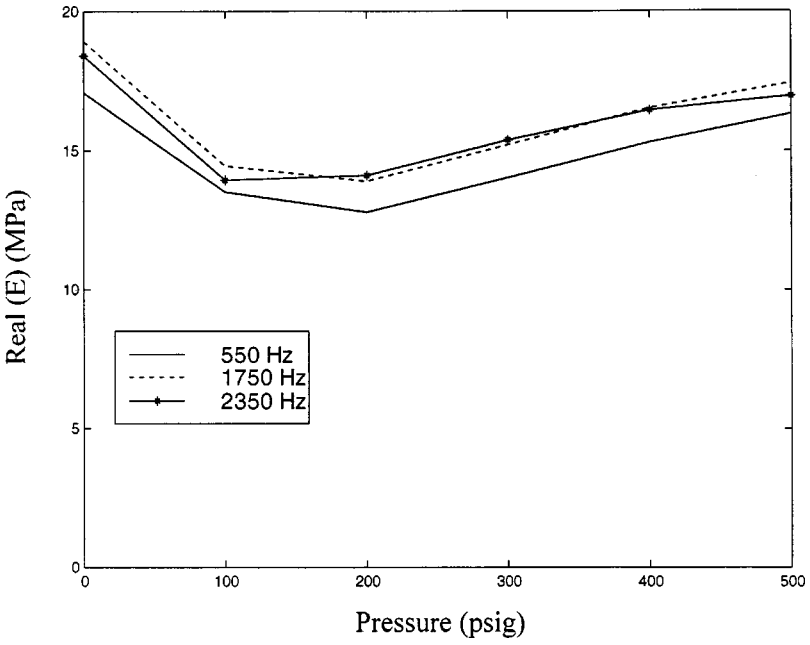


FIG. 11. Young's modulus as a function of pressure for three frequencies: 550, 1700, and 2350 Hz.

deformations could cause the baseplate and the sample to wobble. In one case, the top corner of the sample actually split open almost as if delaminating, indicating that the sample may have been cured in stages. Another source of error is the poorly defined boundary condition between the sample and the baseplate through the thin layer of butyl rubber. In addition, the baseplate itself may cause some shielding against the static pressure and cause some deformation localized at the base of the sample. Whether it is the sample or the measurement technique or a combination of the two, it is clear that with soft materials at high static pressure, the proposed method has its own limitations.

V. CONCLUSIONS

The Young and shear dynamic moduli of viscoelastic materials have been determined from laser vibrometric measurements of the surface motion of a three-dimensional sample excited by a piezoelectric actuator inside a chamber with controllable temperature and static pressure. The moduli were estimated from an inversion code that minimizes the difference between the data and the predictions from a finite element model in which the elastic moduli are the adjustable parameters. The results were compared with those obtained from the resonant rod technique using well-characterized homogeneous samples. The Young's modulus of Lucite was also determined using the laser-based/FEM technique and, within experimental errors, the results compared well with published values. Results were then obtained with microvoided samples in the 0.5–3 kHz frequency range, at temperatures ranging from 7 to 40 °C, and static pressures ranging from ambient to 34 atm (3.45 MPa or 500 psi). It was found that, with rectangular samples, the inversion code does not converge well in a frequency range between the first and second longitudinal modes of the sample. This effect is considerably reduced by using a sample with a U shape, or with lossy samples. The temperature dependence of the moduli is quite strong. In particular, the loss factor was found to increase by a factor of 3 if the temperature decreased from 40 to 7 °C. The dependence of the loss factor on static pressure, however, was small and below the measurement accuracy of our system in the range 0–34 atm [0–500 psi (gauge)]. The real parts of the Young and shear moduli decreased by about 30% when the static pressure was increased from 0 to about 6.8 to 13.6 atm [100–200 psi (gauge)], and then increased again so that, at 34 atm [500 psi (gauge)], they were slightly below their values at ambient conditions. When pressurized, the samples not only shrunk considerably but also deformed nonuniformly revealing the presence of anisotropy in the samples. Under those circumstances, better results were obtained with the rectangular block than with the U-shape sample. The measured frequency dependence of the moduli at high static pressure is not smooth, indicating that the inversion code did not converge in a robust manner throughout the frequency range of interest. Whether it is the sample itself or the measurement technique or a combination of the two, it is clear that the technique does not work very well with soft materials at high static pressure. Despite its limitations, the laser-based technique enables in many cases of practical interest the direct, simultaneous determination of the Young and shear moduli as a function of frequency, temperature, and, to some extent, pressure, without the need for time-temperature shifts, on three-dimensional samples.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the Office of Naval Research, Code 334. Homogeneous samples were provided by Bruce Hartmann of Naval Surface Weapon Center—Carderock Division. The resonant rod experiment was performed by Chad Rasmussen.

APPENDIX: NUMERICAL ESTIMATE OF THE VARIANCE

Let V_k $(k=1,N)$ be complex surface velocities ($N=9$ surface points in our case), θ_j $(j=1,...,M)$ be the elastic moduli of material to be determined ($M=3$ in our case: Young's modulus, loss factor, shear modulus). Define $V_{\text{data},k}$ as the measured surface velocity and $V_{\text{FEM},k}=f_k(\hat{\theta}_j)$ as the surface velocity computed with the FEM code where $\hat{\theta}_j$ are the values of the moduli estimated by the inversion procedure. Also define θ_j^* as the exact values of the moduli (unknowns). Let

$$V_{\text{data},k}-V_{\text{FEM},k}=\varepsilon_k, \quad \hat{\theta}_j-\theta_j^*=\varepsilon_j', \tag{A1}$$

where ε_k is the error which combines measurement error, theoretical model error, and error due to the difference ε_j' between the exact and estimated values of the moduli. It follows that, to first order in ε_j',

$$\varepsilon_k=\frac{\partial f_k(\hat{\theta}_j)}{\partial \theta_j}\varepsilon_j'. \tag{A2}$$

The sensitivity matrix, X, is therefore defined as the N by M matrix

$$X=\begin{bmatrix} \dfrac{\partial f_1(\hat{\theta})}{\partial \theta_1} & \cdots & \dfrac{\partial f_1(\hat{\theta})}{\partial \theta_M} \\ \cdots & \cdots & \cdots \\ \dfrac{\partial f_N(\hat{\theta})}{\partial \theta_1} & \cdots & \dfrac{\partial f_N(\hat{\theta})}{\partial \theta_M} \end{bmatrix}. \tag{A3}$$

It can be shown that, in general, the variance–covariance matrix of the estimates is[32]

$$\text{Var}=[X'X]^{-1}\sigma^2, \tag{A4}$$

where, if the model is appropriate, we can express σ^2 in terms of the sum of the residuals ε_k from

$$\sigma^2=\frac{1}{N-M}\sum_{k=1}^{N}\varepsilon_k^2 \tag{A5}$$

provided than $N \gg M$. The estimates can be regarded as a normal distribution (N_0) with variance Var, centered around θ^*,

$$\hat{\theta}\cong N_0[\theta^*,\sigma^2(X'X)^{-1}]. \tag{A6}$$

[1] R. N. Capps, ''Elastomeric materials for acoustical applications,'' Naval Research Laboratory—Underwater Sound Reference Detachment, 15 September, 1989.
[2] W. M. Madigosky and G. F. Lee, ''Automated dynamic Young's modulus and loss factor measurements,'' J. Acoust. Soc. Am. **66**, 345–349 (1979).
[3] W. M. Madigosky and Gilbert F. Lee, ''Improved resonance technique for materials characterization,'' J. Acoust. Soc. Am. **73**, 1374–1377 (1983).
[4] ANSI S2.22-1998, ''Resonance method for measuring the dynamic mechanical properties of viscoelastic materials,'' American National Standards Institute, published through the Acoustical Society of America, New York, NY, 1998.
[5] L. F. Nielsen, N. J. Wismer, and S. Gade, ''Improved method for complex modulus estimation,'' Sound Vib. **34**, 20–24 (2000).
[6] ANSI S2.23-1998, ''Single cantilever beam method for measuring the dynamic mechanical properties of viscoelastic materials,'' American National Standards Institute, published through the Acoustical Society of America, New York, NY, 1998.
[7] Rheometric Scientific, Inc., Dynamic Mechanical Thermal Analyzer IV, 1996.
[8] M. L. Williams, R. F. Landel, and J. D. Ferry, ''The temperature dependence of relaxation mechanisms in amorphous polymers and other glass-forming liquids,'' J. Am. Chem. Soc. **77**, 3701–3706 (1955).
[9] J. Piquette, ''Determination of the complex dynamic bulk modulus of elastomers by inverse scattering,'' J. Acoust. Soc. Am. **77**, 1665–1673 (1985).
[10] R. L. Willis, T. S. Stone, and Y. H. Berthelot, ''A laser-based experimental-numerical technique for evaluating the bulk and shear dynamic moduli of viscoelastic materials,'' J. Acoust. Soc. Am. **102**, 3549–3555 (1997).
[11] J. Jarzynski, ''Mechanisms of sound attenuation in materials,'' in *Sound and Vibration Damping with Polymers*, edited by R. D. Corsaro and L. H. Sperling (The American Chemical Society, Washington, DC, 1990), pp. 167–207.
[12] E. H. Kerner, ''The elastic and thermo-elastic properties of composite media,'' Proc. Phys. Soc. London **69**, 808–813 (1956).
[13] I. A. Chaban, ''Self-consistent field approach to the calculation of the effective parameters of microinhomogeneous media,'' Sov. Phys. Acoust. **10**, 298–304 (1965).
[14] G. T. Kuster and M. N. Toksöz, ''Velocity and attenuation of seismic waves in two-phase Media. I. Theoretical formulations,'' Geophysics **39**, 587–606 (1974).
[15] G. C. Gaunaurd and H. Überall, ''Resonance theory of the effective properties of perforated solids,'' J. Acoust. Soc. Am. **71**, 282–295 (1982).
[16] G. C. Gaunaurd and H. Überall, ''Erratum: 'Resonance theory of the effective properties of perforated solids,''' J. Acoust. Soc. Am. **73**, 372 (1983).
[17] G. C. Gaunaurd and H. Überall, ''Author's reply to recent comments on our effective medium theory,'' J. Acoust. Soc. Am. **82**, 696–699 (1983).
[18] F. H. Kerr, ''The scattering of a plane elastic wave by spherical elastic inclusions,'' Int. J. Eng. Sci. **30**, 169–186 (1992).
[19] J. D. Eshelby, ''Elastic inclusions and inhomogeneities,'' Prog. Solid Mech. **2**, 89–98 (1961).
[20] M. Berveiller, F. O. Fassi, and A. Hihi, ''The problem of two plastic and heterogeneous inclusions,'' Int. J. Eng. Sci. **25**, 691–709 (1987).
[21] M. Cherkaoui, H. Sabar, and M. Berveiller, ''Micromechanical approach of the coated inclusion problem and applications to composite materials,'' J. Eng. Mater. Technol. **116**, 274–278 (1994).
[22] P. C. Waterman and R. Truell, ''Multiple scatter inclusions,'' J. Appl. Phys. **62**, 1613–1615 (1987).
[23] V. K. Varadan, Y. Ma, and V. V. Varadan, ''A multiple scattering theory for elastic wave propagation in discrete random media,'' J. Acoust. Soc. Am. **77**, 375–385 (1985).
[24] W. H. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, 1986), pp. 289–301.
[25] J. R. Hutchinson and S. D. Zillmer, ''Vibration of a free rectangular parallelepiped,'' J. Appl. Mech. **50**, 123–130 (1983).
[26] R. H. Myers and D. C. Montgomery, *Response Surface Methodology* (Wiley, New York, 1995).
[27] M. Giovagnoni, ''On the direct measurement of the dynamic Poisson's ratio,'' Mech. Mater. **17**, 33–46 (1994).
[28] In this particular experiment, the sample was actually glued to the aluminum base with the cyanoacrylate bond because these calibration samples do not substantially shrink under static pressure thus alleviating the need for the butyl rubber compliant bond.
[29] N. Lagakos, J. Jarzynski, J. H. Cole, and J. A. Bucaro, ''Frequency and temperature dependence of elastic moduli of polymers,'' J. Appl. Phys. **59**, 4017–4031 (1986).
[30] For this particular experiment, the surface velocities were measured at nine points, since the estimate of the variance requires that the number of surface points be much greater than the number of unknown parameters (three in our case) being evaluated.
[31] G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experiments* (Wiley, New York, 1978), Chap. 14.
[32] J. L. Buchanan, ''Numerical solutions for the dynamic moduli of a viscoelastic bar,'' J. Acoust. Soc. Am. **81**, 1775–1786 (1987).
[33] G. Gaunaurd, E. Callen, and J. Barlow, ''Pressure effects on the dynamic effective properties of resonating perforated elastomers,'' J. Acoust. Soc. Am. **76**, 173–177 (1984).

Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation

Radha Kalluri
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Speech and Hearing Sciences Program, Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139

Christopher A. Shera
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114, Speech and Hearing Sciences Program, Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139, and Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02115

(Received 5 July 2000; revised 6 October 2000; accepted 27 October 2000)

This paper tests key predictions of the ''two-mechanism model'' for the generation of distortion-product otoacoustic emissions (DPOAEs). The two-mechanism model asserts that lower-sideband DPOAEs constitute a mixture of emissions arising not simply from two distinct cochlear locations (as is now well established) but, more importantly, by two fundamentally different mechanisms: nonlinear distortion induced by the traveling wave and linear coherent reflection off pre-existing micromechanical impedance perturbations. The model predicts that (1) DPOAEs evoked by frequency-scaled stimuli (e.g., at fixed f_2/f_1) can be unmixed into putative distortion- and reflection-source components with the frequency dependence of their phases consistent with the presumed mechanisms of generation; (2) The putative reflection-source component of the total DPOAE closely matches the reflection-source emission (e.g., low level stimulus-frequency emission) measured at the same frequency under similar conditions. These predictions were tested by unmixing DPOAEs into components using two completely different methods: (a) selective suppression of the putative reflection source using a third tone near the distortion-product frequency and (b) spectral smoothing (or, equivalently, time-domain windowing). Although the two methods unmix in very different ways, they yield similar DPOAE components. The properties of the two DPOAE components are consistent with the predictions of the two-mechanism model. [DOI: 10.1121/1.1334597]

PACS numbers: 43.64.Bt, 43.64.Jb, 43.64.Yp [BLM]

I. INTRODUCTION

Mammalian otoacoustic emissions (OAEs) have generally been regarded as originating through nonlinear electromechanical distortion (e.g., Kemp, 1978, 1997, 1998; Probst *et al.,* 1991; Allen and Neely, 1992; Allen and Lonsbury-Martin, 1993; Patuzzi, 1996). Shera and Guinan (1999), however, argue that OAEs arise by at least two fundamentally different mechanisms within the cochlea. These differences in mechanism, they suggest, can profitably be used to define an ''OAE family tree.'' The mechanism-based taxonomy groups emissions into two basic types: distortion-source emissions, which arise by nonlinear distortion induced by the traveling wave, and reflection-source emissions, which arise via linear reflection off pre-existing micromechanical impedance perturbations (Shera and Zweig, 1993b; Zweig and Shera, 1995). This distinction between distortion- and reflection-source emissions differs from the ''wave-'' and ''place-fixed'' dichotomy maintained by Kemp and Brown (1983) in that the latter was introduced and developed within an integrated framework that views all OAEs as manifestations of cochlear mechanical nonlinearity. The mechanism-based taxonomy, by contrast, emphasizes the fundamental differences between linear (reflection-source) and nonlinear (distortion-source) emission mechanisms.

The analysis underlying the taxonomy predicts that the two types of OAEs mix to form the evoked emissions measured in the ear canal. In any given measurement, the different emission types contribute in degrees dependent on species, stimulus parameters, and cochlear state. As an example of the process, Shera and Guinan suggest that the generation of lower-sideband distortion-product otoacoustic emissions (DPOAEs) can be understood in terms of the mixing of the two OAE types. Much of DPOAE fine structure apparently arises through the interference of emissions originating from two distinct cochlear locations (e.g., Kim, 1980; Kemp and Brown, 1983). Although the ''two-place model'' for DPOAEs now appears well established (e.g., Gaskill and Brown, 1990; Brown *et al.*, 1996; Engdahl and Kemp, 1996; Brown and Beveridge, 1997; Heitmann *et al.*, 1998; Fahey and Allen, 1997; Siegel *et al.*, 1998), the taxonomy identifies the two interfering emission components as arising not simply from two distinct locations, but, more importantly, via two different mechanisms.

The proposed generation process is illustrated in Fig. 1.

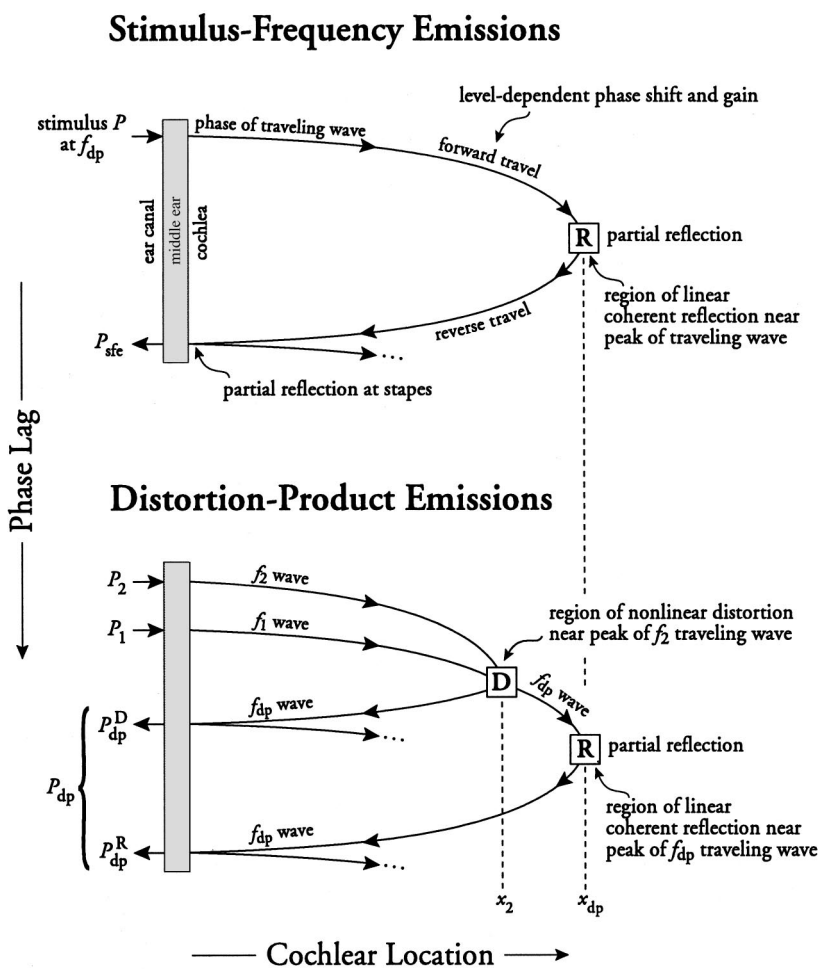


FIG. 1. Schematic diagram of the two-mechanism model. The figure illustrates the generation of SFOAEs (top) and DPOAEs (bottom) at low sound levels. Each panel shows phase lag relative to stimulus phase (lag increasing downward) of forward- and backward-traveling waves vs cochlear location. At low stimulus levels, SFOAEs (P_{sfe}) result from coherent reflection (**R**) in the region near the peak of the traveling-wave envelope. For DPOAEs, the primary traveling waves produce a region of nonlinear distortion (**D**), located near the peak of the f_2 wave (at x_2), where nonlinear distortion generates traveling waves at the frequency f_{dp} that travel in both directions (shown here for the case $f_{dp}=2f_1-f_2$, where f_{dp} equals the SFOAE frequency shown in the top panel). The backward-traveling wave propagates to the ear canal (where it appears as the distortion-source emission, P_{dp}^D). The forward-traveling wave propagates to its characteristic place (at x_{dp}), where it undergoes partial reflection (**R**) near the peak of its wave envelope, generating a second backward-traveling wave that propagates to the ear canal (the reflection-source emission, P_{dp}^R). The two types of emission combine to produce the DPOAE measured in the ear canal ($P_{dp}=P_{dp}^D+P_{dp}^R$). For simplicity, phase shifts due to propagation through the middle ear and reflection by the stapes are not shown. Adapted, with permission, from Shera and Guinan (1999).

The primary traveling waves, at frequencies f_1 and f_2 (with $f_2>f_1$), interact to produce a region of nonlinear distortion (**D**), located near the peak of the f_2 wave, which creates energy at distortion-product frequencies. In particular, traveling waves at the frequency $f_{dp}=2f_1-f_2$ are generated that travel in both directions. The backward-traveling wave propagates to the ear canal, where it appears as a distortion source emission. The forward-traveling wave propagates to its characteristic place, where it undergoes partial reflection (**R**) near the peak of its wave envelope, generating a second backward-traveling wave that propagates to the ear canal (a reflection-source emission). The two types of emission mix in the ear canal.[1]

The proposed model thus predicts that the two components originate not simply from two different regions of the cochlea but—more significantly—by two fundamentally different mechanisms. Similar predictions emerge from recent modeling studies (e.g., Talmadge *et al.*, 1998, 1999; Mauermann *et al.*, 1999a, b). Based on nonlinear cochlear models that meet the requirements detailed by the theory of coherent reflection filtering for the generation of realistic reflection emissions (Shera and Zweig, 1993b; Zweig and Shera, 1995),[2] these studies incorporate both classes of emission-generating mechanisms (i.e., nonlinear distortion and linear coherent reflection). The primary goal of the experiments reported here was to test this *two-mechanism model* for DPOAE generation.

According to the analysis underlying the taxonomy, distortion- and reflection-source emissions manifest very different frequency dependencies in their phase. In a nutshell, the argument runs roughly as follows:

Distortion-source OAEs: Nonlinear distortion depends upon the interaction between the two primary traveling waves. When produced using frequency-scaled stimuli, the combined excitation pattern of the primary traveling waves simply translates along the cochlear partition as the stimulus frequencies are varied. This approximate translation invariance (or ''shift similarity'') follows from local scaling symmetry and the logarithmic form of the cochlear frequency-position map. Approximate shift similarity ensures that the amplitudes and phases of the primary waves—and hence any nonlinear interactions between them—remain nearly invariant in a coordinate system that moves with the spatial envelope of the f_2 traveling wave as the frequencies are swept (Fig. 2, left). OAEs generated by frequency-scaled nonlinear distortion therefore manifest a nearly constant phase.

Reflection-source OAEs: According to the theory of coherent reflection filtering (Zweig and Shera, 1995), reflection-source emissions are generated when a forward-traveling wave reflects off ''random'' perturbations in the mechanics of the cochlea. The phase of each scattered wavelet depends on the phase of the forward-traveling wave at the location of scattering. Since the micromechanical impedance perturbations are fixed in space (unlike sources of nonlinear

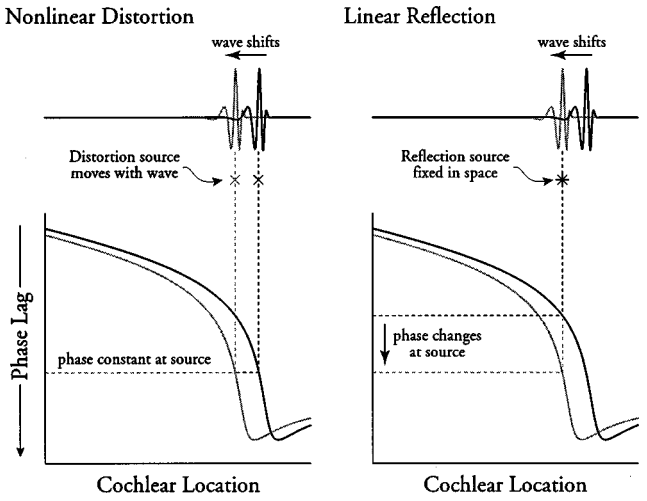


FIG. 2. Schematic diagram illustrating the consequences of scaling for the phase of distortion- and reflection-source emissions. The left-hand panel shows a snapshot of the f_2 traveling wave at two different frequencies (top) and the corresponding phase lags (bottom) vs cochlear location. The two frequencies are denoted by black and gray lines, respectively. For simplicity, the f_1 traveling waves are not shown. Distortion sources result from nonlinear interaction between the primary traveling waves. The sources illustrated here ($\times$) are idealized as points at the peak of the f_2 traveling wave. When the frequency ratio f_2/f_1 is held fixed during the measurement sweep, the primary traveling waves (and thus the resulting distortion sources) simply shift along the cochlear partition, maintaining a nearly constant relative phase relationship as the stimulus frequencies are varied. Note, for example, that the phases of the primary traveling waves at the distortion source remain constant as the frequency is increased and the wave pattern shifts ($\leftarrow$) along the partition. As a result, the phases of all resulting distortion products are essentially independent of frequency. The right-hand panel shows a similar diagram for a reflection source (e.g., a perturbation in the mechanics of the cochlea). Since the perturbation (*) is fixed in space the phase of the wave scattered by the perturbation changes considerably ($\downarrow$) as the stimulus frequency is varied. Consequently, the phases of OAEs generated by linear reflection vary rapidly with frequency.

distortion, which move with the excitation pattern as the frequency changes), the phase of the incident wave at each perturbation changes as the frequency of the forward-traveling wave is varied (Fig. 2, right). Consequently, OAEs generated by linear reflection manifest a phase that rotates rapidly with frequency.

In this paper we apply this reasoning to test the two principal predictions of the two-mechanism model, as suggested by the taxonomy and framework presented in Fig. 1. Specifically, we test the predictions that

(1) The total distortion-product emission, P_{dp}, represents the sum of distortion- and reflection-source components, P_{dp}^{D} and P_{dp}^{R}

$$P_{dp}=P_{dp}^{D}+P_{dp}^{R}, \tag{1}$$

where the components P_{dp}^{D} and P_{dp}^{R} manifest frequency dependencies in their phase consistent with their presumed mechanisms of generation. Specifically, the model predicts that when P_{dp} is evoked using frequency-scaled stimuli (e.g., with the ratio f_2/f_1 fixed), the phase of P_{dp}^{D} should be essentially independent of frequency, whereas the phase of P_{dp}^{R} should rotate rapidly.

(2) The putative reflection-source component, P_{dp}^{R}, of the total DPOAE closely matches the reflection emission measured at the same frequency under similar conditions. According to the taxonomy, stimulus-frequency emissions (SFOAEs) evoked at low stimulus levels are nearly pure reflection emissions (see Fig. 1). We thus test the prediction that

$$P_{dp}^{R}\approx P_{sfe}, \tag{2}$$

where P_{sfe} is the SFOAE at the same frequency. Once stimulus parameters have been adjusted to yield comparable overall emission levels, the predicted match includes the frequency dependence of both the amplitude (or spectral shape) and the phase.[3]

Testing these predictions requires a technique for unmixing the total DPOAE into putative distortion- and reflection-source components. Initially, we adopt an experimental approach based on selective suppression that exploits the spatial separation of the presumed distortion- and reflection-source regions within the cochlea. To explore the sensitivity of our results to variations in the methodology of unmixing, we compare our results obtained using suppression to an alternative unmixing procedure based on spectral smoothing or time-domain windowing. A preliminary account of this work has been presented elsewhere (Kalluri and Shera, 2000).

II. UNMIXING VIA SELECTIVE SUPPRESSION

Reference to Fig. 1 suggests that one can separate the two components, P_{dp}^{D} and P_{dp}^{R}, of the total DPOAE pressure if the reflection-source emission originating from the **R** region can be eliminated. The unmixing procedure would then be to (1) measure the total emission, P_{dp}, using frequency-scaled stimuli; (2) eliminate the **R** component and remeasure the DPOAE to obtain the pure distortion-source component, P_{dp}^{D}; and (3) compute the reflection-source component, P_{dp}^{R}, by subtraction, $P_{dp}^{R}=P_{dp}-P_{dp}^{D}$.

The spatial separation of the two source regions within the cochlea suggests trying to eliminate the **R** component by introducing a third, suppressor tone at a nearby frequency. The suppressor would act by reducing the amplitude of the wavelets incident upon and/or scattered back from the **R** region. Suppression techniques for separating OAE sources originating at different spatial locations in the cochlea were pioneered by Kemp and Brown (1983) and later refined by others (e.g., Heitmann *et al.*, 1998; Siegel *et al.*, 1998). The selective suppression strategy for unmixing yields the following estimates of P_{dp} and its components:

$$P_{dp}=P_{ec}(f_{dp}) \quad \text{(measured at fixed } f_2/f_1\text{)}; \tag{3}$$

$$P_{dp}^{D}\approx P_{ec}(f_{dp})\big|_{\text{with suppressor}} \quad \text{(with suppressor at } f_s\approx f_{dp}\text{)}; \tag{4}$$

$$P_{dp}^{R}\approx P_{dp}-P_{dp}^{D}. \tag{5}$$

In these expressions, $P_{ec}(f)$ denotes the complex ear-canal pressure at frequency f resulting from stimulation at primary frequencies f_1 and f_2.

A. Measurement methods

We measured emissions from one ear of each of four normal hearing humans. Treatment of human subjects was in accordance with protocols approved by the Human-Studies Committee at the Massachusetts Eye and Ear Infirmary. All measurements were performed with subjects reclining comfortably in a sound-proofed, vibration-isolated chamber (Ver *et al.*, 1975). Stimulus waveforms were generated and responses acquired and averaged digitally using a custom-built data-acquisition system. Acoustic signals were transduced using a Etymotic Research ER-10c DPOAE probe system supplemented with an ER-3A earphone whose sound-delivery tube was threaded through the ER-10c foam eartip. *In situ* earphone calibrations were performed at regular intervals throughout all measurement sessions. The calibrations were used to guarantee that the stimulus tones had constant level and zero starting phase in the ear canal at all frequencies. Real-time artifact rejection was implemented by comparing the time waveforms in successive data buffers before adding them to the averaged responses. In these and other respects, the methods and equipment used to obtain both SFOAEs and DPOAEs are generally similar to those described elsewhere (Shera and Guinan, 1999). We briefly summarize relevant differences here and provide detailed descriptions in the Appendix.

1. Measurement of DPOAEs

We measured distortion-product emissions at the frequency $2f_1-f_2$ using frequency-scaled stimuli (i.e., using frequency sweeps performed with the primary-frequency ratio, f_2/f_1, held constant). The measurements reported here were obtained using primary levels of $\{L_1,L_2\}=\{60,45\}$ dB SPL at the frequency ratio $f_2/f_1=1.2$. To ensure that our ability to maintain a constant f_2/f_1 ratio during the sweep was not systematically compromised by the frequency quantization imposed by digital stimulus generation, we modified our data-acquisition system to allow the sampling frequency to vary between measurement points. This flexibility enabled us to choose f_1 and f_2 so that the ratio f_2/f_1 varied by less than a thousandth of a percent between measurements (at our typical frequency spacing of about 15 Hz). The resulting sampling frequencies varied by less than 3% about the nominal value (59.94 kHz).

To allow any multiple reflections that might be present within the cochlea to settle into an approximately steady-state response, we measured $P_{\rm dp}$ and $P_{\rm dp}|_{\rm suppressed}$ over time intervals ($\approx$136 ms) much longer than the estimated round-trip travel time for cochlear waves ($\approx$10–15 ms). To guard against possible systematic variations in emission amplitude over time that might invalidate the unmixing procedure (e.g., due to efferent feedback or to changes in earphone calibration caused by subject movement or by temperature variations), we interleaved measurements of $P_{\rm dp}$ and $P_{\rm dp}|_{\rm suppressed}$ in time and averaged multiple repetitions (typically $n=64$). We set the suppressor frequency approximately 44 Hz below the distortion-product frequency (e.g., Siegel *et al.*, 1998; Dreisbach and Siegel, 1999). The suppressor level was adjusted (separately for each subject) to minimize the DPOAE fine structure while leaving the mean DPOAE amplitude (as averaged over several fine-structure periods) largely unchanged (cf. Heitmann *et al.*, 1998). The suppressor level chosen in this way was generally in the range 50–55 dB SPL.

2. Measurement of SFOAEs

We measured stimulus-frequency emissions using the suppression method (e.g., Guinan, 1990; Kemp *et al.*, 1990; Brass and Kemp, 1991, 1993; Souter, 1995; Shera and Guinan, 1999). In this method, the emission is obtained as the complex (or vector) difference between the ear-canal pressure at the probe frequency measured first with the probe tone alone and then in the presence of a stronger suppressor tone at a nearby frequency. Thus, the SFOAE pressure at the probe frequency, $P_{\rm sfe}(f_{\rm p})$, is defined as

$$P_{\rm sfe}(f_{\rm p}) \equiv P_{\rm ec}(f_{\rm p}) - P_{\rm ec}(f_{\rm p})|_{\rm with\ suppressor\ at\ } f_{\rm s}\approx f_{\rm p}. \tag{6}$$

In the measurements reported here, the suppressor frequency was approximately 44 Hz below the probe ($f_{\rm s}\approx f_{\rm p}-44\,{\rm Hz}$). To prevent contamination from the considerable cross talk between output channels of the ER-10c, we generated the suppressor tone using a separate ER-3A earphone whose sound-delivery tube was threaded through the foam eartip. Unless otherwise noted, the probe and suppressor levels, $L_{\rm p}$ and $L_{\rm s}$, were 40 and 55 dB SPL, respectively (Shera and Guinan, 1999). Exploratory measurements at other nearby probe levels indicate that the spectral shape and phase of $P_{\rm sfe}$ are not strong functions of intensity at these levels. We found that probe levels of 40 dB SPL gave emission levels generally comparable to those of $P_{\rm dp}^{\rm R}$, especially after introduction of the primary-mimicking tone described below. As with the DPOAE measurements, we interleaved measurements of $P_{\rm ec}(f_{\rm p})$ and $P_{\rm ec}(f_{\rm p})|_{\rm suppressed}$ in time to minimize artifacts that might contaminate the difference. The measurement frequency resolution, approximately 15 Hz between points, was always sufficient to prevent ambiguities in phase unwrapping.

In some experiments, we measured SFOAEs in the presence of an additional continuous tone. The idea was to measure $P_{\rm sfe}$ under conditions matching as closely as possible those present during the measurement of $P_{\rm dp}^{\rm R}$. Thus, we introduced the additional tone at a frequency and level corresponding to the f_1 primary used in the measurement of DPOAEs. In terms of the probe frequency, the frequency of this additional tone (the ''f_1-primary mimicker'') was therefore given by $f_1=f_{\rm p}/(2-r)$, where r denotes the f_2/f_1 ratio we wished to mimic. We denote the SFOAE measured in the presence of the f_1-primary mimicking tone by $P_{\rm sfe}^1$:

$$P_{\rm sfe}^1 \equiv P_{\rm sfe}|_{\rm with\ } f_1 \text{ mimicker}. \tag{7}$$

B. Results: Unmixing via selective suppression

1. P_{dp} and its components, P_{dp}^{D} and P_{dp}^{R}

Typical measurements of the total DPOAE and its components estimated using suppression are shown in Fig. 3. To illustrate the variation across subjects, we show results for three of our four subjects (those for whom the most data are available); similar results were obtained in the fourth subject.

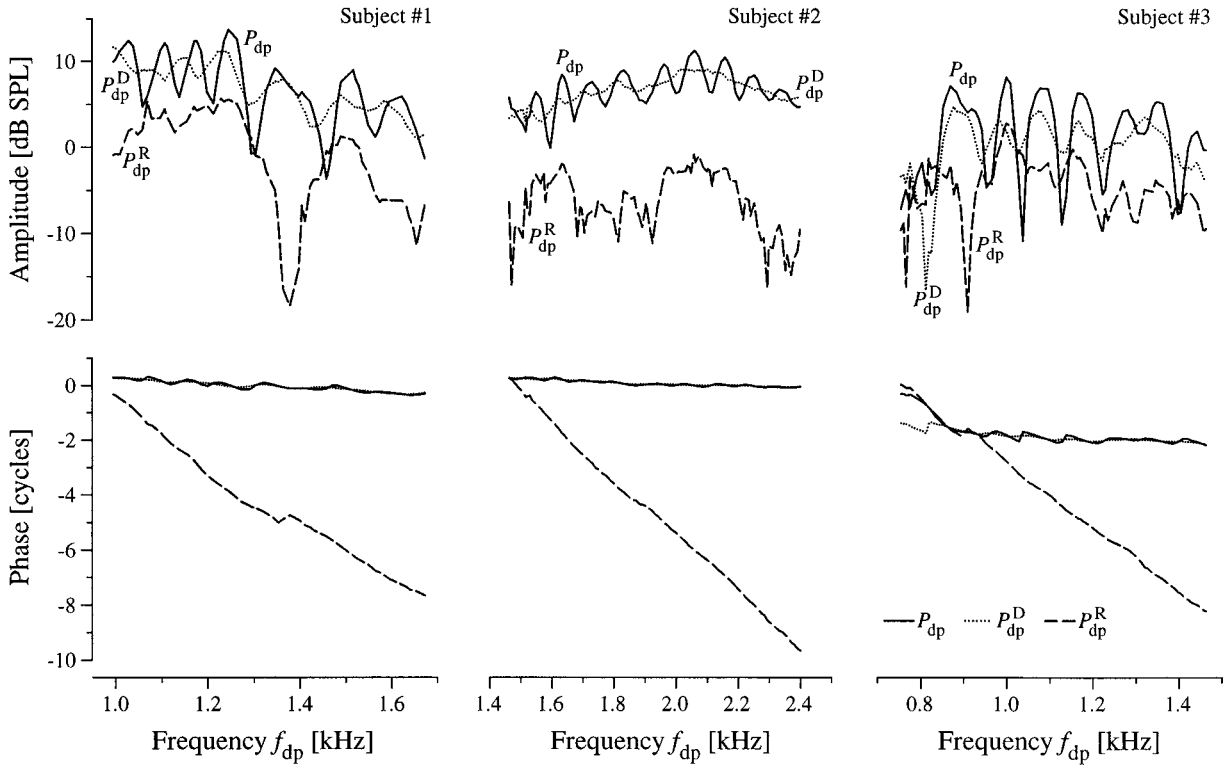


FIG. 3. The DPOAE P_{dp} and its estimated distortion- and reflection-source components, P_{dp}^{D} and P_{dp}^{R}, obtained using suppression. The figure shows typical measurements of the amplitude (top) and phase (bottom) of the $2f_1-f_2$ DPOAE and its components measured using a frequency-scaled stimulus (i.e., the primary frequencies f_1 and f_2 were swept with their ratio held constant at the value $f_2/f_1=1.2$). Left to right, the panels show data from subjects #1, #2, and #3, respectively; similar results were obtained in the fourth subject. In each case, the total DPOAE (solid line) was unmixed using a suppressor tone near the distortion-product frequency, f_{dp}. Although the phases of the estimated P_{dp}^{D} components (dotted lines) vary less than a period, the phases of the estimated P_{dp}^{R} components (dashed lines) fall through many cycles over the same frequency range, in agreement with Prediction #1. The measurement noise floor was approximately -25 dB SPL and the frequency resolution was always sufficient to prevent ambiguities in phase unwrapping. Stimulus levels for subject #1: $\{L_1,L_2,L_s\}=\{60,45,50\}$ dB SPL. Stimulus levels for subject #2: $\{L_1,L_2,L_s\}=\{60,45,50\}$ dB SPL. Stimulus levels for subject #3: $\{L_1,L_2,L_s\}=\{60,45,55\}$ dB SPL.

In each case, the putative distortion-source component, P_{dp}^{D}, is essentially a smoother version of the total DPOAE in which much of the quasiperiodic fine structure apparent in both the amplitude and phase of P_{dp} has been eliminated. In agreement with Prediction #1, the phase of P_{dp}^{D} is nearly constant, varying by less than a period. By contrast, the phase of the reflection-source component, P_{dp}^{R}, falls through many cycles (typically eight or more) over the same frequency range. These different frequency dependencies imply generation by fundamentally different mechanisms: The nearly constant phase of P_{dp}^{D} is consistent with generation by frequency-scaled nonlinear distortion and the rapidly rotating phase of P_{dp}^{R} with generation by linear coherent reflection (Shera and Guinan, 1999).

The fine-structure manifest in the total DPOAE arises because of alternating constructive and destructive interference between the two components, P_{dp}^{D} and P_{dp}^{R}, caused by the systematic rotation of their relative phase, a consequence of the very different slopes of their phase versus frequency functions. Thus, the components P_{dp}^{D} and P_{dp}^{R} ''beat'' against each other, producing an oscillatory interference pattern. In other words, DPOAE fine structure arises because DPOAEs are mixtures of emissions with distinctly different properties that reflect their different mechanisms of generation.

2. Comparison between P_{dp}^{R} and P_{sfe}

According to Prediction #2, the reflection-source component of the total DPOAE, P_{dp}^{R} (Fig. 1, lower panel), should closely match other reflection-source emissions measured under comparable conditions (e.g., SFOAEs at low stimulus levels as in Fig. 1, upper panel). Figure 4 tests this prediction by comparing P_{dp}^{R} and P_{sfe} measured in the same ear. In agreement with predictions, the phase slopes of P_{sfe} and P_{dp}^{R} are nearly identical. In addition, both P_{sfe} and P_{dp}^{R} have similar amplitude features (e.g., a deep notch near 1.4 kHz). These similarities support the idea that P_{sfe} and P_{dp}^{R} are generated by a similar mechanism. Note that deep spectral notches such as that apparent near 1.4 kHz are predicted by the theory of coherent reflection filtering (cf. Fig. 11 of Zweig and Shera, 1995). In the model, such notches arise from random spatial fluctuations in the irregularities that scatter the wave. At some frequencies, wavelets scattered from different locations within the scattering region combine nearly out of phase and cause near cancellation of the net reflected wave.

Although the overall match between P_{dp}^{R} and P_{sfe} is good—especially when one considers the substantial differences in the way that the two emissions are evoked and measured—details of the spectral shape (e.g., the precise lo-

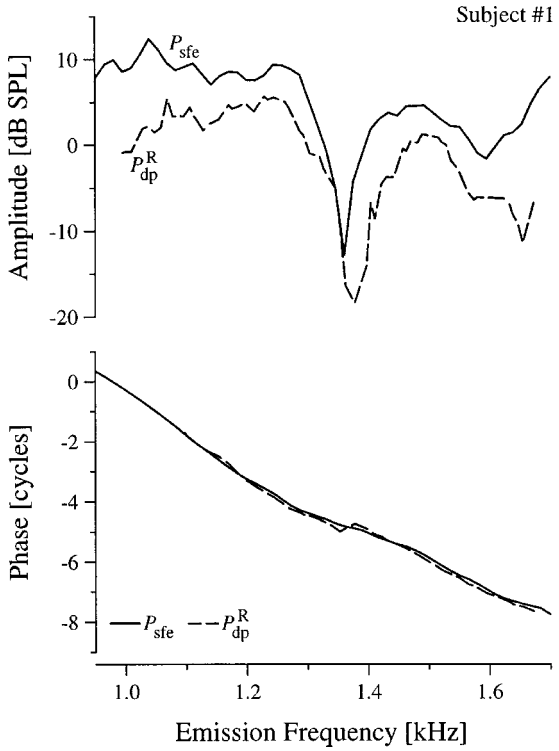


FIG. 4. Comparison between the estimated reflection-source component, P_{dp}^{R}, and the SFOAE, P_{sfe}. The figure shows the amplitude (top) and phase (bottom) of P_{dp}^{R} (dashed line), the reflection-source component of the total DPOAE obtained in Fig. 3 for subject #1. Shown for comparison is P_{sfe} (solid line), the SFOAE measured in the same subject at a probe level of 40 dB SPL. Note the considerable agreement in both amplitude and phase (e.g., P_{dp}^{R} and P_{sfe} have similar amplitude notches and phase slopes). SFOAE stimulus levels: $\{L_p, L_s\} = \{40,50\}$ dB SPL.

cation of the notch) do not match perfectly. Do these discrepancies suggest important differences between P_{dp}^{R} and P_{sfe} and their mechanisms of generation? Or do they reflect differences in measurement conditions that influence the magnitude and/or phase of the traveling-wave energy scattered back from **R**? For example, the primaries present during the DPOAE measurement may suppress the traveling wave near the f_{dp} place, thereby affecting the frequency dependence of P_{dp}^{R}.

3. Mimicking suppression by the primaries

To address these questions, we modified our P_{sfe}-measurement paradigm to better mimic the intracochlear conditions under which P_{dp}^{R} originated. Specifically, we measured P_{sfe} in the presence of an additional tone whose frequency and level were chosen to match those of the f_1 primary used during the measurement of P_{dp}^{R} (see Methods above). We mimic the f_1 primary because we expect it to have the greater effect; the f_1 primary is both closer in frequency to f_{dp} and higher in level than the f_2 primary. We define P_{sfe}^{1} as the value of P_{sfe} measured in the presence of the f_1-primary mimicker.

Measurements of P_{sfe}^{1} are shown and compared to those of P_{dp}^{R} in Fig. 5. The match between the two putative reflection-source emissions is now much closer. This result is consistent with the idea that the differences in Fig. 4 reflect differing intracochlear stimulus conditions; differences in the mechanisms of emission generation are not required. Thus, the similarity in both magnitude and phase between P_{dp}^{R} and P_{sfe}^{1} is in agreement with Prediction #2 and provides strong support for the model. Note that the changes in the overall amplitude and spectral shape of P_{sfe} caused by the f_1-primary mimicker suggest that the primaries have a significant effect on the reflection-source component of the DPOAE (presumably via suppression of the wave incident upon and/or reflected back from the **R** region).

III. UNMIXING VIA SPECTRAL SMOOTHING

A potential difficulty with suppression-based unmixing is that the suppressor tone, introduced with the intent of selectively suppressing the reflection-source component, may inadvertently modify the response in other ways. For example, the suppressor tone may also suppress the distortion-source component (either directly, or through its effects on the primaries) or ''catalyze'' the generation of additional distortion-sources at the frequency f_{dp} (Fahey *et al.*, 2000). As a test of these possibilities, and to investigate the sensitivity of our conclusions to the method of unmixing, we repeated our analysis using a completely different method. This method—spectral smoothing (or its equivalent, time windowing)—was suggested by the correspondence, in a linear system, between phase slope in the frequency domain and latency in the time domain (e.g., Papoulis, 1962). As unmixed by suppression, the two components P_{dp}^{D} and P_{dp}^{R} of P_{dp} have very different phase slopes, evidently reflecting fundamental differences in their mechanisms of generation. Consequently, if we apply Fourier analysis to our frequency-domain measurements of P_{dp}, we expect to see two components of very different latencies in the corresponding ''latency-domain response:''[4] namely, a short-latency component corresponding to P_{dp}^{D} and a long-latency component corresponding to P_{dp}^{R}. Thus, our suppression results suggest that the two components of P_{dp} should be separable using signal-processing strategies based on appropriate windowing in the latency domain. Techniques for analyzing OAEs in this way were introduced by Shera and Zweig (1993a; Zweig and Shera, 1995), who applied them to the study of SFOAEs; similar methods have since been applied to other emissions (e.g., Stover *et al.*, 1996; Brown *et al.*, 1996; Fahey and Allen, 1997; Knight and Kemp, 2000b; Ren *et al.*, 2000).

Multiplication by a window in the latency domain corresponds to convolution with a smoothing function in the frequency domain. Although the two approaches are entirely equivalent, we refer to the technique as ''spectral smoothing'' rather than ''time windowing'' because viewing the process in the frequency domain yields equations for the components that are more directly analogous to those of the suppression method [cf. Eqs. (3)–(5)]. The spectral smoothing strategy for unmixing thus yields the following estimates of P_{dp} and its components:

$$P_{dp} = P_{ec}(f_{dp}) \quad \text{(measured at fixed } f_2/f_1\text{)}; \tag{8}$$

$$P_{dp}^{D} \approx P_{dp}|_{\text{smoothed}} \quad \text{(convolved with smoothing filter)}; \tag{9}$$

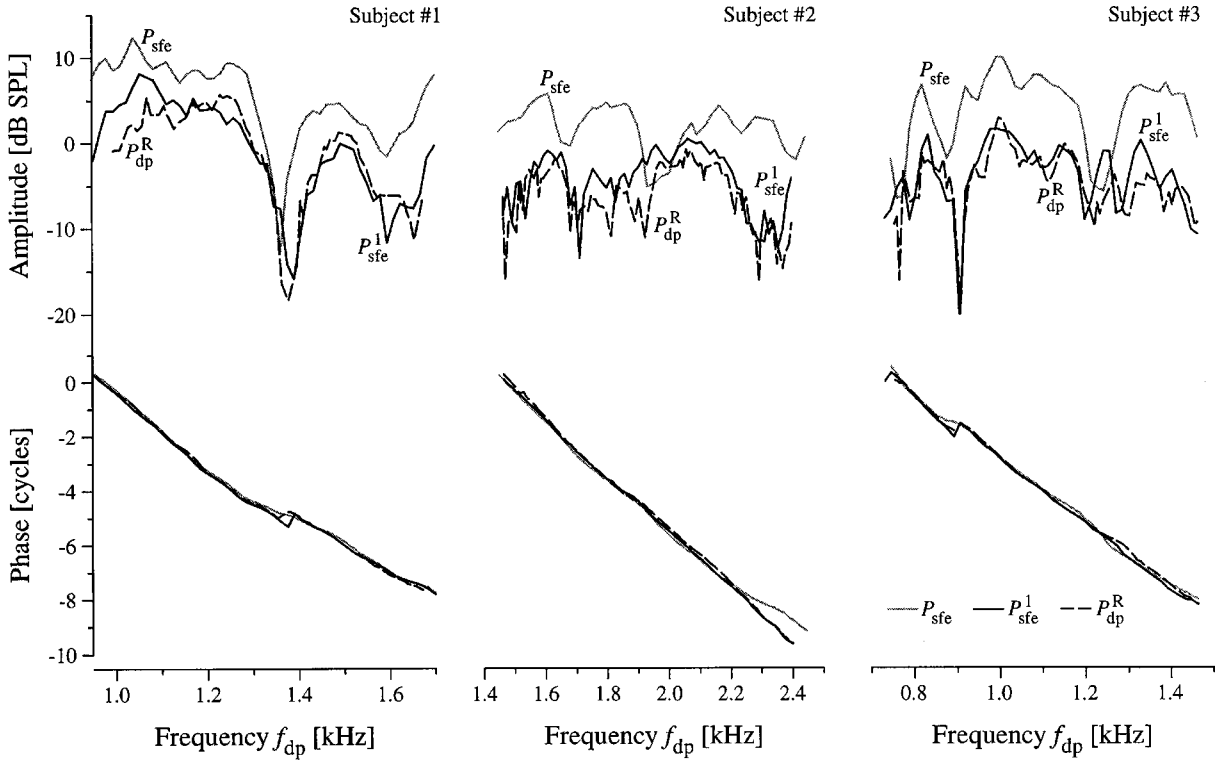


FIG. 5. Comparison between the estimated reflection-source component, P_{dp}^{R}, and the f_1-mimicked SFOAE, P_{sfe}^{1}. The figure compares the amplitude (top) and phase (bottom) of P_{dp}^{R} (dashed line, from Fig. 3) and P_{sfe}^{1} (black solid line), the value of P_{sfe} measured in the presence of an additional tone at the frequency and level of the f_1 primary present during the measurement of P_{dp}^{R}. Left to right, the three panels show data for subjects #1, #2, and #3, respectively. Shown for comparison is P_{sfe} (gray line). The match between the amplitude and phase of P_{dp}^{R} and P_{sfe}^{1} is generally excellent, in agreement with Prediction #2. The differences between P_{sfe} and P_{sfe}^{1} caused by the mimicker suggest that the primaries have a significant effect on the reflection-source component of the DPOAE. SFOAE stimulus levels: $\{L_p, L_s, L_1\} = \{40,55,60\}$ dB SPL.

$$P_{dp}^{R} \approx P_{dp} - P_{dp}^{D}. \tag{10}$$

A. Analysis methods

Measurements of transient-evoked and stimulus-frequency emissions indicate that reflection-emission latency varies with frequency (e.g., Kemp, 1978; Wilson, 1980; Norton and Neely, 1987; Neely *et al.*, 1988; Shera and Guinan, 2000a). This frequency dispersion tends to smear out the reflection-source component in time, making it more difficult to separate by windowing. To help compensate for this dispersion, it proves helpful to work in the log-frequency domain. Consequently, we perform Fourier transforms with respect to the dimensionless frequency variable[5]

$$\nu \equiv -\log(f/f_{ref}), \tag{11}$$

where f_{ref} is a reference frequency taken, for convenience, as the maximum frequency of hearing. Fourier transformation with respect to a log-frequency variable, suggested by the approximate local scaling symmetry of cochlear mechanics, results in sharper, more well-defined peaks in the Fourier-conjugate latency domain (Zweig and Shera, 1995; Knight and Kemp, 2000b).[6] The conjugate dimensionless latency variable, here denoted τ, represents emission latency expressed in periods of the emission frequency (Zweig and Shera, 1995).[7]

Unmixing by smoothing involves convolving P_{dp} with a smoothing function, S, of finite bandwidth (e.g., a Gaussian)[8]:

$$P_{dp}|_{smoothed} \equiv S * P_{dp}, \tag{12}$$

where $*$ denotes the operation of convolution. The convolution is equivalent to a multiplication (or windowing) in the τ domain. Thus,

$$P_{dp}|_{smoothed} = F^{-1}\{\hat{S} \times F\{P_{dp}\}\}, \tag{13}$$

where $F\{\cdot\}$ represents the operation of Fourier transformation (with respect to ν),[9] $F^{-1}\{\cdot\}$ the inverse transformation (with respect to τ), and the window, $\hat{S}(\tau)$, is the Fourier transform of S:

$$\hat{S} \equiv F\{S\}. \tag{14}$$

Separation of P_{dp} into meaningful components requires choosing the smoothing function (or, equivalently, the shape and duration of the latency window) appropriately. Ideally, the window $\hat{S}(\tau)$ should have a sharp cutoff in the τ domain—to cleanly separate emission components of different latencies—but avoid extensive spreading (or ''ringing'') in the frequency response (smoothing function). To approximate these desired characteristics we employ one of a class of ''recursive exponential filters'' (Shera and Zweig, 1993a).[10] The recursive-exponential filters are entire functions and have no poles, discontinuities, or other undesirable features in the complex plane to contribute large oscillations to the smoothing function.

In practice, measurements are only available over a finite frequency range, and the smoothing operation is compli-

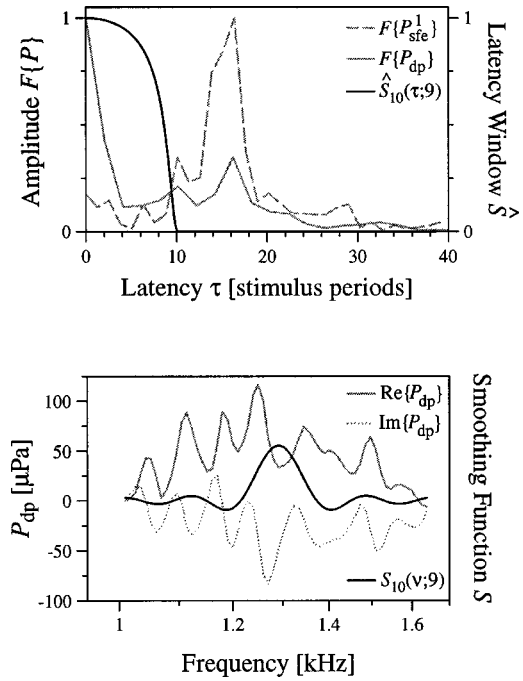


FIG. 6. The smoothing function and corresponding latency window. The figure shows both latency-domain (top) and corresponding frequency-domain representations (bottom) of P_{dp} and the matched smoothing function, S. The top panel shows the amplitudes of the Fourier transforms $F\{P_{dp}\}$ (solid gray line) and $F\{P^1_{sfe}\}$ (dashed gray line) vs τ, the emission latency expressed in stimulus periods. The τ-domain representations of both P_{dp} and P^1_{sfe} show a strong peak centered at a latency of about 15 cycles. Shown for comparison (solid black line) is the tenth-order recursive exponential filter, $\hat{S}_{10}(\tau;\tau_{cut})$, with a cutoff time of $\tau_{cut}=9$ used in subsequent analysis to separate short- and long-latency components of P_{dp}. The bottom panel shows frequency-domain representations of P_{dp} and S. Note that the frequency axis is logarithmic (linear in ν). The real and imaginary parts of P_{dp} are shown with the solid and dotted gray lines, respectively. A linear ramp has been subtracted from P_{dp} to render the function periodic on a cylinder. The black line shows the smoothing function $S_{10}(\nu;9)$ which, when convolved with P_{dp}, yields our estimate of P^D_{dp}. Note that the vertical scale for S, dependent on the number of points in our numerical Fourier transform, is not especially illuminating and has been left unspecified. The measurements of P_{dp} and P^1_{sfe} are from Figs. 3 and 5, respectively (subject #1).

cated by end effects. Throughout this paper, the analyzed frequency range was chosen to include an approximately integral number of spectral cycles, and smoothing was performed using periodic boundary conditions (the data were effectively wrapped around a cylinder). When necessary, a linear ramp was subtracted, and subsequently restored after smoothing, to remove any discontinuity at the ''seam.'' The estimate of P^D_{dp} so obtained was then discarded at each end over a frequency interval equal to the approximate bandwidth of the smoothing function.[11]

1. Determining the window duration

Unmixing via windowing (spectral smoothing) requires specification of the duration of the time window (bandwidth of the smoothing function) used to separate components with different latencies. The suppression studies reported above indicate that the long-latency component, P^R_{dp}, closely matches the characteristics of reflection emissions measured under comparable conditions (e.g., P^1_{sfe}). Consequently, an estimate of the appropriate window duration can be obtained from measurements of SFOAEs evoked at low stimulus levels. Analysis of such measurements indicates that in the 1–2-kHz range, reflection emissions are delayed by an average of about 15 periods of the stimulus frequency with a spread of roughly ±35% (Zweig and Shera, 1995; Shera and Guinan, 2000a). Multiplication by a window of duration $\tau_{cut}=8-9$ periods might therefore be expected to cleanly remove reflection-source components in this frequency range.

Figure 6 corroborates this analysis using our measurements of P_{dp} and P^1_{sfe}. Both short- and long-latency components are clearly apparent in the Fourier transform $F\{P_{dp}\}$, the latency-domain representation of P_{dp}. [By contrast, the long-latency component is almost entirely absent in the Fourier transform of P^D_{dp} obtained by suppression (not shown).] As expected, the long-latency component in $F\{P_{dp}\}$, centered at a latency of about 15 cycles, coincides with the peak in $F\{P^1_{sfe}\}$. The tenth-order recursive exponential filter, $\hat{S}_{10}(\tau;\tau_{cut})$, with a cutoff latency of $\tau_{cut}=9$ periods, is shown for comparison. In subsequent analysis, we use $\hat{S}_{10}(\tau;\tau_{cut}=9)$ to separate the short- and long-latency components of P_{dp}.

B. Results: Unmixing via spectral smoothing

1. P_{dp} and its components, P^D_{dp} and P^R_{dp}, revisited

Typical measurements of P_{dp} and its components unmixed by spectral smoothing are shown with the components obtained by suppression in Fig. 7. Qualitatively, the two methods unmix P_{dp} into similar components. For example, the estimates of P^R_{dp} obtained by the two methods have nearly identical phases and manifest similar frequency dependence in their amplitude curves. There are, of course, differences in the details. For example, the distortion-source components, P^D_{dp}, obtained by suppression unmixing have larger fine structure than the same components obtained by smoothing. We examine this issue further in the next section. Despite differences in detail, the qualitative agreement between the estimated components indicates that our tests of the two-mechanism model are not especially sensitive to the method of unmixing.

IV. ERRORS DUE TO INCOMPLETE UNMIXING

We explore in Fig. 8 the effects of varying key parameters in each of our two unmixing paradigms. For unmixing by suppression, the top panels show how estimates of P^D_{dp} and P^R_{dp} depend on suppressor level, L_s; for unmixing by smoothing (time windowing), the bottom panels show the dependence on the duration of the latency window, τ_{cut} (or, equivalently, the bandwidth of the smoothing function). Note how the fine-structure oscillations in P^D_{dp} (left) increase toward the bottom of each plot (i.e., at lower values of L_s or longer τ_{cut}). By contrast, the fine-structure oscillations in P^R_{dp} (right) increase towards the top (i.e., at higher values of L_s or shorter τ_{cut}).

These systematic trends can be understood using a simple model of the unmixing process. Let the model pressure P_{dp} be the sum of two components, D and R, with very different phase slopes. As a consequence of this difference, D and R beat against each other, producing an oscillatory

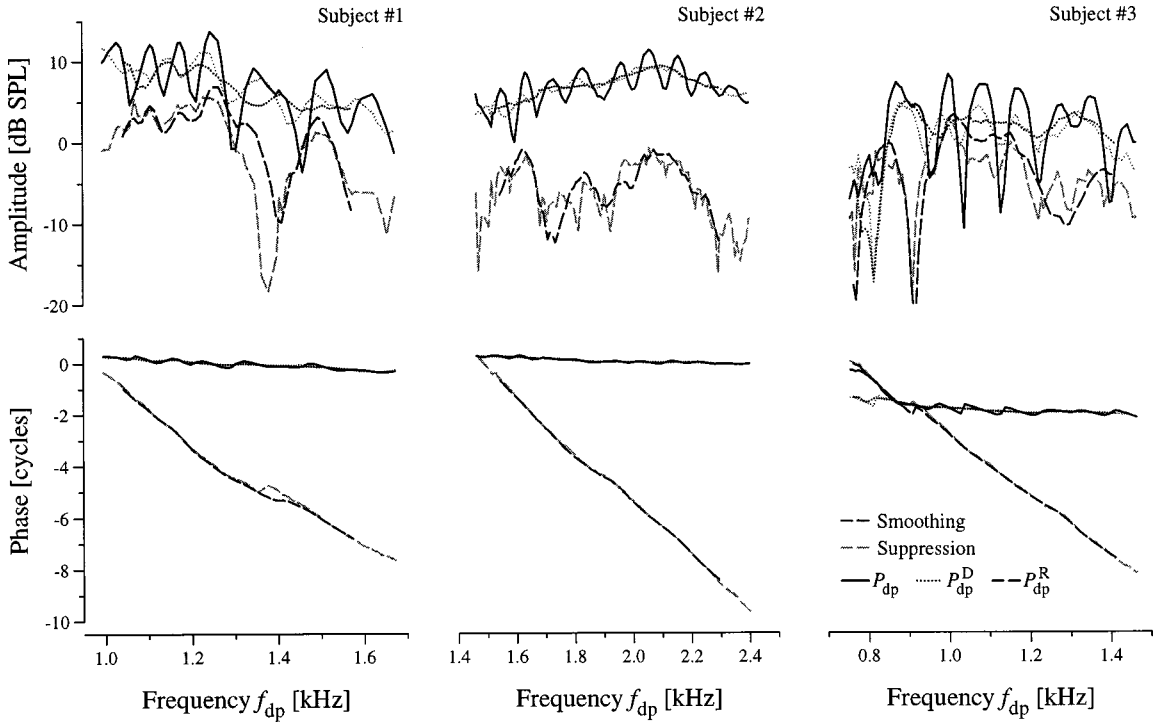


FIG. 7. The DPOAE P_{dp} and its estimated distortion- and reflection-source components, P^D_{dp} and P^R_{dp}, obtained using spectral smoothing (time windowing). The figure shows the amplitude (top) and phase (bottom) of P_{dp} from Fig. 3. Left to right, the three panels show data from subjects #1, #2, and #3, respectively. In each case, the total DPOAE (solid black line) was unmixed as described in the text using the tenth-order recursive exponential filter, $\hat{S}_{10}(\tau;9)$. The two components, P^D_{dp} (dotted black line) and P^R_{dp} (dashed black line), are qualitatively similar to those obtained using suppression (gray lines).

interference pattern in the amplitude and phase of P_{dp}. Imagine now that we attempt to unmix the components experimentally; let our estimates of the two components be denoted P^D_{dp} and P^R_{dp}, respectively. Perfect unmixing would yield values $P^D_{dp}=D$ and $P^R_{dp}=R$. In general, however, unmixing is incomplete, and the estimates contain contributions from both D and R:

$$\begin{pmatrix} P^D_{dp} \\ P^R_{dp} \end{pmatrix} = \begin{pmatrix} 1-\delta & \rho \\ \delta & 1-\rho \end{pmatrix} \begin{pmatrix} D \\ R \end{pmatrix}, \tag{15}$$

where the complex, frequency-dependent coefficients δ and ρ quantify the unmixing errors. Note that the coefficients satisfy the constraint $P^D_{dp}+P^R_{dp}=D+R$. Although perfect unmixing requires $\delta=\rho=0$, acceptable results occur with $|\delta| \ll 1$ and $|\rho| \ll 1$.

The unmixing errors δ and ρ depend on unmixing parameters such as the level of the suppressor and the duration of the latency window. To explicate the trends in Fig. 8, we consider three special cases of incomplete unmixing:

(1) Case $\delta=0$ and $\rho\neq 0$ so that

$$P^D_{dp}=D+\rho R; \tag{16}$$

$$P^R_{dp}=(1-\rho)R. \tag{17}$$

For suppression-based unmixing, this case results from using a weak suppressor that leaves D unchanged but only incompletely removes the R component from the mix (i.e., undersuppression); in the smoothing case, it corresponds to under-smoothing (i.e., to using too narrow a smoothing function or too long a latency window). Since the resulting estimate of P^D_{dp} appears contaminated by the R component, the magnitude $|P^D_{dp}|$ should oscillate with frequency. These features are found in the figure: at smaller values of L_s and longer values of τ_{cut}, the estimates L^D_{dp} manifest considerable fine structure.

(2) Case $\delta\neq 0$ and $\rho=0$ so that

$$P^D_{dp}=(1-\delta)D; \tag{18}$$

$$P^R_{dp}=R+\delta D. \tag{19}$$

Here, the suppressor is strong enough to completely remove the R component, but in so doing it modifies the D; for smoothing, this case results from oversmoothing (i.e., using an overly broad smoothing function or too short a latency window). In this case, the estimate P^R_{dp} is contaminated with part of the D component, and its magnitude should therefore oscillate. These features occur in Fig. 8: at the largest values of L_s and shortest values of τ_{cut}, the estimates L^R_{dp} show evidence of fine structure.

(3) Case $\delta\neq 0$ and $\rho\neq 0$ so that

$$P^D_{dp}=(1-\delta)D+\rho R; \tag{20}$$

$$P^R_{dp}=(1-\rho)R+\delta D. \tag{21}$$

In this more general case, the suppressor is neither strong enough to eliminate the R component nor weak enough not to affect the D component. For smoothing, this case results from a temporal overlap between the D and R components in the latency domain. In this situation, both

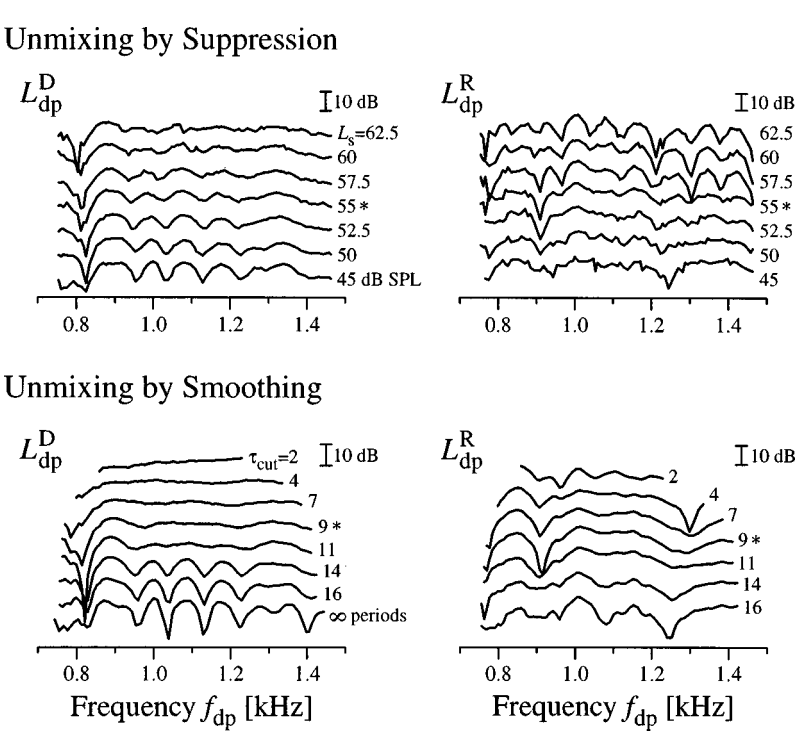


FIG. 8. Changes in estimates of $|P_{dp}^{D}|$ and $|P_{dp}^{R}|$ with variations in the parameters of the unmixing process. The figure shows the levels L_{dp}^{D} (left) and L_{dp}^{R} (right) obtained when the parameters for the suppression- and smoothing-based unmixing are varied systematically. The top panels show the results obtained by varying the level of the suppressor tone, L_s. The bottom panels show the results of varying the duration of the latency window, τ_{cut}. The original, unsmoothed measurement of L_{dp} is shown for comparison in the bottom left ($\tau_{cut}=\infty$). Note that because the estimates of P_{dp}^{D} and P_{dp}^{R} were discarded at each end over a frequency interval equal to the bandwidth of the smoothing function (see Sec. IV A), the estimates cover a more limited frequency range at smaller values of τ_{cut}. In all panels, the different curves have been offset from one another for clarity. Unmixing parameters used earlier in the paper ($L_s=55$ dB SPL and $\tau_{cut}=9$ periods) are marked with an asterisk. Data from subject #3 with stimulus levels of $\{L_1, L_2\}=\{60,45\}$ dB SPL.

P_{dp}^{D} and P_{dp}^{R} will show fine structure oscillations, as seen in Fig. 8 at certain intermediate values of L_s and τ_{cut}.

Our results with suppression unmixing suggests that there is no ''ideal'' suppressor level valid over a range of frequencies that simultaneously eliminates the reflection-source component while leaving the distortion-source essentially unaffected. Figure 8, for example, shows some residual fine structure in both P_{dp}^{R} and P_{dp}^{D} at this subject's ''optimal'' suppressor level of $L_s=55$ dB SPL. The choice of suppressor level involves a trade-off between minimizing $|\delta|$ and minimizing $|\rho|$, with their sum inevitably finite. With proper choice of the windowing function, the prospects for near ideal unmixing by spectral smoothing (time windowing) appear brighter. Judging by the almost negligible amplitude of the fine structure obtained at intermediate values of τ_{cut}, unmixing by smoothing appears able to effect a cleaner separation between the two components than is possible using suppression.

A. Estimating δ and ρ

We illustrate the trade-off between δ and ρ and give a feel for the suppression-based unmixing errors in Fig. 9, which shows estimates of $|\delta|$ and $|\rho|$ for three different suppressor levels. Computation of δ and ρ requires knowledge of D and R; the estimates in Fig. 9 were computed by substituting for D and R the components obtained by spectral smoothing (with $\tau_{cut}=9$). Since the two equations represented in matrix form in Eq. (15) are not independent (but are related by $P_{dp}^{D}+P_{dp}^{R}=D+R$), an additional constraint is necessary to determine δ and ρ uniquely. Since two parameters are available, a natural choice is to match both P_{dp}^{D} and its frequency derivative, $P_{dp}^{D'}$. We thus obtain values of δ and ρ by solving the pair of simultaneous equations

$$P_{dp}^{D}=(1-\delta)D+\rho R; \tag{22}$$

$$P_{dp}^{D'}=(1-\delta)D'+\rho R'. \tag{23}$$

The values of δ and ρ obtained in this way vary with frequency; at every point, the coefficients δ and ρ are chosen to match to the curve P_{dp}^{D}, both its value and its derivative, as closely as possible.[12] Because of the constraint $P_{dp}^{D}+P_{dp}^{R}=D+R$, these same coefficients also provide a match to P_{dp}^{R} and its derivative.

Since the true components D and R are not known, the unmixing errors δ and ρ shown in Fig. 9 were computed relative to the components obtained by spectral smoothing; they therefore provide only a rough guide to the actual errors. The results are, however, generally consistent with expectations based on the three special cases of Eq. (15) considered above. Note, for example, that $|\delta|$ and $|\rho|$ vary in opposite directions with changes in suppressor level. At the largest suppressor level, $|\delta|$ is relatively large and $|\rho|$ relatively small (corresponding to the expectations for strong suppression outlined in case #2 above). Similarly, at the smallest suppressor level, the relative magnitudes of δ and ρ are reversed (weak suppression, as in case #1). At the ''optimal'' suppressor level, the errors $|\delta|$ and $|\rho|$ are intermediate between these extremes. Not surprisingly, $|\delta|$ and $|\rho|$ can become large in frequency regions where the total DPOAE is itself poorly determined (e.g., near 0.8 kHz where $|P_{dp}|$ is relatively close to the noise floor) and/or where the estimated components P_{dp}^{D} and P_{dp}^{R} change rapidly (e.g., near notches of P_{dp}^{R}). Overall, however, the unmixing errors are fairly small for intermediate suppressor levels (typically $|\delta|\sim 0.1$ and $|\rho|\sim 0.2-0.3$). These findings corroborate the qualitative visual impression that the two methods, selective suppression and spectral smoothing, unmix into generally similar components.

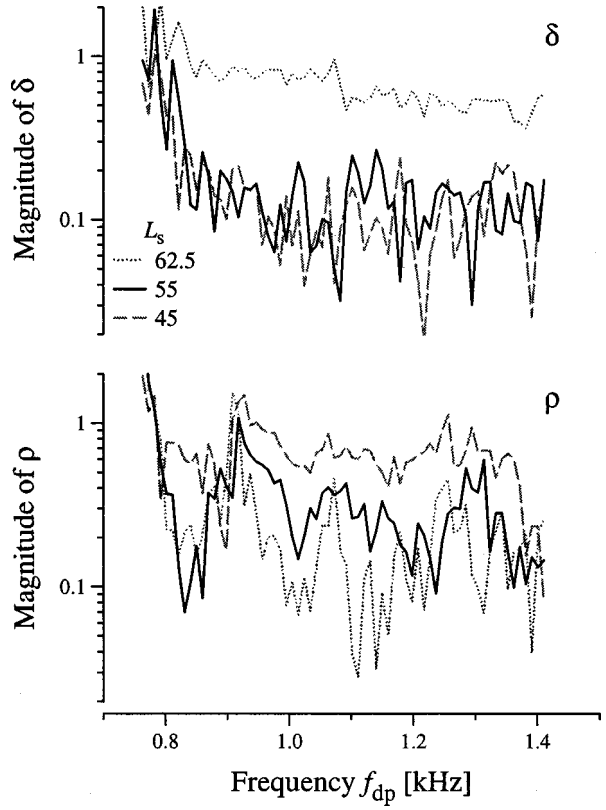


FIG. 9. Estimates of the unmixing errors δ and ρ at different suppressor levels. The figure shows magnitudes of the unmixing errors δ and ρ computed as the solution to Eqs. (22) and (23). The components P_{dp}^{D} and P_{dp}^{R} obtained by spectral smoothing (with $\tau_{cut}=9$) were used as estimates of D and R. Results for three different suppressor levels were computed using the data whose magnitudes are shown in the top panels of Fig. 8.

V. DISCUSSION

In this paper we tested the two key predictions of the two-mechanism model of DPOAE generation by successfully unmixing DPOAEs into components, P_{dp}^{D} and P_{dp}^{R}, with characteristics indicative of fundamentally different mechanisms of generation (i.e., nonlinear distortion vs linear reflection). In agreement with Prediction #1, the phase of the putative distortion-source component (P_{dp}^{D}) is nearly constant, whereas the phase of the reflection-source component (P_{dp}^{R}) varies rapidly with frequency. These differing phase slopes imply fundamental differences in the respective mechanisms of emission generation. In particular, the two slopes are consistent with generation by nonlinear distortion (P_{dp}^{D}) and linear coherent reflection (P_{dp}^{R}), respectively (Shera and Guinan, 1999). Furthermore, in agreement with Prediction #2, the spectral shape and phase of the reflection-source component closely match those of the SFOAE measured at the same frequency under comparable conditions (i.e., with the addition of an f_1-primary mimicker). Changes in the SFOAE caused by the mimicker suggest that the primaries have a significant influence on the reflection-source component of the DPOAE, presumably via suppression. To investigate the robustness of our conclusions, we unmixed DPOAE sources using two completely different methods: (a) selective suppression of the reflection source using a third tone near the distortion-product frequency, and (b) spectral smoothing (or, equivalently, time-domain windowing). Although the two methods unmix in very different ways, explicit analysis of the unmixing errors demonstrates that they yield similar DPOAE components, indicating that our results are not especially sensitive to the method of unmixing.

A. Source mechanism versus source location

The quasiperiodic fine structure often evident in DPOAE spectra is now generally regarded as resulting from the alternating constructive and destructive interference between backward-traveling waves originating in two separate regions of the cochlea (Kim, 1980). The physics underlying the interference pattern has generally been understood as follows (e.g., Brown *et al.*, 1996; Stover *et al.*, 1996; Fahey and Allen, 1997): Because the two sources are spatially separated, backward-traveling waves originating at the more apical location must travel further to reach the ear canal than waves originating at the basal location. Consequently, waves from the apical source are delayed relative to the basal source; in the frequency domain, this delay corresponds to a frequency-dependent phase shift. Thus, the relative phase of the two waves rotates with frequency, alternately passing through plus and minus 1. This rotation of relative phase creates the interference pattern—known as DPOAE fine structure—when the two waves are combined in the ear canal. Kim (1980) originally referred to the two DPOAE sources as the ''primary-place source'' and the ''characteristic-place source,'' and considerable evidence now suggests that the two backward-traveling waves do indeed originate at these locations (e.g., Furst *et al.*, 1988; Gaskill and Brown, 1990; Brown *et al.*, 1996; Engdahl and Kemp, 1996; Brown and Beveridge, 1997; Talmadge *et al.*, 1998, 1999; Heitmann *et al.*, 1998; Fahey and Allen, 1997; Siegel *et al.*, 1998; Mauermann *et al.*, 1999a, 1999b).

We demonstrate here, however, that this place-based nomenclature—and the conceptual model that underlies it (e.g., Brown *et al.*, 1996; Stover *et al.*, 1996; Fahey and Allen, 1997)—although apparently accurate in its specification of the locations of wave origin, fails to capture the critical distinction between the two sources. As suggested by Shera and Guinan (1999), the *fundamental distinction between the two sources is evidently not source* **location**, *but source* **mechanism**. Indeed, only by incorporating both classes of emission-generating mechanisms (i.e., nonlinear distortion and linear coherent reflection) have models been able to account for the known phenomenology of DPOAE fine structure (e.g., Talmadge *et al.*, 1998, 1999; Mauermann *et al.*, 1999a). Accordingly, our terminology distinguishes the two components not by their place of origin, but by their mechanism of generation (i.e., distortion- versus reflection-source components).

Our results support the two-mechanism model of DPOAE generation. To illustrate, consider how our experimental results would have differed if both sources in Fig. 1 had been distortion sources like **D**. When probed with the frequency-scaled stimuli used here, both sources would then have generated backward-traveling waves with phases essentially independent of frequency. Consequently, the relative phase of the waves from the two sources would have been

nearly constant, and no oscillatory fine structure would have appeared in the ear-canal pressure spectrum. *Note that this constancy of relative phase would have occurred despite the fact that the two waves originate at different spatial locations within the cochlea*. In other words, although the reflection-source region at **R** is further from the stapes than the distortion-source region at **D**, the difference in phase slope characterizing emissions from these two sources is not due to the differing locations of the **D** and **R** regions. Rather, contrary to standard assumption, phase slopes are ultimately determined by mechanisms of emission generation. For example, the theory of coherent reflection filtering (Shera and Zweig, 1993b; Zweig and Shera, 1995) implies that reflection-emission latency is determined not by the distance a wave travels to reach its characteristic place but by the characteristics of cochlear tuning it finds when it gets there (Shera and Guinan, 2000a, 2000b).

B. Comparison with other work

The experiments reported here were designed specifically to test Predictions #1 and #2 and therefore differ from most other studies of DPOAE components (e.g., Brown *et al.*, 1996; Stover *et al.*, 1996; Siegel *et al.*, 1998) in their use of frequency-scaled stimuli (i.e., fixed f_2/f_1). According to the analysis underlying the model (Shera and Guinan, 1999), distortion and reflection mechanisms yield qualitatively different phase behavior (i.e., nearly constant phase vs rapidly rotating phase) when emissions are evoked with frequency-scaled stimuli. Similar qualitative differences in phase are not found using other measurement paradigms, and the underlying differences in mechanism can therefore be considerably less transparent. For example, much more rapid phase rotation occurs when distortion emissions are measured using stimulus paradigms (e.g., fixed f_1, fixed f_2, or fixed f_{dp}) for which the cochlear wave pattern is not simply translated along the cochlear partition (e.g., Kimberley *et al.*, 1993; O'Mahoney and Kemp, 1995; Shera *et al.*, 2000). Unmixing DPOAEs measured using constant f_2/f_1-ratio sweeps, rather than one of the more common alternative paradigms, thus greatly facilitates recognition of the two emission mechanisms. By increasing the difference in phase slope between the distortion- and reflection-source components, our use of frequency-scaled stimuli also facilitates unmixing of the two components using spectral smoothing (time windowing) by maximizing the separation between the two components when the data are transformed into the ''latency domain'' using Fourier analysis.

Our tests of Prediction #2 contrast sharply with the findings of Brown *et al.* (1996), who performed a DPOAE unmixing analysis using a smoothing technique and compared the resulting ''DP residual'' (their analog of P_{dp}^{R}) to measurements of SFOAEs. Although they noted similarities in the phase slopes, they found ''little correspondence in the magnitude across frequency of the DP residual and SF[OA]E.'' Their reported discrepancy between emission components conflicts with earlier work (Kemp and Brown, 1983), which found at least qualitative agreement between the SFOAE and the DPOAE component believed to originate at the distortion-product place (as obtained, in this case, using a suppression paradigm). In contrast with these results, we find excellent agreement, both between DPOAE components unmixed via different paradigms and between P_{dp}^{R} and corresponding SFOAEs. Unfortunately, Brown *et al.* (1996) do not specify their smoothing algorithm in the detail necessary to enable a direct comparison with our method.[13] We note, however, that in our experiments the addition of the f_1-primary mimicker often improved the agreement between the magnitudes of P_{dp}^{R} and P_{sfe} considerably (cf. Fig. 5). This result indicates that suppressive and other effects of the primaries on P_{dp}^{R} must be taken into account in any such comparison.

C. Region of validity of the two-mechanism model

The tests of Predictions #1 and #2 reported here, together with more limited data at other (low to moderate) primary levels and at frequency ratios f_2/f_1 in the range 1.1–1.3, establish the validity of the two-mechanism model in humans for the DPOAE measurement parameters in common use [i.e., low to moderate sound-pressure levels with $L_1 \geqslant L_2$ and primary frequency ratios $f_2/f_1 \approx (f_2/f_1)_{optimal}$]. Knight and Kemp (2000a) provide a test of Prediction #1 over a broad range of frequency ratios ($1.01 \leqslant f_2/f_1 \leqslant 1.5$) in an unmixing analysis of their stunning $\{f_1, f_2\}$-area map (Knight and Kemp, 2000b). Their results, based on time windowing of DPOAEs measured using primary levels $L_1 = L_2 = 70$ dB SPL, are consistent with the two-mechanism model and indicate that the relative amplitudes of the components P_{dp}^{D} and P_{dp}^{R} vary systematically with f_2/f_1. Whether Prediction #2 also applies over a similarly broad range of parameter values remains an important open question.

Described and tested here in the frequency domain, Prediction #1 of the two-mechanism model evidently also applies in the time domain. Combining phase-rotation averaging (Whitehead *et al.*, 1996) with an elegant pulsed-primary technique, Talmadge *et al.* (1999) provide strong support for model predictions that amount, in effect, to time-domain analogs of Prediction #1. Since the responses involved arise in a nonlinear system, this conclusion is nontrivial. Time-domain tests of Prediction #2 await further experiment.

The validity of the model at high intensities also remains to be investigated. For example, at higher levels of intracochlear distortion, the emission evoked by the forward-traveling distortion component may contain, in addition to contributions from coherent reflection, significant energy from distortion-source waves created by nonlinear distortion (e.g., Withnell and Yates, 1998). Furthermore, the two emission sources may also begin to mix in ways more complicated than simple linear summation.[14] For example, the strength of the micromechanical impedance perturbations that scatter the traveling wave may depend on the local amplitude of basilar-membrane vibration.

D. Methods of unmixing

Our success at unmixing using two completely different methods (suppression and smoothing) demonstrates the robustness of our conclusions to the method of unmixing. The

two methods unmix in very different ways, and the systematic errors each introduces are presumably quite different. Whereas the suppression method separates components based on their differential modification by an external tone, the spectral-smoothing (or time-domain windowing) method separates components based on latency in the ''time-domain response'' obtained using Fourier analysis.[15] Despite these differences, the two methods unmix the total emission into rather similar components (at least for $f_2/f_1=1.2$ and low to moderate primary levels).[16] Whether the two methods yield similar results at other f_2/f_1 ratios and/or at higher stimulus levels remains an important open question. Differences between the methods would not be surprising at f_2/f_1 ratios close to 1—although the spectral-smoothing method does not depend on spatial separation of source regions in the cochlea, the ability of the suppression method to selectively eliminate one of the sources presumably deteriorates as the two sources draw closer to one another as f_2/f_1 approaches 1.

An advantage of the spectral-smoothing method is that it requires measurement of only a single quantity (namely, P_{dp}, whereas the suppression method requires both P_{dp} and $P_{\mathrm{dp}}|_{\mathrm{suppressed}}$). Unlike the suppression method, the smoothing method therefore allows each measurement of P_{dp} to serve as its own control against possible systematic changes (e.g., variations in overall emission level due to efferent effects) that may occur during the course of the measurement. In the suppression studies reported here, we sought to minimize these potential problems by interleaving measurements of P_{dp} and $P_{\mathrm{dp}}|_{\mathrm{suppressed}}$ in time. Although the spectral smoothing method depends only on P_{dp}, it requires knowledge of P_{dp} at multiple frequencies. Indeed, the method works best if applied to measurements that span a relatively wide frequency range (i.e., many periods of the microstructure) with good frequency resolution (i.e., many points per period). In addition, because of uncertainties introduced near the end points due to incomplete knowledge of P_{dp} outside the measured interval, the smoothing method requires measurements over an interval slightly larger than the desired frequency range. The suppression method, by contrast, imposes no such constraints; suppression unmixing requires measurement of P_{dp} and $P_{\mathrm{dp}}|_{\mathrm{suppressed}}$ only at the actual frequency (or frequencies) of interest.

E. Implications of unmixing DPOAEs

Uncontrolled mixing may be a substantial source of subject-dependent variability in DPOAE measurements. Indeed, our results imply that the interpretation of DPOAE responses appears doubly confounded. First, DPOAEs are mixtures of emissions originating from at least two different regions in the cochlea. This ''spatial blurring,'' now widely recognized, compromises the frequency selectivity of DPOAE measurements (e.g., Heitmann *et al.*, 1998). Second, DPOAEs are mixtures of emissions arising by fundamentally different mechanisms. This ''mechanistic blurring,'' established here, compromises the etiological specificity of DPOAE measurements. For although both distortion- and reflection-source emissions share a common dependence on propagation pathways from the cochlea to the ear canal, and are therefore both sensitive to modifications of that pathway (e.g., to middle-ear pathology or to reductions in cochlear amplification caused by damage to outer hair cells), their respective mechanisms of generation—and hence their dependence on underlying parameters of cochlear mechanics—remain fundamentally distinct. For example, whereas distortion-source emissions presumably depend on the form and magnitude of cochlear nonlinearities (e.g., on the effective ''operating point'' along hair-cell displacement-voltage transduction functions), reflection-source emissions depend strongly on the size and spatial arrangement of micromechanical impedance perturbations (e.g., on variations in hair-cell number and geometry). Distortion-product unmixing, using techniques such as those employed here, should therefore improve the power and specificity of DPOAEs as noninvasive probes of cochlear function.

ACKNOWLEDGMENTS

We thank our experimental subjects for their considerable patience and gratefully acknowledge the efforts of Paul Fahey, John Guinan, William Peake, and two anonymous reviewers, who provided valuable comments on the manuscript. This work was supported by Grants Nos. R01 DC03687 and T32 DC00038 from the NIDCD, National Institutes of Health.

APPENDIX MEASUREMENT METHODS

This Appendix describes in more detail the methods used to obtain the emission measurements reported here.

1. Measurement of DPOAEs

Distortion-product otoacoustic emissions at the frequency $2f_1-f_2$ were measured using frequency-scaled stimuli (i.e., with the ratio f_2/f_1 held constant). At each measurement frequency the acoustic stimulus had the form

$$\mathrm{stimulus}=\underbrace{\langle XX\cdots X\rangle}_{\#\geqslant M}, \tag{A1}$$

where X represents a periodic (5×4096)-sample ($\approx$342 ms[17]) segment consisting of three components:

$$X=\begin{cases}\pi_1^1\Pi_2^1\Pi_3^1\pi_4^1\pi_5^1\pi_6^1\Pi_7^1\pi_8^1 & (\text{primary earphone \#1})\\ \pi_1^2\Pi_2^2\pi_3^2\pi_4^2\pi_5^2\Pi_6^2\Pi_7^2\pi_8^2 & (\text{primary earphone \#2}).\\ o_1O_2O_3\langle{}_4\sigma_5\Sigma_6\Sigma_7\rangle_8 & (\text{suppressor earphone})\end{cases} \tag{A2}$$

Each component consisted of four long intervals (uppercase) and four short intervals (lowercase and angled brackets). The long intervals were each 4096 samples ($\approx$68 ms) in duration. The primary segments, Π_i^1 and Π_i^2, contained an integral number of periods of the primary frequencies, f_1 and f_2, respectively. The suppressor segments, Σ_i, contained an integral number of cycles of the suppressor frequency, f_s. The zero segments, O_i, were identically zero throughout. Waveform phases were adjusted, using information from the calibration procedure, so that each stimulus had zero (cosine) phase in the ear canal at the beginning of segment Π_2. The short intervals were one fourth the duration of the long intervals (i.e., 1024 samples or $\approx$17 ms) and did not, in gen-

eral, contain an integral number of periods of the corresponding waveform. The short intervals π_i, σ_i, and o_1 allowed for response settling time and contained segments of the primary, suppressor, or zero waveforms, respectively. The short intervals $\{\langle_4,\rangle_8\}$ were used to ramp the suppressor tone {on, off} using the {first, second} half of the Blackman window. The three components of X were synchronized and presented simultaneously through three separate earphones. Note that whereas the primary tones played continuously during the measurement, the suppressor tone cycled on and off repeatedly due to alternation of the zero and suppressor waveforms. Interleaving the measurements of P_{dp} and $P_{\mathrm{dp}}|_{\mathrm{suppressed}}$ in this way helps to minimize possible artifacts due to systematic variations over time (e.g., due to subject movement, drifts in earphone calibration, efferent feedback, etc.). Unless otherwise noted, the primary levels $\{L_1,L_2\}$ were {60, 45} dB SPL, respectively. Primary levels were chosen in approximate accordance with the formula $L_1 \approx 0.4L_2+39$ dB SPL, which tracks the ''ridge'' in the L_1L_2 plane that maximizes the $2f_1-f_2$ emission for $f_2/f_1 \approx 1.2$ (Kummer *et al.*, 1998).

Measurements were made versus probe frequency by sweeping the primaries and suppressor from high frequencies to low, with $f_s=f_{\mathrm{dp}}+\Delta f_s$ and $\Delta f_s=-44$ Hz. The periodic segments X were played repeatedly until M artifact-free responses were collected. In these measurements, M was typically 64 so that at each frequency the total stimulus duration was therefore $\gtrsim 64\times 342\,\mathrm{ms}\approx 22\,\mathrm{s}$. To reduce unwanted transients the probe waveform was ramped on and off by pre- and postpending two additional segments [indicated by the angled brackets $<$ and $>$ in Eq. (A1)] with envelopes of half Blackman windows with 2.5-ms rise and fall times. After digitizing the resulting ear-canal pressure, responses to all primary-alone segments (i.e., all segments Π_2 and Π_3) were averaged to form Y_{p}; similarly, the responses to all probe+suppressor segments (i.e., all segments Π_6 and Π_7) were averaged to form $Y_{\mathrm{p+s}}$. From these averaged response waveforms, the complex amplitudes of the f_{dp} components of the ear-canal pressure, denoted $P_{\mathrm{dp}}=P_{\mathrm{ec}}(f_{\mathrm{dp}})$ and $P_{\mathrm{dp}}^{\mathrm{D}}=P_{\mathrm{ec}}(f_{\mathrm{dp}})e^{-2\pi i\Delta N\Delta T f_{\mathrm{dp}}}|_{\mathrm{suppressed}}$, were extracted using Fourier analysis. The complex exponential compensates for the phase shift in the probe due to the time interval, $\Delta N\Delta T$, between the primary-alone and primary+suppressor segments. Here, ΔT is the sampling interval (reciprocal of the sampling rate), and ΔN represents the total number of these intervals that separate the two segments:

$$\Delta N=\#\mathrm{samples}(\Pi_2\Pi_3\pi_4\pi_5)=2\tfrac{1}{2}\times 4096=10\,240. \tag{A3}$$

Note that when the two segments are separated by an integral number of periods of the f_{dp} waveform, the phase shift modulo 2π is zero. The complex quantity $P_{\mathrm{dp}}^{\mathrm{R}}(f_{\mathrm{dp}})$ was then obtained as

$$P_{\mathrm{dp}}^{\mathrm{R}}=P_{\mathrm{dp}}-P_{\mathrm{dp}}^{\mathrm{D}}. \tag{A4}$$

2. Measurement of SFOAEs

Stimulus-frequency emissions were measured using the suppression method detailed elsewhere (Shera and Guinan, 1999). In some experiments, we measured SFOAEs in the presence of an additional continuous tone (the ''f_1-primary mimicker'') at a frequency and level corresponding to the f_1 primary in the measurement of DPOAEs detailed above.

At each measurement frequency the acoustic stimulus had the form given by Eq. (A1), with X representing a periodic (5×4096)-sample ($\approx$342 ms) segment consisting of three components:

$$X=\begin{cases}\pi_1\Pi_2\Pi_3\pi_4\pi_5\Pi_6\Pi_7\pi_8 & (\text{probe earhone})\\ o_1O_2O_3\langle_4\sigma_5\Sigma_6\Sigma_7\rangle_8 & (\text{suppressor earphone})\\ \mu_1M_2M_3\mu_4\mu_5M_6M_7\mu_8 & (\text{primary-mimicker earphone})\end{cases}. \tag{A5}$$

Each component consisted of four long (uppercase) and four short (lowercase and angled brackets) intervals. The long intervals were each 4096 samples ($\approx$68 ms) in duration and contained an integral number of periods of the probe (Π_i), suppressor (Σ_i), zero (O_i), or primary mimicker (M_i) waveforms, respectively. The phase of the probe waveform was adjusted, using information from the calibration procedure, so that the stimulus had zero (cosine) phase in the ear canal at the beginning of segment Π_2. The short intervals were one fourth the duration of the long intervals (i.e., 1024 samples or $\approx$17 ms) and did not, in general, contain an integral number of periods of the corresponding waveform. The short intervals π_i, σ_i, o_1, and μ_i allowed for response settling time and contained segments of the probe, suppressor, zero, and mimicker waveforms, respectively. The short intervals $\{\langle_4,\rangle_8\}$ were used to ramp the suppressor tone {on, off} using the {first, second} half of the Blackman window. The three components of X were synchronized and presented simultaneously through three separate earphones. Note that whereas the probe and primary mimicker tones played continuously during the measurement, the suppressor tone cycled on and off repeatedly due to alternation of the zero and suppressor waveforms. The probe and suppressor levels $\{L_{\mathrm{p}},L_{\mathrm{s}}\}$ were generally {40, 55} dB SPL. The primary mimicker was presented at a frequency and level corresponding to the f_1 primary in the measurement of DPOAEs (i.e., at a frequency equal to $f_1=f_{\mathrm{p}}/(2-r)$, where r denotes the f_2/f_1 ratio we wished to mimic, and a typical level of 60 dB SPL).

Other features of the stimulus paradigm and the subsequent data analysis used to compute P_{sfe}^1 are analogous to the measurement of DPOAEs detailed above and have been described elsewhere (Shera and Guinan, 1999).

[1]Note that for brevity this simple synopsis neglects contributions to the total reflection-source emission arising from multiple internal reflection within the cochlea (i.e., from multiple cycles of partial reflection at the stapes and linear coherent reflection within the **R** region).

[2]In a nutshell, the theory says that given ''almost any'' arrangement of micromechanical impedance perturbations (i.e., an arrangement with the appropriate spatial-frequency content, such as perturbations that are randomly and densely distributed), a model will produce realistic reflection emissions whenever the peak region of the traveling wave has a slowly varying wavelength and an envelope that is simultaneously both tall and broad.

[3]Some differences (e.g., in phase) between $P_{\mathrm{dp}}^{\mathrm{R}}$ and P_{sfe} are, of course, expected because the initial sources of forward-traveling cochlear waves at the emission frequency are at different spatial locations in the two cases (i.e., at the distortion-source region, **D**, for $P_{\mathrm{dp}}^{\mathrm{R}}$, and at the stapes for P_{sfe}).

[4]We put ''latency-domain response'' in quotes because the signal we obtain

by Fourier transforming the frequency response does not correspond with the time-domain impulse response of the system.

[5]The minus sign in Eq. (11) has the effect of converting a forward Fourier transform (with respect to ν) into an inverse transform (with respect to $\log f/f_{\text{ref}}$). We work with forward Fourier transforms (with respect to ν) for conceptual and numerical convenience.

[6]Measurements of tone-burst-evoked OAE and ABR latency (Neely *et al.*, 1988), as well as measurements of SFOAE group delay (Shera and Guinan, 2000a), all indicate a gradual breaking of scaling symmetry in the basal turns of the mammalian cochlea. For near-optimal compensation for traveling-wave dispersion, the measurements suggest working with the variable $-\sqrt{f/f_{\text{ref}}}$ (see also Shera *et al.*, 2000).

[7]Our τ scale differs from the time scale employed by Knight and Kemp (2000b), who apply a log-frequency transformation and plot the resulting Fourier transforms against an axis they call ''normalized time.'' Their normalized time has units of milliseconds and was computed based on the mean frequency step size in the log-frequency scale. The two scales differ by a multiplicative factor inversely proportional to the geometric mean frequency of the analyzed data.

[8]Note that unlike the more familiar case of time-domain filtering, the oscillatory function to be removed occurs here in the frequency response. In its reversal of the roles usually played by time and frequency, the technique used here is similar to cepstral analysis (Bogert *et al.*, 1963), although we work with a log-frequency variable, ν, and analyze P_{dp} rather than $\log(P_{\text{dp}})$. [In cepstral analysis, one takes the logarithm of the frequency response in order to decompose a presumed product of spectra into a sum. In our application, the pressure P_{dp} is represented directly as a sum of components (Prediction #1); taking the logarithm is therefore both unnecessary and undesirable.]

[9]To perform our transforms numerically, we resampled our measurements of P_{dp} at equal intervals in log frequency using cubic spline interpolation. Because our sampling rate was variable, our measurements of P_{dp} were not equally spaced in linear frequency.

[10]The nth order recursive-exponential filtering window is defined by (Shera and Zweig, 1993a)

$$\hat{S}_n(\tau;\tau_{\text{cut}}) \equiv 1/\Gamma_n(\lambda_n \tau/\tau_{\text{cut}}),$$

where the parameter τ_{cut} is the cutoff time (length of the window) and the function $\Gamma_n(\tau)$ is defined recursively:

$$\Gamma_{n+1}(\tau) = e^{\Gamma_n(\tau)-1}, \text{ with } \Gamma_1(\tau) = e^{\tau^2}.$$

The window $\hat{S}_n(\tau;\tau_{\text{cut}})$ has a maximum value of 1 at $\tau=0$; the scale factor λ_n is chosen so that the window falls to the value $1/e$ at $\tau=\tau_{\text{cut}}$:

$$\lambda_n = \sqrt{\gamma_n}, \text{ where } \gamma_{n+1} = \ln(\gamma_n + 1) \text{ with } \gamma_1 = 1.$$

Note that the first-order filtering window is a simple Gaussian; in the limit $n\to\infty$, $\hat{S}_n$ approaches a rectangular (or boxcar) window. For intermediate n (e.g., the value $n=10$ used here), $\hat{S}_n$ has a much sharper cutoff than standard windowing functions (e.g., the Hamming, Blackman, etc.) and considerably less ''ringing'' in the smoothing function than the simple boxcar.

[11]The smoothing function has approximate width (Shera and Zweig, 1993a)

$$\Delta\nu = \Delta f/f \geqslant 1/\pi\tau_{\text{cut}}.$$

[12]The quantities δ and ρ are thus analogous to the osculating parameters used in the theory of linear differential equations (e.g., Mathews and Walker, 1964).

[13]To smooth their frequency-domain measurements, Brown *et al.* (1996) used a 101-point moving average (evidently tailored to a frequency spacing between points of approximately 1.2 Hz) but fail to specify the shape of their smoothing function. If all points in the moving average were weighted equally (i.e., if the smoothing function were rectangular), the corresponding time-domain window, a sinc function, would have been nonmonotonic, oscillating about zero with a period of roughly 8.25 ms.

[14]We are reminded here of the dialectic described by Levins and Lewontin (1985): ''A necessary step in theoretical work is to make distinctions. But whenever we divide something into mutually exclusive and jointly all-encompassing categories, it turns out on further examination that these opposites interpenetrate.''

[15]To unmix P_{dp} into two components we used a window with a ''low-pass'' characteristic in the time domain. The technique is easily generalized to the unmixing of multiple components with different latencies (e.g., by using multiple ''bandpass'' windows centered at different latencies or a succession of ''low-pass'' windows with different cutoffs).

[16]Working with SFOAEs at low stimulus levels, Shera and Zweig (1993a) established a similar equivalence between SFOAEs extracted using the vector-subtraction method (Kemp and Chum, 1980) and the method of spectral smoothing. The vector-subtraction method exploits the nonlinear saturation of the SFOAE—or ''self-suppression'' of the traveling wave (e.g., Kanis and de Boer, 1993)—at higher stimulus levels.

[17]Because we varied our sampling rate between measurement points, corresponding stimulus durations varied by up to $\pm 3\%$.

Allen, J. B., and Lonsbury-Martin, B. L. (**1993**). ''Otoacoustic emissions,'' J. Acoust. Soc. Am. **93**, 568–569.

Allen, J. B., and Neely, S. T. (**1992**). ''Micromechanical models of the cochlea,'' Phys. Today **45**, 40–47.

Bogert, B. P., Healy, M. J. R., and Tukey, J. W. (**1963**). ''The frequency analysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking,'' *Proceedings of the Symposium on Time Series Analysis*, edited by M. Rosenblatt (Wiley, New York), pp. 209–243.

Brass, D., and Kemp, D. T. (**1991**). ''Time-domain observation of otoacoustic emissions during constant tone stimulation,'' J. Acoust. Soc. Am. **90**, 2415–2427.

Brass, D., and Kemp, D. T. (**1993**). ''Suppression of stimulus frequency otoacoustic emissions,'' J. Acoust. Soc. Am. **93**, 920–939.

Brown, A. M., and Beveridge, H. A. (**1997**). ''Two components of acoustic distortion: Differential effects of contralateral sound and aspirin,'' in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. L. Hecht-Poinar (World Scientific, Singapore), pp. 219–225.

Brown, A. M., Harris, F. P., and Beveridge, H. A. (**1996**). ''Two sources of acoustic distortion products from the human cochlea,'' J. Acoust. Soc. Am. **100**, 3260–3267.

Dreisbach, L. E., and Siegel, J. H. (**1999**). ''Level and phase relationships of distortion-product otoacoustic emission sources with varied primary frequency ratios in humans,'' Assoc. Res. Otolaryngol. Abs. **22**, 392.

Engdahl, B., and Kemp, D. T. (**1996**). ''The effect of noise exposure on the details of distortion product otoacoustic emissions in humans,'' J. Acoust. Soc. Am. **99**, 1573–1587.

Fahey, P. F., and Allen, J. B. (**1997**). ''Measurement of distortion product phase in the ear canal of the cat,'' J. Acoust. Soc. Am. **102**, 2880–2891.

Fahey, P. F., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**2000**). ''Nonlinear interactions that could explain distortion product interference response areas,'' J. Acoust. Soc. Am. **108**, 1786–1802.

Furst, M., Rabinowitz, W. M., and Zurek, P. M. (**1988**). ''Ear canal acoustic distortion at $2f_1-f_2$ from human ears: Relation to other emissions and perceived combinations tones,'' J. Acoust. Soc. Am. **84**, 215–221.

Gaskill, S. A., and Brown, A. M. (**1990**). ''The behavior of the acoustic distortion product, $2f_1-f_2$, from the human ear and its relation to auditory sensitivity,'' J. Acoust. Soc. Am. **88**, 821–839.

Guinan, J. J. (**1990**). ''Changes in stimulus frequency otoacoustic emissions produced by two-tone suppression and efferent stimulation in cats,'' in *Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer, New York), pp. 170–177.

Heitmann, J., Waldman, B., Schnitzler, H. U., Plinkert, P. K., and Zenner, H.-P. (**1998**). ''Suppression of distortion product otoacoustic emissions (DPOAE) near $2f_1-f_2$ removes DP-gram fine structure—Evidence for a secondary generator,'' J. Acoust. Soc. Am. **103**, 1527–1531.

Kalluri, R., and Shera, C. A. (**2000**). ''Are DPOAEs a mixture of emissions generated by different mechanisms?'' Assoc. Res. Otolaryngol. Abs. **23**, 480.

Kanis, L. J., and de Boer, E. (**1993**). ''Self-suppression in a locally active nonlinear model of the cochlea,'' J. Acoust. Soc. Am. **94**, 3199–3206.

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T. (**1997**). ''Otoacoustic emissions in perspective,'' in *Otoacoustic Emissions: Clinical Applications*, edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 1–21.

Kemp, D. T. (**1998**). ''Otoacoustic emissions: Distorted echoes of the cochlea's travelling wave,'' in *Otoacoustic Emissions: Basic Science and Clinical Applications*, edited by C. I. Berlin (Singular Group, San Diego).

Kemp, D. T., and Brown, A. M. (**1983**). ''An integrated view of cochlear mechanical nonlinearities observable from the ear canal,'' in *Mechanics of Hearing*, edited by E. de Boer and M. A. Viergever (Martinus Nijhoff, The Hague), pp. 75–82.

Kemp, D. T., and Chum, R. A. (**1980**). ''Observations on the generator mechanism of stimulus frequency acoustic emissions—Two tone suppres-

sion,'' in *Psychophysical Physiological and Behavioural Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft University Press, Delft), pp. 34–42.
Kemp, D. T., Brass, D., and Souter, M. (**1990**). ''Observations on simultaneous SFOAE and DPOAE generation and suppression,'' in *Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer, New York), pp. 202–209.
Kim, D. O. (**1980**). ''Cochlear mechanics: Implications of electrophysiological and acoustical observations,'' Hear. Res. **2**, 297–317.
Kimberley, B. P., Brown, D. K., and Eggermont, J. J. (**1993**). ''Measuring human cochlear traveling wave delay using distortion product emission phase responses,'' J. Acoust. Soc. Am. **94**, 1343–1350.
Knight, R., and Kemp, D. T. (**2000a**). ''Separation of 'wave' and 'place' fixed $2f_1-f_2$ DPOAE,'' Assoc. Res. Otolaryngol. Abs. **23**, 987.
Knight, R. D., and Kemp, D. T. (**2000b**). ''Indications of different distortion product otoacoustic emission mechanisms from a detailed f_1,f_2 area study,'' J. Acoust. Soc. Am. **107**, 457–473.
Kummer, P., Janssen, T., and Arnold, W. (**1998**). ''The level and growth behavior of the $2f_1-f_2$ distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss,'' J. Acoust. Soc. Am. **103**, 3431–3444.
Levins, R., and Lewontin, R. C. (**1985**). *The Dialectical Biologist* (Harvard University Press, Cambridge).
Mathews, J., and Walker, R. (**1964**). *Mathematical Methods of Physics* (Benjamin, New York).
Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (**1999a**). ''Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of the frequency ratio f_2/f_1,'' J. Acoust. Soc. Am. **106**, 3473–3483.
Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (**1999b**). ''Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss,'' J. Acoust. Soc. Am. **106**, 3484–3491.
Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). ''Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli,'' J. Acoust. Soc. Am. **83**, 652–656.
Norton, S. J., and Neely, S. T. (**1987**). ''Tone-burst-evoked otoacoustic emissions from normal-hearing subjects,'' J. Acoust. Soc. Am. **81**, 1860–1872.
O'Mahoney, C. F., and Kemp, D. T. (**1995**). ''Distortion product otoacoustic emission delay measurement in human ears,'' J. Acoust. Soc. Am. **97**, 3721–3735.
Papoulis, A. (**1962**). *The Fourier Integral and its Applications* (McGraw-Hill, New York).
Patuzzi, R. (**1996**). ''Cochlear micromechanics and macromechanics,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 186–257.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). ''A review of otoacoustic emissions,'' J. Acoust. Soc. Am. **89**, 2027–2067.
Ren, T., Nuttall, A. L., and Parthasarathi, A. A. (**2000**). ''Quantitative measure of multicomponents of otoacoustic emissions,'' J. Neurosci. Methods **96**, 97–104.
Shera, C. A., and Guinan, J. J. (**1999**). ''Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs,'' J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., and Guinan, J. J. (**2000a**). ''Frequency dependence of stimulus-frequency-emission phase: Implications for cochlear mechanics,'' in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, K. Ohyama, and T. Koike (World Scientific, Singapore), pp. 381–387.
Shera, C. A., and Guinan, J. J. (**2000b**). ''Reflection-emission phase: A test of coherent reflection filtering and a window on cochlear tuning,'' Assoc. Res. Otolaryngol. Abs. **23**, 545.
Shera, C. A., and Zweig, G. (**1993a**). ''Noninvasive measurement of the cochlear traveling-wave ratio,'' J. Acoust. Soc. Am. **93**, 3333–3352.
Shera, C. A., and Zweig, G. (**1993b**). ''Order from chaos: Resolving the paradox of periodicity in evoked otoacoustic emission,'' in *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 54–63.
Shera, C. A., Talmadge, C. L., and Tubis, A. (**2000**). ''Interrelations among distortion-product phase-gradient delays: Their connection to scaling symmetry and its breaking,'' J. Acoust. Soc. Am. **108**, 2933–2948.
Siegel, J. H., Dreisbach, L. E., Neely, S. T., and Spear, W. H. (**1998**). ''Vector decomposition of distortion-product otoacoustic emission sources in humans,'' Assoc. Res. Otolaryngol. Abs. **21**, 347.
Souter, M. (**1995**). ''Stimulus frequency otoacoustic emissions from guinea pig and human subjects,'' Hear. Res. **90**, 1–11.
Stover, L. J., Neely, S. T., and Gorga, M. P. (**1996**). ''Latency and multiple sources of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **99**, 1016–1024.
Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). ''Modeling otoacoustic emission and hearing threshold fine structures,'' J. Acoust. Soc. Am. **104**, 1517–1543.
Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). ''Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **105**, 275–292.
Ver, I. L., Brown, R. M., and Kiang, N. Y. S. (**1975**). ''Low-noise chambers for auditory research,'' J. Acoust. Soc. Am. **58**, 392–398.
Whitehead, M. L., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**1996**). ''Visualization of the onset of distortion-product otoacoustic emissions and measurement of their latency,'' J. Acoust. Soc. Am. **100**, 1663–1679.
Wilson, J. P. (**1980**). ''Evidence for a cochlear origin for acoustic re-emissions, threshold fine-structure and tonal tinnitus,'' Hear. Res. **2**, 233–252.
Withnell, R. H., and Yates, G. K. (**1998**). ''Enhancement of the transient-evoked otoacoustic emission produced by the addition of a pure tone in the guinea pig,'' J. Acoust. Soc. Am. **104**, 344–349.
Zweig, G., and Shera, C. A. (**1995**). ''The origin of periodicity in the spectrum of evoked otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.

Spontaneous otoacoustic emissions and relaxation dynamics of long decay time OAEs in audiometrically normal and impaired subjects

R. Sisto[a)]
Dipartimento Igiene del Lavoro, ISPESL, Via Fontana Candida, 1, 00040 Monte Porzio Catone (Roma), Italy

A. Moleti[b)]
Dipartimento di Fisica, Università di Roma ''Tor Vergata,'' Via della Ricerca Scientifica, 1, 00133 Roma, Italy

M. Lucertini[c)]
Italian Air Force—CVS Aerospace Medicine Department, Pratica di Mare AFB, 00040 Pomezia (Roma), Italy

(Received 23 June 2000; revised 27 October 2000; accepted 7 November 2000)

The relationship between hearing loss, detected by measuring the audiometric threshold shift, and the presence of long-lasting otoacoustic emissions, has been studied in a population of 66 adult males, by analyzing the cochlear response in the 80 ms following the subministration of a click stimulus. Most long-lasting OAEs are also recognizable as Synchronized Spontaneous OAEs (SSOAEs). The OAE characteristic decay times were evaluated according to the model by Sisto and Moleti [J. Acoust. Soc. Am. **106**, 1893 (1999)]. The starting hypothesis, confirmed by the results, is that long decay time and large equilibrium amplitude are both manifestations of the effectiveness of the active feedback mechanism. The prevalence and frequency distribution of long-lasting OAEs, and of their SSOAE subset, have been separately analyzed for normal and impaired ears. No long-lasting OAE was found within the hearing loss frequency range, but several were found in impaired ears outside the hearing loss range, both at lower and higher frequencies. This result suggests that the correlation between the presence of long-lasting OAEs and good cochlear functionality be local in the frequency domain. The monitor of the OAE decay time is proposed as a new possible method for early detecting hearing loss in populations exposed to noise. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1336502]

PACS numbers: 43.64.Jb, 43.64.Kc, 43.64.Wn [BLM]

I. INTRODUCTION

Spontaneous otoacoustic emissions (SOAEs) were predicted by Gold (1948) as instabilities in the active feedback mechanism, which allows the cochlea to perform quantum-limited measures in the proximity of the auditory threshold (Bialek and Wit, 1984). The experimental discoveries by Kemp (1978) opened the way to extensive observations of acoustic signals generated by the cochlea both in the presence and absence of external stimuli, evoked OAEs [Transiently Evoked OAEs (TEOAEs) and Distortion Product OAEs (DPOAEs)] and SOAEs, respectively (reviewed in Probst *et al.*, 1991). Properly named SOAEs are measured by recording the signal in the ear canal with no external stimulus, and averaging a large number of spectra. Synchronized Spontaneous Otoacoustic Emissions (SSOAEs) are measured by synchronously averaging in the time domain a large number of relatively long (typically 80 ms) data streams, each recorded after an identical click stimulus, and computing the spectrum of the average waveform. It is quite obvious that SOAE and SSOAE measuring techniques are not equivalent, because the SSOAE signal is the response to an external stimulus. However, as the observation time is typically much longer than the decay time of TEOAEs, the SSOAE spectrum is dominated by nonzero equilibrium amplitude OAEs (SOAEs). The study of OAEs has produced important advances in two main directions: the understanding of hearing physiology and the clinical detection of hearing impairment related to cochlear functionality. In the case of serious hearing loss, a correlation is observed with total absence of OAEs (Probst *et al.*, 1987). According to these findings, pass–fail tests based on the reproducibility of TEOAEs have been developed, and are currently applied to the neonatal hearing screening (Norton, 1993; White *et al.*, 1994). Recently, the Recurrence Quantification Analysis (RQA) technique (Eckmann, 1983; Zbilut and Webber, 1992) has been applied to the study of TEOAEs (Zimatore *et al.*, 2000), and RQA parameters have been correlated to TEOAE reproducibility.

In the case of moderate hearing loss, not sufficient to completely suppress the evoked cochlear response, much work has been done about the correlation between spectral characteristics of OAEs and audiometric threshold shift (Probst *et al.*, 1987). The audiometric threshold profile was correlated (Lucertini *et al.*, 1996; Kowalska and Sulkowski,

[a)]Electronic mail: sisto@igiene-ispesl.castelliromani.it
[b)]Electronic mail: moleti@roma2.infn.it
[c)]Electronic mail: csv.rmas@tin.it

1997; Vinck *et al.*, 1998), to the spectral distribution of TEOAEs. The affected audiometric frequencies [with hearing loss (HL) >25 dB] are correlated to low signal-to-noise ratio (snr) in the TEOAE spectrum. DPOAEs have also been correlated with the audiometric profile (Moulin *et al.*, 1994), as regards both the DPOAE spectral amplitude and the DPOAE threshold (minimum amplitude of the primary tones needed to evoke the DP). It should be reminded here that standard tonal audiometric techniques, which consist of a few monochromatic threshold measurements, are often affected by threshold microstructure. High resolution audiometry is in principle able to partially overcome the problem, but it is a very expensive and time consuming technique. All these observations suggest the possibility of developing objective tests, which could support standard audiometry, particularly for subjects who could likely simulate or dissimulate hearing damage.

A further promising application of OAEs, which will be discussed in this work, is the monitoring of cochlear functionality in populations exposed to noise or other ototoxic agents. The goal of this research is the early detection of functional alterations that would not be detectable by standard low resolution audiometric techniques. The variation of the TEOAE amplitude and of the DPOAE threshold as a consequence of exposure to noise (Attias and Bresloff, 1996; Engdahl *et al.*, 1996; Mansfield *et al.*, 1999) and to ototoxic drugs (Furst *et al.*, 1995) has been measured with high sensitivity also in subjects that would not show any variation of the audiometric threshold. However, for follow-up purposes in subjects at risk for cochlear hearing loss, both TEOAE and DPOAE changes associated to exposure should result larger than the standard test-retest variability. Unfortunately, relatively large test–retest variabilities have been reported in various studies for TEOAEs (Avan *et al.*, 1997; Harris *et al.*, 1991; Marshall and Heller, 1996) and DPOAEs (Roede *et al.*, 1993). Thus from a clinical point of view, a final diagnostic decision making process, when a slight change of OAE parameters is observed (Lucertini *et al.*, 1998; Arnold *et al.*, 1999), can result in quite a difficult task.

The cochlear sites associated with the SOAE frequencies seem to have the highest amplification by the outer hair cell feedback system. This idea is supported by the observation that SOAE frequencies correspond (within 10 Hz) to minima in the hearing threshold microstructure (Schloth, 1983). It has also been observed that the hearing threshold at SOAE frequencies is always less than 20 dB HL (Probst *et al.*, 1987). Moulin *et al.* (1991) reported that no SOAE was present in ears with hearing loss of more than 10 dB at 1 kHz. They concluded that the presence of a SOAE is an indication that no hearing loss higher than 10 dB at 1 kHz is present. McFadden and Mishra (1993) demonstrated a significant correlation between the average hearing threshold and the presence of a minimum number of SOAEs. In the present work it will be shown that an analogous relationship exists indeed, but that the correlation between SSOAE (and, more generally, long-lasting OAE) presence and cochlear functionality is local in the frequency domain, rather than global.

The same SOAE cochlear sites have also been shown to be vulnerable by ototoxic agents. There is large experimental evidence that SOAEs decrease in amplitude, or even disappear, as a consequence of exposure to noise (Norton *et al.*, 1989; Furst *et al.*, 1992) and of aspirin administration (Long and Tubis, 1988; Long *et al.*, 1991).

Although long-lasting OAEs, and among them, SSOAEs, are measurable only in a limited fraction of subjects (about 76% and 70%, respectively, in the present work), the evidence that the cochlear sites associated to these OAEs are correlated both to good cochlear functionality and to high sensitivity to ototoxic agents suggests to propose the measure of long-lasting OAE parameters as a method for early monitoring cochlear damage in exposed populations. This work focuses on the analysis of the time evolution of long-lasting OAEs, which dominate the long-time click-evoked cochlear response, with the goal of finding some measurable physical quantity to be used for monitoring cochlear functionality in exposed populations.

SSOAE data, due to their relatively long (80-ms) recording time, are suitable for analyzing the relaxation dynamics of the long decay time lines. According to the limit-cycle oscillator equation proposed by Sisto and Moleti (1999) for describing OAEs, the cochlear response to a click stimulus at the OAE frequencies is predicted to follow three phases: a short (a few ms) first phase of hyperbolic decay, a slower exponential decay phase, and, finally, a stationary phase at the limit-cycle equilibrium amplitude (the SOAE amplitude). The duration of the exponential phase and the SOAE amplitude, which are dependent on the oscillator parameters, determine which physical quantities are well measurable. If the SOAE level is high, or the damping coefficient is large, the exponential decay phase is often absent, or too short to allow a reliable measure of the characteristic decay time. If the SOAE level is below or near to the local noise level, the evaluation of the equilibrium amplitude results impossible or uncertain, while it may still be possible to get a good measure of the decay time. Thus long-lasting OAEs can have either a measurable equilibrium amplitude (in this case they belong to the SSOAE subset) or a measurable exponential decay characteristic time, or both. The SSOAE subset has been separately analyzed, because the existing literature reports the correlation between hearing functionality and SOAE only, so it is necessary to select this subset of the long-lasting OAEs, to allow for comparison with earlier studies.

The monitor of cochlear functionality could be based, in principle, either on the measure of the SOAE equilibrium amplitude or of the OAE decay time. Optimized data acquisition and analysis procedures would likely be different in the two cases. Both methods should be tested by evaluating on a statistically large database, for each parameter (amplitude and decay time), the sensitivity to cochlear damage and the test–retest fluctuations. The last are due to both systematic uncertainties and intrinsic physiological fluctuations. In principle, the monitor of the decay time would have the obvious advantage, with respect to the monitor of the SOAE absolute amplitude, of using differential measurements, which are not affected by the systematic uncertainties that are typical of the measure of absolute physical quantities.

This point could become particularly important when long-time monitoring of exposed populations is concerned, because, in such a case, measurements would likely be performed in nonidentical environmental conditions and by different operators. In this work, to preliminarily check the practical applicability of both methods, it has been evaluated what is the typical fraction of subjects showing at least one clearly detectable SOAE, and the typical fraction of subjects with at least one exponentially decaying OAE, whose decay time is accurately measurable. An ''ad hoc'' data analysis method has been developed to evaluate the decay time of OAEs, after administration of the click, in a population of 66 adult males. In Sec. II the population is described, and the basic data acquisition and analysis techniques used in this work are reviewed. In Sec. III the mathematical model used for describing the OAE response is briefly recalled. In Sec. IV the results are shown. A discussion of the results is presented in Sec. V.

II. METHOD

In this work SSOAE spectra and associated 80-ms waveforms recorded using the ILO-96 system (Otodynamics Ltd.), and pure-tone audiograms from both ears of 66 young adult (20–30 yrs.) males, have been analyzed. All subjects underwent an otoscopy with removal of debris and wax from the ear canal, and an impedance test, before being analyzed for OAEs, to rule out possible bias due to external and/or middle ear dysfunctions.

Pure-tone audiograms were recorded in an acoustically shielded room. The audiometric test frequencies were 0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz. The examined ear was conventionally defined ''normal'' if no threshold shift higher than 20 dB was found over the whole frequency range.

For 39 of the 132 ears the audiograms showed threshold shifts of more than 20 dB. These ears were defined ''impaired.'' For impaired ears, the limits of the hearing loss frequency range (f_{L1}, f_{L2}) were defined as the audiogram lowest and highest frequencies with a threshold shift of more than 20 dB.

SSOAE recordings were obtained using the standard ILO-96 system, which performs synchronous averages of a large number of data streams of duration $T=80$ ms, each recorded after an identical click stimulus, and computes the spectrum of the average wave form. It is well known that the spectral lines found by this technique (SSOAEs) are strongly correlated to the true SOAEs (Burns *et al.*, 1998; Sisto and Moleti, 1999), which are actually measured in absence of the external stimulus. This happens because, after presenting the click stimulus, the excitation at frequencies that do not correspond to SOAEs gives a negligible contribution to the spectrum, because these excitations decay on a timescale of a few ms, which is much shorter than T.

In this work, the prevalence and frequency distribution of SSOAEs have been studied, and, for impaired ears, the SSOAE frequencies have been correlated to the frequency range of the audiometrically detected hearing loss. The SSOAE spectra of the whole 80-ms recordings, with frequency resolution $\Delta f \sim 12$ Hz, were used to select and record frequency and amplitude of SSOAEs whose amplitude exceeded the local noise by more than 5 dB. This empirical SSOAE selection criterion was tested by an independent visual inspection of all the spectra by one of the authors, with several years of clinical experience in the visual analysis of SSOAE spectra. The empirical method yielded the same SSOAE list as the visual inspection method, with the exception of a few lines, which had excaped visual identification, and whose inclusion would negligibly affect the statistical meaning of the results. A total of 218 SSOAEs were found.

The SSOAE recordings were also processed off-line by a dedicated software developed in LabVIEW (National Instruments). The 80-ms waveform was divided in seven 50% overlapping time intervals of duration $T \sim 20$ ms. Each interval was Hanning windowed and FFT analyzed. This choice of the time intervals duration and spacing was made for maintaining a frequency resolution sufficient to separate most nearby OAEs, while getting a number of independent time intervals sufficient to get a precise determination of the linear damping coefficient also for relatively fast-decaying OAEs. The first time interval was not used because it is perturbed both by linear ringing and by the OAE delay, which is significant at low frequencies (Tognola *et al.*, 1997). The resulting six low resolution spectra allow for roughly estimating the characteristic decay time of the excited spectral lines in the 80 ms following the stimulus. As expected, a large number of spectral lines showed strong excitation by the stimulus and approximately exponential decay, sometimes approaching a constant equilibrium amplitude. These lines are sometimes called resonant TEOAEs in literature (Talmadge *et al.*, 1998). In this analysis, OAEs were conventionally defined ''long-lasting'' if their spectral line amplitude exceed local noise for at least 40 ms after the click stimulus subministration. This quite arbitrary (because it depends also on the noise level) selection criterion yielded a total number of 284 long-lasting OAEs. Most of them (218 out of 284, 77%) were also identified as SSOAEs, by the selection criterion discussed above, while, in the other cases, the corresponding SSOAE was not measurable in the spectrum above the local noise floor.

Statistical analysis of the fraction of subjects (and of ears) with at least one long-lasting OAE (and at least one SSOAE), and of the average number of long-lasting OAEs (SSOAEs) per subject (ear), in the normal and impaired subjects (ears) has been performed, to investigate if a correlation of these quantities to audiometric impairment was present. The correlation between hearing impairment and the presence of long-lasting OAEs (SSOAEs) has been estimated by the Pearson's χ^2 test. The test yields a number, $P_{n-i}(\chi^2)$, which estimates the probability for erroneously rejecting the null hypothesis assuming independence between hearing impairment and presence of long-lasting OAEs (SSOAEs). The correlation between hearing impairment and the distribution of the number of long-lasting OAEs (SSOAEs) per subject (ear) has been measured by the student's t-test. This test yields $P_{n-i}(t)$, which is the probability of getting a value of the student's t parameter larger than that corresponding to the measured distributions of the number of long-lasting OAEs (SSOAEs) in the two populations (normal and impaired), if the two data sets belong to the same probability

distribution. For both tests, a 95% confidence level ($P<0.05$) has been conventionally assumed to be statistically significant. The same statistical tests have also been applied to evaluate the differences between right and left ears.

III. MODEL

In this study, the time evolution of OAEs after a click stimulus has been analyzed. Two typical time evolutions were observed: (a) approximately constant amplitude in the 80 ms after the stimulus, and (b) approximately exponential decay, often approaching a constant equilibrium value. This observation supports previous results, obtained for neonates on a smaller data set, and the modeling of OAEs by a particular class of embedded nonlinear oscillators, proposed by Sisto and Moleti (1999). In that work it was demonstrated that the Van der Pol oscillator (Van der Pol, 1927), which is generally used in embedded oscillator models (Tubis and Talmadge, 1998), cannot explain the exponential decay regime, and the following oscillator was proposed:

$$\ddot{X}+\left(aX^2+b-\frac{c}{\langle X^2\rangle}\right)\dot{X}+\omega_0^2X=0. \tag{1}$$

The response to an external click stimulus, by using the multiple-scale analysis method (Bender and Orszag, 1978), is given by:

$$X(t)=2R(t)\sin(\omega_0 t+\theta_0), \tag{2}$$

with

$$R^2(t)=\frac{R_\infty^2+R_L^2(1+r_3)r_4e^{-b't}}{1-r_4e^{-b't}}, \tag{3}$$

where $b'=b+(2ac/b)$; $R_\infty^2\approx(c/2b)$; $R_L^2=(b/2a)$, and

$$r_1=\frac{R_\infty^2}{R_0^2};\quad r_2=\frac{R_L^2}{R_0^2};\quad r_3=\frac{R_\infty^2}{R_L^2};$$
$$r_4=\frac{1-r_1}{1+r_1+r_2};\quad \begin{cases} R_0=\dfrac{h_S\tau_S}{2\omega_0} \\ \theta_0\cong\dfrac{\omega_0\tau_s}{2}, \end{cases} \tag{4}$$

where h_s, τ_s, and ω_0 are, respectively, the stimulus amplitude, duration and angular frequency. The behavior of the oscillator described by Eq. (1), after presenting the click, consists of a short, strongly damped phase, followed by a phase with exponential decay, and, finally, the system reaches an equilibrium amplitude that is identified with the SOAE level. R_0, R_∞, and R_L are, respectively, the initial amplitude, the equilibrium amplitude, and the threshold amplitude for the exponential decay onset. Constant amplitude SSOAEs are also predicted by Eq. (1), for $ac/b^2\gg1$. In this case, the equilibrium SOAE amplitude is rapidly reached through the competition of the two nonlinear components of the second term of Eq. (1), without the onset of an intermediate exponential decay phase.

TABLE I. Global audiometric behavior of hearing impaired ears. For each audiometric band, the hearing loss mean, standard deviation, minimum and maximum values are listed.

f (kHz)	0.25	0.5	1	2	3	4	6	8
mean h. l.	15	15	17	20	33	42	38	29
st. dev.	2	4	7	14	17	20	21	21
min h. l.	10	5	0	10	10	5	10	10
max h. l.	20	30	40	70	60	75	80	90

As the spectral data used in this work have a 10-ms time resolution, the first strongly damped phase is not visible, because the amplitude reaches the value R_L in a time that is typically much shorter than 10 ms. After that, the first term starts becoming negligible, so the oscillator behavior is well described by:

$$\ddot{X}+b\left[1-\frac{c}{b\langle X^2\rangle}\right]\dot{X}+\omega_0^2X=0, \tag{5}$$

and the data can be fitted to solution of this simplified equation, which is:

$$R^2(t)=R_\infty^2+(R_0^2-R_\infty^2)e^{-bt}. \tag{6}$$

Here R_0 is just a phenomenological parameter, which is not directly connected to the stimulus intensity only, and which can be roughly identified with the amplitude R_L defined above for the complete Eq. (1).

As already mentioned, SSOAEs and the other long-lasting OAEs together dominate the click-evoked signal. It may be shown that the evoked signal is often well reproduced by a superposition of the responses of embedded oscillators corresponding to few OAEs only. These responses are analytically predictable by deterministic equations like Eqs. (1)–(3). Thus it is not surprising at all that a large degree of deterministic structuring was found, by using the RQA technique without any physical hypothesis (Zimatore *et al.*, 2000), in click-evoked signals, revealing the existence of an underlying ''autonomous'' dynamical system.

TABLE II. SSOAE statistics. A distinction is made between normal and impaired subjects. N is the number of subjects, $N_{S>1}$ the number of subjects with at least one SSOAE, and N_{SSOAE} the total number of SSOAEs detected in each subject category. Subjects with unilateral and bilateral hearing loss (UHL and BHL) are also separately analyzed.

Subjects	N	$N_{S>1}$	$N_{S>1}/N$ (%)	$P_{n-i}(\chi^2)$	N_{SSOAE}	N_{SSOAE}/N	$P_{n-i}(t)$
Normal ($n-n$)	36	26	72	n.s.	135	3.75	n.s.
Impaired	30	20	67		83	2.77	
UHL ($n-i$ or $i-n$)	21	13	62		58	2.76	
BHL ($i-i$)	9	7	78		25	2.78	
Total	66	46	70		218	3.30	

TABLE III. SSOAE statistics. Notation as in Table II. Here the distinction is made between normal and impaired ears and between right and left ears.

Ears	N	$N_{S>1}$	$N_{S>1}/N(\%)$	$P_{n-i}(\chi^2)$	N_{SSOAE}	N_{SSOAE}/N	$P_{n-i}(t)$
n	93	55	59	n.s.	170	1.83	n.s.
i	39	19	49		48	1.23	(0.088)
Total	132	74	56	$P_{R-L}(\chi^2)$	218	1.67	$P_{R-L}(t)$
R	66	41	62	n.s.	138	2.09	0.015
L	66	33	50		80	1.21	
R_n	47	29	62	n.s.	108	2.30	0.038
L_n	46	26	57		62	1.35	
R_i	19	12	63	n.s.	30	1.58	n.s.
L_i	20	7	35		18	0.90	

IV. RESULTS

Table I shows the global audiometric behavior of the hearing impaired ears analyzed in this work. The impaired ears were equally distributed between left and right (19 right ears and 20 left ears were found impaired). The same statistical analysis has been performed on all long-lasting OAEs and on the SSOAE subset. Results are separately shown for SSOAEs only, also because they can be directly compared to other studies. SSOAE statistics are summarized in Table II, separately for normal and impaired subjects, and in Table III, where a distinction is done between normal and impaired ears, and between right and left ears. Analogous results for all long-lasting OAEs (including SSOAEs) are in Tables IV and V.

As shown in Table II, there is no significant difference between normal and impaired subjects (unilateral and bilateral) as regards the fraction of subjects with at least one measurable SSOAE. The observed difference in the average number of SSOAE per subject, which is larger in normal subjects, is not statistically relevant, due to the large variance of the distributions. A significant fraction (30%, 20/66) of the examined subjects showed absence of any measurable SSOAE in both ears. Any monitoring technique based on SSOAEs would not be conceivable for such subjects. The measured SSOAE prevalence (70%) is similar to that found for SOAE in adults by other authors (Penner and Zhang, 1997; Burns *et al.*, 1992).

As shown in Table III, considering separately the ears of each subject, the fraction of ears with at least one measurable SSOAE is slightly larger for normal ears than for impaired ears, and for right ears than for left ears. Both differences are statistically not significant. A more evident difference exists in the same sense for the average number of SSOAE per ear, which is fully significant only for the $R-L$ comparison, yielding $P_{R-L}(t)=0.015$. Thus as regards SSOAEs only, the difference between normal and impaired ears is not statistically significant, and it is smaller than that due to the well-known asymmetry between right and left ears (Burns *et al.*, 1992).

By considering all long-lasting OAEs, the correlation is more significant. As shown in Table IV, the fraction of subjects with long-lasting OAEs is slightly larger for normal subjects but the difference is not statistically significant, while a significant difference is found in the average number of long-lasting OAEs per subject, with $P_{n-i}(t)=0.036$. As shown in Table V, considering separately the ears of each subject, the fraction of ears with at least one measurable long-lasting OAE is significantly larger for normal ears than for impaired ears, with $P_{n-i}(\chi^2)=0.021$. The average number of long-lasting OAEs per ear is even more significantly larger in normal ears $P_{n-i}(t)=0.006$. Right ears have more long-lasting OAEs than left ears, as a consequence of their generally larger OAE response, but, in this case, the differences between normal and impaired ears are more significant than those between the right and left ears.

Tables III and V also report the data separately for the right&normal, right&impaired, left&normal, and left&impaired population subsets. The $R-L$ and $n-i$ asymmetries are maintained in the comparison between the corresponding subsets, but the statistical significance obviously drops, due to the reduced samples.

Such statistical analyses may be defined by global analysis, since the ear is considered globally impaired if a threshold shift larger than 20 dB is measured at any of the audiometric frequencies. Summarizing, global statistical analyses, presented in Tables II–V show a weak correlation between hearing impairment and SSOAE presence, while a clearer correlation is obtained by including in the analysis the other

TABLE IV. Long-lasting OAE statistics. A distinction is made between normal and impaired (also separating unilateral from bilateral impairment) subjects. N is the number of subjects, $N_{L>1}$ the number of subjects with at least one long-lasting OAE, and N_{LOAE} the total number of long-lasting OAEs detected in each subject category.

Subjects	N	$N_{L>1}$	$N_{L>1}/N(\%)$	$P_{n-i}(\chi^2)$	N_{LOAE}	N_{LOAE}/N	$P_{n-i}(t)$
Normal ($n-n$)	36	29	81	n.s.	189	5.25	0.036
Impaired	30	21	70		95	3.17	
UHL ($n-i$ or $i-n$)	21	14	67		68	3.23	
BHL ($i-i$)	9	7	78		27	3.0	
Total	66	50	76		284	4.30	

TABLE V. Long-lasting OAE statistics. Notation as in Table IV. Here the distinction is made between normal and impaired ears and between right and left ears.

Ears	N	$N_{L>1}$	$N_{L>1}/N(\%)$	$P_{n-i}(\chi^2)$	N_{LOAE}	N_{LOAE}/N	$P_{n-i}(t)$
n	93	65	70	0.021	230	2.47	0.006
i	39	19	49		54	1.38	
Total	132	84	64	$P_{R-L}(\chi^2)$	284	2.15	$P_{R-L}(t)$
R	66	46	70	n.s.	175	2.65	0.017
L	66	38	58		109	1.65	
R_n	47	34	72	n.s.	141	3.00	0.045
L_n	46	31	67		89	1.93	
R_i	19	12	63	n.s.	34	1.79	n.s.
L_i	20	7	35		20	1.00	

long-lasting OAEs. The correlation increases by localizing the impairment definition from subjects to ears. It is interesting to note that SSOAEs were found in 59% (55/93) of the normal ears, but also in 49% (19/39) of the impaired ears. This last fraction could seem to be surprisingly high, if the relation between SOAE presence and good cochlear functionality had to be interpreted in a global sense. The comparison between the impaired ear audiograms and the SSOAE spectra has clarified this issue. No SSOAE was found within the hearing loss frequency range (f_{L1}, f_{L2}), but they were often found in impaired ears at frequencies both lower and higher than the hearing loss frequency range. An example is shown in Fig. 1, where the SSOAE spectrum is shown superimposed on the audiogram, for two impaired ears. In impaired ears, SSOAEs were found at frequencies either lower than $0.8f_{L1}$ or higher than $1.4f_{L2}$, as shown in Fig. 2, where the histogram is plotted of the ratio between the SSOAE frequency and the nearest audiometrically impaired frequency. In the same figure the same distribution is shown also for the whole set of long-lasting OAEs. A logarithmic scale is used in Fig. 2 to give a visual representation of the data, which is reminiscent of the position difference along the cochlear membrane between the OAE source and the impaired sites. In fact, according to Greenwood tonotopicity mapping (Greenwood, 1990), the logarithm of a frequency ratio is roughly proportional to a physical distance along the cochlea.

The small number of long-lasting OAEs at frequencies higher than the hearing loss range, shown in Fig. 2, is probably due to the frequency distribution of such OAEs. This distribution is shown in Fig. 3, separately for normal and impaired ears, for SSOAEs only and for the other long-lasting OAEs. As expected for adult subjects (Burns *et al.*, 1992), the distribution is peaked at low frequency, between 1 and 2 kHz, while the audiometric threshold shifts were typically found at higher frequencies (see Table I and Fig. 4). A global statistical comparison of the average number of long-lasting OAEs in impaired ranges to that in normal ranges would be biased by this asymmetry. Such a comparison, which is necessary to provide a quantitative statistical evaluation of the new result that is very clearly shown in Fig. 2,

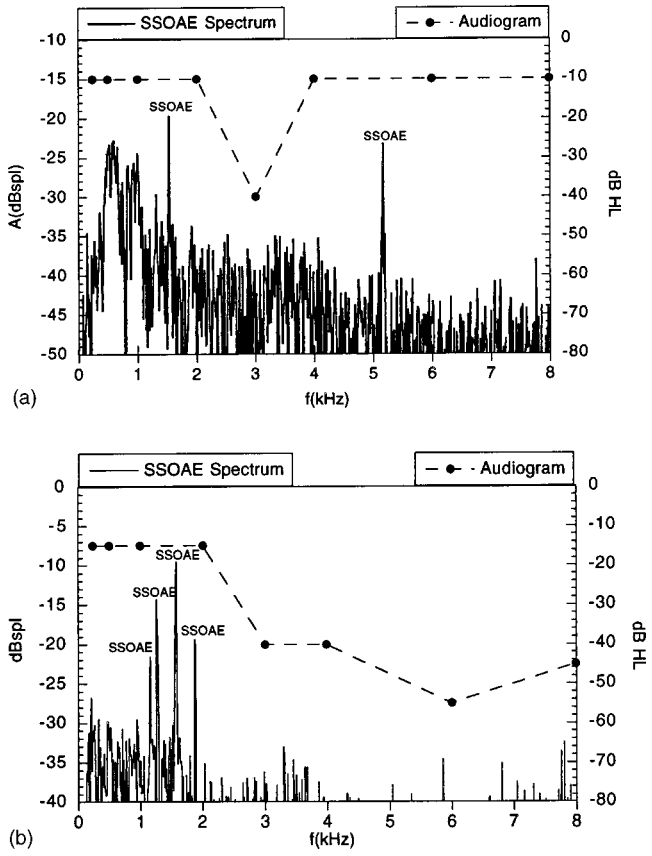


FIG. 1. Audiograms and SSOAE spectrum for two impaired ears, showing SSOAEs at frequencies both lower and higher than the impairment frequency range. SSOAE frequencies are indicated.

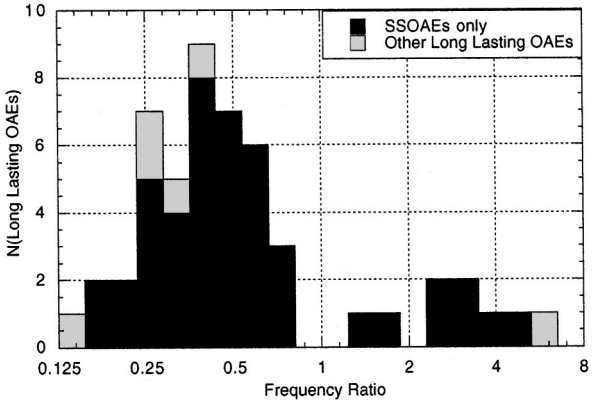


FIG. 2. Long-lasting OAE (and SSOAE only) relative frequency distribution in impaired ears. The OAE frequency is normalized to the nearest limit of the audiometric hearing loss frequency range ($f_n = f/f_{L1}$ if $f < f_{L1}$, $f_n = f/f_{L2}$ if $f > f_{L2}$).

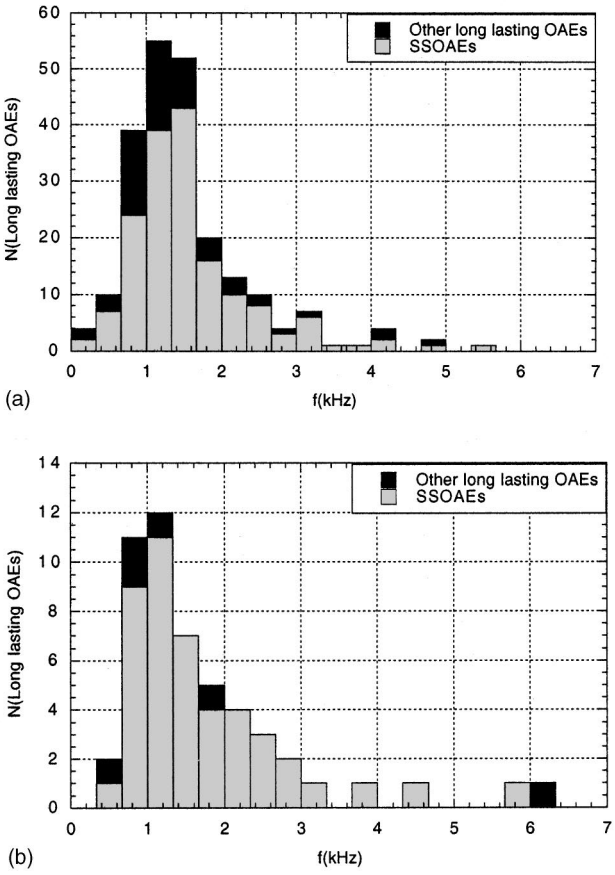


FIG. 3. Long-lasting OAE (and SSOAE only) absolute frequency distribution for normal (a) and impaired ears (b).

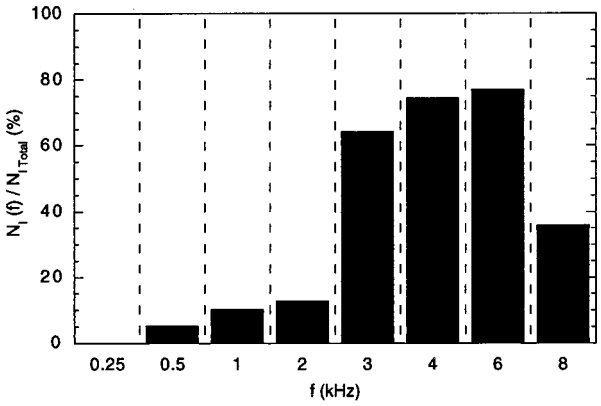


FIG. 4. Frequency distribution of the audiometric hearing loss in the examined impaired ears. The ratio between the number of ears with HL≥20 dE at a given audiometric frequency and the total number of impaired ears is reported.

must be made separately for each audiometric band. As already mentioned, audiometric data are not able to provide information about the fine scale structure of threshold shift. Nevertheless, just for statistical purposes, audiometric band limits can be conventionally assumed to lie halfway between successive audiometric frequencies. Then each band may be defined impaired or normal according to the audiometric threshold value at its central frequency. With such a definition of audiometric bands, it is possible to select the ears according to their being normal or impaired in a given audiometric band. The number of ears with at least one long-lasting OAE in a given band was different from zero only for the subset of ears that were normal in that band. The result is shown and statistically analyzed in Table VI. Despite having reduced the data sample size, the statistical significance is still high, as expected, in the central frequency bands, where both the fraction of locally impaired ears and the OAE frequency distribution are sufficiently large. One should also remind that the bands are independent, so the global result is quite interesting. For this reason, some values of $P_{n-i}(\chi^2)$ higher than the conventional 0.05 threshold have also been reported in parentheses.

In Fig. 5(a) the time evolution of the spectral amplitude after the click stimulus is shown for the long-lasting OAEs of one of the examined ears. Both exponentially decaying OAEs and constant amplitude SSOAEs are visible. In Fig. 5(b) are shown the OAE spectra of the same ear, in seven successive time intervals, with a 10 ms time step. The characteristic decay time of the spectral lines was computed by fitting the 80 ms OAE time evolution to Eq. (6), to which the above described cutting, windowing, and FFT procedure had been applied, as done for the real data.

The fit parameters R_0, b, and R_∞ were found for 176 OAEs, found in 69 of the examined ears. The characteristic decay time is given by $\tau=2/b$. The statistical distribution of the linear damping coefficients b is shown in Fig. 6, for the two ear populations, while in Fig. 7 the distribution of the quality factor $Q=f/\Delta f=2f/b$ is shown. As shown in Fig. 6, there is no significant difference between the typical decay

TABLE VI. Local correlation between hearing impairment and long-lasting OAE (and SSOAE only) presence in normal and impaired audiometric bands. $N_{S>1}$ and $N_{L>1}$ indicate the number of ears showing, respectively, at least one SSOAE and at least one long-lasting OAE, in the given frequency range.

f	N	$N_{S>1}$	$N_{L>1}$	N	$N_{S>1}$	$N_{L>1}$	N	$N_{S>1}$	$N_{L>1}$
	0.375–0.75 kHz			0.75–1.5 kHz			1.5–2.5 kHz		
$n(f)$	130	14	22	128	52	64	127	37	43
$i(f)$	2	0	0	4	0	0	5	0	0
$P_{n-i}(\chi^2)$	⋯	n.s.	n.s.	⋯	n.s. (0.10)	0.049	⋯	n.s. (0.15)	n.s. (0.11)
	2.5–3.5 kHz			3.5–5 kHz			5–7 kHz		
$n(f)$	107	14	15	103	6	9	102	3	4
$i(f)$	25	0	0	29	0	0	30	0	0
$P_{n-i}(\chi^2)$	⋯	0.056	0.047	⋯	n.s.	n.s. (0.10)	⋯	n.s.	n.s.

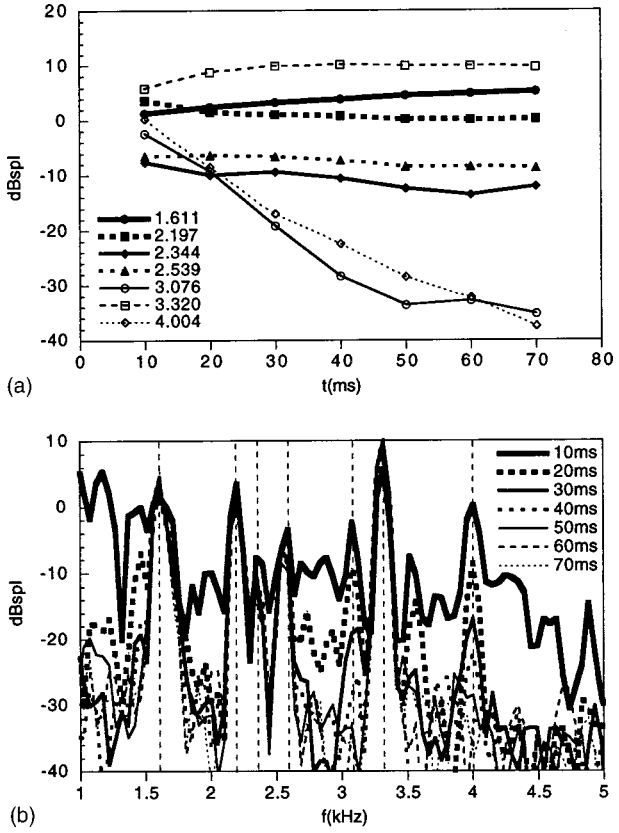


FIG. 5. Relaxation dynamics of long-lasting OAEs in the 80 ms following the stimulus presentation, in one of the examined ears. Spectral amplitude versus time for the selected OAEs (a), and spectra at different times after the click (b). Both exponentially decaying and constant amplitude OAEs are visible.

times of the OAEs observed in normal and impaired ears. This is not surprising, from a local point of view, because they all belong to audiometrically normal frequency bands.

The fraction of subjects showing at least one long-lasting OAE with a measurable decay time turned out to be 67%, which is comparable to the 70% of subjects showing at least one detectable SSOAE. These numbers are necessary to preliminarily evaluate the practical use of monitoring methods based on the measure of the OAE decay time or the SSOAE amplitude.

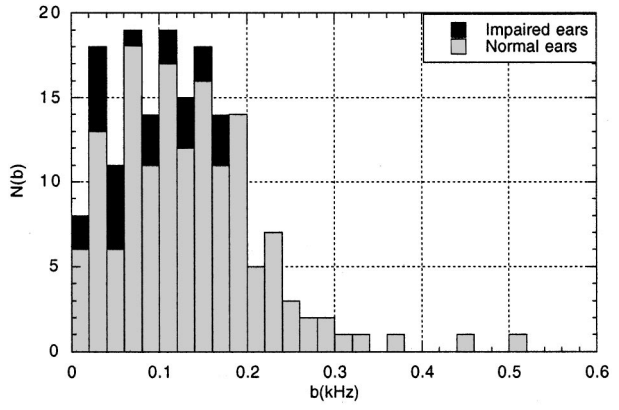


FIG. 6. Statistical distribution of the OAE linear damping parameter *b* for normal and impaired ears.

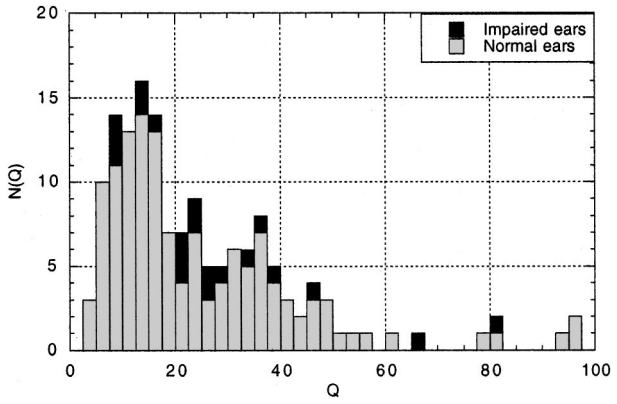


FIG. 7. Statistical distribution of the OAE quality factor *Q* for normal and impaired ears.

V. DISCUSSION

The very clear evidence that long-lasting OAEs exist also in impaired ears, but only outside the hearing impaired cochlear region, is a new important result. It suggests that the presence of a long-lasting OAE is an indicator of local good cochlear functionality and that the audiogram threshold shift is mainly due to local cochlear damage. This observation, which is limited by the coarse resolution of the audiometric bands, also suggests that the damage should poorly affect the overall cochlear transmission and amplification properties at other frequencies, which allow for OAEs, generated at other cochlear sites, to be produced and effectively transmitted to the outer ear. Figure 2 and Table VI present a large-scale frequency-sensitive confirmation of the general hypothesis that long-lasting OAEs are correlated to good cochlear functionality. This hypothesis, which was already confirmed on a smaller frequency scale by the observed correlation between hearing threshold fine structure minima and SOAE frequencies (Schloth, 1983; Furst *et al.*, 1992), is important in order to accept the general idea of monitoring the hearing functionality by measuring some characteristic parameters of long-lasting OAEs. The underlying hypothesis is that the cochlear membrane regions associated to such OAEs are more sensitive than the others, due to the local higher performances of the active feedback amplification system based on the outer hair cells motility. As the outer hair cells are easily damaged by ototoxic agents, such as noise, it is expected that the effect of noise should be early detectable by monitoring long-lasting OAE parameters.

On the other hand, the global correlation between hearing impairment and SSOAE presence was not statistically significant. This correlation has become significant by including in the analysis also the other long-lasting OAEs without a measurable equilibrium SOAE amplitude. The local correlation in separate audiometric bands was significant in the middle frequency range. All the above results confirm that a local relationship exists between cochlear damage and absence of long-lasting OAEs, but the correlation is perhaps not so tight to predict that the presence of these OAEs could provide a useful criterion for screening purposes. However, there is a good chance of effectively using the long-lasting OAE parameters to monitor cochlear functionality in ex-

posed subjects. Indeed, for monitoring purposes, it is not important to show that a difference exist between normal and impaired populations. It is important to show that, for a given subject, the disappearing, or a large parameter variation, of long-lasting OAEs actually be an early indicator of cochlear damage. The observation that in all the examined impaired subjects long-lasting OAEs were present only outside the impaired range is an encouraging, but rather indirect, confirmation of this hypothesis. It is still necessary to prove, in exposed subjects monitored on a long time, the sensitivity and specificity of long-lasting OAE parameter variation to hearing loss, also estimating the test-restest intrinsic fluctuations of these parameters. This study has just begun.

The observation of high-Q decaying OAEs is an interesting result itself for the modeling of cochlear mechanics. It is also interesting to comment on the observation that constant amplitude SSOAEs and exponentially decaying OAEs are found to coexist in the same ear, and in the same frequency range, as already observed in neonates (Sisto and Moleti, 1999). This evidence shows that very different dynamical properties are associated to nearby cochlear sites. Thus OAE dynamical properties are mainly determined, independently for each frequency, by the local cochlear parameters, rather than by the overall cochlear transmission. This observation is also supported by the fact that embedded oscillator models (Tubis and Talmadge, 1998) are very effective indeed for describing OAE phenomenology. The measured prevalence of OAEs with a measurable decay time (67% of subjects in the examined population) suggests that the proposed method for early detecting hearing loss by monitoring the OAE decay time could be effectively applied to a significant fraction of the population, with the advantage of using differential measurements, which are not affected by the systematic errors associated to any absolute amplitude measurement. It should also be stressed that the measuring and analysis techniques adopted in this work are still not fully optimized for decay time measurements. An optimized data acquisition system and analysis procedure (e.g., based on other time-frequency techniques) could be designed to apply the measurement of decaying OAE parameters to the largest possible fraction of the population, and to minimize the uncertainty on the decay time evaluation.

VI. CONCLUSIONS

In this work, the relationship between hearing loss, detected by measuring the audiometric threshold shift, and the presence of SSOAEs (and more generally, long-lasting OAEs) has been clarified, showing that such OAEs are associated to local (in the frequency domain) good cochlear functionality. No long-lasting OAE was found in the impaired range for all impaired ears, but they were often found at lower and/or higher frequencies. This result suggests that hearing loss at a given frequency be due to damage of the localized cochlear amplifier. The observation of large OAE response at frequencies apical to the audiometric loss range is particularly interesting, suggesting that the damage does not significantly affect the cochlear amplification and transmission for lower frequencies.

The comparison between audiometrically impaired and normal ears has shown no statistically significant difference between the two populations, as regards the fraction of ears with at least one measurable SSOAE, while the difference was significant for the whole set of long-lasting OAEs. The observation in the same ears of stationary and exponentially decaying OAEs, which can both be modeled by the oscillator of Eq. (1), is in agreement with the unitary interpretation of such OAEs as the response of a particular class of active non linear oscillators, proposed by Sisto and Moleti (1999).

A possible new method is suggested, for early detecting hearing loss in exposed populations, which is based on the measure of the characteristic decay time of exponentially decaying OAEs. Such decaying OAEs with a measurable decay time were found in a large fraction (67%) of the examined subjects, while SSOAEs were found in 70% of the subjects. The technique for estimating the decay time may be optimized, and the effectiveness of the method has still to be verified by testing its sensitivity and specificity on a large database. The proposed method could be used together with the monitoring of SSOAEs' absolute amplitude, and would have, in principle, the advantage of using differential measurements, which are not affected by the systematic errors associated to the measure of absolute quantities.

Arnold, D. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1999**). ''High-frequency hearing influences lower-frequency distortion-product otoacoustic emissions,'' Arch. Otolaryngol. Head Neck Surg. **125**, 215–222.

Attias, J., and Bresloff, I. (**1996**). ''Noise induced temporary otoacoustic emission shifts,'' J. Basin Clin. Physiol. Pharmacol. **7**, 221–233.

Avan, P., Elbez, M., and Bonfils, P. (**1997**). ''Click-evoked otoacoustic emissions and the influence of high-frequency hearing losses in humans,'' J. Acoust. Soc. Am. **101**, 2771–2777.

Bender, C. M., and Orszag, S. A. (**1978**). *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York).

Bialek, W. S., and Wit, H. P. (**1984**). ''Quantum limits to oscillator stability: Theory and experiments on otoacoustic emissions from the human ear,'' Phys. Lett. A **104**, 173–178.

Burns, E. M., Arehart, K. H., and Campbell, S. L. (**1992**). ''Prevalence of spontaneous otoacoustic emissions in neonates,'' J. Acoust. Soc. Am. **91**, 1571–1575.

Burns, E. M., Keefe, D. H., and Ling, R. (**1998**). ''Energy reflectance in the ear canal can exceed unity near spontaneous otoacoustic emission frequencies,'' J. Acoust. Soc. Am. **103**, 462–474.

Eckmann, J. P., Oliffson Kamphorst, S., and Ruelle, D. (**1987**). ''Recurrence plots of dynamical systems,'' Europhys. Lett. **4**, 973–977.

Engdahl, B., Woxen, O., Arnesen, A. R., and Mair, I. W. (**1996**). ''Transient evoked otoacoustic emissions as screening for hearing losses at the school for military training,'' Scand. Audiol. **25**, 71–78.

Furst, M., Reshef, I., and Attias, J. (**1992**). ''Manifestations of intense noise stimulation on spontaneous otoacoustic emission and threshold microstructure: Experiment and model,'' J. Acoust. Soc. Am. **91**, 1003–1014.

Gold, T. (**1948**). ''Hearing II. The physical basis of the action of the cochlea,'' Proc. R. Soc. B. **135**, 492–498.

Greenwood, D. D. (**1990**). ''A cochlear frequency position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Harris, F. P., Probst, R., and Wenger, R. (**1991**). ''Repeatability of transiently evoked otoacoustic emissions in normally hearing humans,'' Audiology **30**, 135–141.

Kemp, D. T. (**1978**). ''Stimulated otoacoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Kowalska, S., and Sulkowski, W. (**1997**). ''Measurements of click-evoked otoacoustic emissions in industrial workers with noise-induced hearing loss,'' Int. J. Occup. Med. Environ. Health **10**, 441–459.

Long, G. R., and Tubis, A. (**1988**). ''Modification of spontaneous and evoked otoacoustic emissions and associated psychoacoustic microstructure by aspirin consumption,'' J. Acoust. Soc. Am. **84**, 1343–1353.

Long, G. R., Tubis, A., and Jones, K. L. (**1991**). ''Modeling synchronization and suppression of spontaneous otoacoustic emissions using Van der Pol oscillators: Effects of aspirin administration,'' J. Acoust. Soc. Am. **89**, 1201–1212.

Lucertini, M., Bergamaschi, A., and Urbani, L. (**1996**). ''Transient evoked otoacoustic emissions in occupational medicine as an auditory screening test for employment,'' Br. J. Audiol. **30**, 79–88.

Lucertini, M., Tufarelli, D., and Urbani, L. (**1998**). ''Influence of a 6–8 kHz audiometric notch on transient evoked otoacoustic emissions,'' Eur. Arch. Otorhinolaryngol. **255**, 491–494.

Mansfield, J. D., Baghurst, P. A., and Newton, V. E. (**1999**). ''Otoacoustic emissions in 28 young adults exposed to amplified music,'' Br. J. Audiol. **33**, 211–222.

Marshall, L., and Heller, L. M. (**1996**). ''Reliability of transient-evoked otoacoustic emissions,'' Ear Hear. **17**, 237–254.

McFadden, D., and Mishra, R. (**1993**). ''On the relation between hearing sensitivity and otoacoustic emission,'' Hear. Res. **71**, 208–213.

Moulin, A., Collet, L., Delli, D., and Morgon, A. (**1991**). ''Spontaneous otoacoustic emissions and sensory-neural hearing loss,'' Acta Oto-Laryngol. **111**, 835–841.

Norton, S. J. (**1993**). ''Application of transient evoked otoacoustic emissions to pediatric populations,'' Ear Hear. **14**, 64–73.

Norton, S. J., Mott, J. B., and Champlin, C. A. (**1989**). ''Behavior of spontaneous otoacoustic emissions following intense ipsilateral acoustic stimulation,'' Hear. Res. **38**, 243–258.

Penner, M. J., and Zhang, T. (**1997**). ''Prevalence of spontaneous otoacoustic emissions in adults revisited,'' Hear. Res. **103**, 28–34.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). ''A review of otoacoustic emissions,'' J. Acoust. Soc. Am. **89**, 2027–2067.

Probst, R., Lonsbury-Martin, B. L., Martin, G. K., and Coats, A. C. (**1987**). ''Otoacoustic emissions in ears with hearing loss,'' Am. J. Otolaryngol. **8**, 73–81.

Roede, J., Harris, F. P., Probst, R., and Xu, L. (**1993**). ''Repeatability of distorsion product otoacoustic emissions in normally hearing humans,'' Audiology **32**, 273–281.

Schloth, E. (**1983**). ''Relation between spectral composition of spontaneous otoacoustic emissions and fine structure of threshold in quiet,'' Acustica **53**, 250–256.

Sisto, R., and Moleti, A. (**1999**). ''Modeling otoacoustic emissions by active non linear oscillators,'' J. Acoust. Soc. Am. **106**, 1893–1906.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). ''Modeling otoacoustic emission and hearing threshold fine structures,'' J. Acoust. Soc. Am. **104**, 1517–1543.

Tognola, G., Grandori, F., and Ravazzani, P. (**1997**). ''Time-frequency distributions of click-evoked otoacoustic emissions,'' Hear. Res. **106**, 112–122.

Tubis, A., and Talmadge, C. L. (**1998**). ''Ear canal reflectance in the presence of spontaneous otoacoustic emissions. I. Limit-cycle oscillator model,'' J. Acoust. Soc. Am. **103**, 454–461.

Van der Pol, B. (**1927**). ''Forced oscillations in a circuit with non-linear resistance,'' Philos. Mag. **3**, 65–80.

Vinck, B. M., Van Cauwenberge, P. B., Corthals, P., and De Vel, E. (**1998**). ''Multi-variant analysis of otoacoustic emissions and estimation of hearing thresholds: Transient evoked otoacoustic emissions,'' Audiology **37**, 315–334.

White, K. R., Vohr, B. R., Maxon, A. B., Behrens, T. R., McPherson, M. G., and Mauk, G. W. (**1994**). ''Screening all newborns for hearing loss using transient evoked otoacoustic emissions,'' Int. J. Pediatr. Otorhinolaryngol. **29**, 203–217.

Zbilut, J. P., and Webber, Jr., C. L. (**1992**). ''Embedding and delays as derived from quantification of recurrence plots,'' Phys. Lett. A **171**, 199–203.

Zimatore, G., Giuliani, A., Parlapiano, C., Grisanti, G., and Colosimo, A. (**2000**). ''Revealing deterministic structures in click-evoked otoacoustic emissions,'' J. Appl. Physiol. **88**, 1431–1437.

A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression

Xuedong Zhang
Hearing Research Center and Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

Michael G. Heinz
Hearing Research Center and Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215 and Speech and Hearing Sciences Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Ian C. Bruce
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21205

Laurel H. Carney[a)]
Hearing Research Center and Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

(Received 20 June 2000; accepted for publication 8 November 2000)

A phenomenological model was developed to describe responses of high-spontaneous-rate auditory-nerve (AN) fibers, including several nonlinear response properties. Level-dependent gain (compression), bandwidth, and phase properties were implemented with a control path that varied the gain and bandwidth of tuning in the signal-path filter. By making the bandwidth of the control path broad with respect to the signal path, the wide frequency range of two-tone suppression was included. By making the control-path filter level dependent and tuned to a frequency slightly higher than the signal-path filter, other properties of two-tone suppression were also included. These properties included the asymmetrical growth of suppression above and below the characteristic frequency and the frequency offset of the suppression tuning curve with respect to the excitatory tuning curve. The implementation of this model represents a relatively simple phenomenological description of a single mechanism that underlies several important nonlinear response properties of AN fibers. The model provides a tool for studying the roles of these nonlinearities in the encoding of simple and complex sounds in the responses of populations of AN fibers. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1336503]

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

I. INTRODUCTION

Phenomenological models for auditory-nerve (AN) responses provide a useful tool for studying the representation of simple and complex sounds at the first level of neural coding in the auditory system. These models allow hypothesis testing of the mechanisms that underlie various response properties. They also provide a tool for creating population responses that can be used to quantify the information available to the central nervous system (CNS) for different stimuli. In this study, a phenomenological model for AN responses was developed that focuses on several nonlinear response properties of AN fibers. The motivation for the development of this model was to provide a more accurate and quantitative description of the responses of AN fibers to complex sounds, such as noise-masked stimuli and speech sounds. To study the encoding of complex stimuli, inclusion of nonlinear interactions between frequency components in the stimulus is important.

Nonlinearities that are the focus of this study include the compressive changes in gain and bandwidth as a function of stimulus level, the associated changes in the phase of phase-locked responses, and two-tone suppression. These phenomena have all been related to a single mechanism in the inner ear, often referred to as the cochlear amplifier (Patuzzi and Robertson, 1988; Patuzzi, 1996; Holley, 1996). The challenge of the present study was to develop a model with a single mechanism that produces these different response properties. The development of the model was guided by the data available in the literature, where possible. However, physiological descriptions of nonlinear response properties typically focus on one or two properties at a time, and have been conducted in a number of different species. Therefore, the goal of the present study was to develop a model that captures the key features of several AN nonlinearities, while keeping the model as simple as possible.

Evidence for nonlinear gain in the inner ear was first described in terms of a compressive nonlinearity, or reduction in gain as stimulus level was increased to relatively high levels (Rhode, 1971). More recent studies have demonstrated that the compressive nonlinearity affects responses from as

[a)]Address for correspondence: Laurel H. Carney, PhD, Department of Biomedical Engineering, 44 Cummington St., Boston, MA 02215. Electronic mail: carney@bu.edu

low as 20 dB SPL up to the highest levels tested (110 dB SPL) in the most sensitive ears (Ruggero *et al.*, 1997). Ruggero *et al.* (1992) and Cooper and Rhode (1996) showed that the compressive nonlinearity and two-tone suppression are both affected by the same experimental manipulations, providing evidence that these two nonlinear properties are likely to be due to a single mechanism. Two-tone suppression has previously been described in a number of studies of the AN (e.g., Sachs and Kiang, 1968; Costalupes *et al.*, 1987; Javel *et al.*, 1978, 1983; Delgutte, 1990; Temchin *et al.*, 1997), inner hair cells (IHCs) (Cheatham and Dallos, 1989, 1990, 1992; Nuttall and Dolan, 1993), and basilar membrane (BM) (e.g., Ruggero *et al.*, 1992; Nuttall and Dolan, 1993; Rhode and Cooper, 1993, 1996; Cooper, 1996). Two-tone suppression grows with suppressor tone level at different rates depending upon the frequency of the suppressor with respect to the characteristic frequency (Costalupes *et al.*, 1987; Javel *et al.*, 1983; Delgutte, 1990; Ruggero *et al.*, 1992; Rhode and Cooper, 1996). This property of two-tone suppression will influence the responses of AN fibers to complex wide-band sounds. The present model includes this asymmetrical aspect of two-tone suppression as well as the compressive nonlinearity associated with the cochlear amplifier.

The model described here significantly extends a previous model developed by Carney (1993), which included compression and level-dependent bandwidths and phases, but not realistic two-tone suppression. The wide-band, feed-forward control path in the present model replaces the feedback control mechanism used in the previous model and is critical for including two-tone suppression. Wide-band suppression mechanisms superimposed on the more narrowly tuned excitatory process have previously been suggested as explaining some of the properties of two-tone suppression (Geisler and Sinex, 1980; Delgutte, 1990).

The model proposed here joins several other phenomenological models of basilar membrane and/or AN responses. This model focuses on the nonlinear response properties of auditory-nerve fibers described above, especially the level-dependent phase properties and two-tone suppression. Level-dependent phase properties of AN responses have not been a focus of other modeling studies, but several have addressed the problem of two-tone suppression and related nonlinear response properties.

Several modeling studies have explored combinations of linear filters and memoryless nonlinearities that provide phenomenological representations of responses at the level of the basilar membrane (without inclusion of models for the inner hair cells and IHC-AN synapse). Initial models of this type included the bandpass nonlinearity (BPNL) models of Pfeiffer (1970) and Duifhuis (1976). Goldstein (1990, 1995) extended this approach with a multiple bandpass nonlinear (MBPNL) model, which included two interacting paths, one with a low-pass filter followed by a memoryless nonlinearity, and one with a bandpass filter. Several nonlinear cochlear response properties can be explained by this model due to the level-dependent interaction of the two paths, which can be thought of as representing multiple modes of BM excitation (e.g., Lin and Guinan, 2000).

The MBPNL model describes level-dependent iso-reponse tuning curves that include ''tails,'' as well as several aspects of two-tone rate suppression, including the asymmetry of suppression for low-side versus high-side suppressors, synchrony capture by low-frequency tones, simple-tone interference, and the generation of combination tones (Goldstein, 1990, 1995; Lin and Goldstein, 1995). While there is some overlap between the phenomena described by the present model and by the MBPNL model, there are several important conceptual differences between the two modeling approaches. The MBPNL model consists of parallel pathways comprised of static filters that interact through memoryless nonlinearities, whereas the model presented here consists of a simple bandpass filter with time-varying gain and bandwidth. In addition, most of the simulations presented here include models for the IHC and the IHC-AN synapse; the goal of this model is to provide AN discharge patterns (as opposed to cochlear responses), and comparisons of model output are primarily made to AN response properties described in the literature. The detailed relationship between phases of BM and AN responses, which vary with stimulus frequency, characteristic frequency (CF, the frequency to which an AN fiber is most sensitive), and SPL (e.g., Ruggero and Rich, 1987; Narayan *et al.*, 1998; Cheatham and Dallos, 1999) are beyond the scope of this study.

The level-dependence of the phase of BM (e.g., Geisler and Rhode, 1982; Ruggero *et al.*, 1997), IHC (Cheatham and Dallos, 1998), and AN (e.g., Anderson *et al.*, 1971) responses to tones at frequencies above and below CF is a nonlinear response property that is potentially important for the encoding of complex sounds (e.g., Carney, 1994). This property cannot be described by the MBPNL model. The study presented here examines the relationship between level-dependent phase and two-tone suppression and illustrates that a single mechanism can be used to incorporate both properties in a phenomenological model. Note that the model presented here does not attempt to include all of the properties explained by the MBPNL and other models. In particular, tails of tuning curves and simple-tone interference, which can be explained by MBPNL models, are not addressed by this model. Also, this model does not include asymmetrical filter shapes, which are the focus of Irino and Patterson's (1997) auditory filter model that includes a level-dependent ''glide'' of the instantaneous frequency of AN impulse responses as a function of time. Actual BM and AN responses have glides that are level-independent (e.g., Recio *et al.*, 1996; de Boer and Nuttall, 1997; Carney *et al.*, 1999), and the direction of the glide varies with CF in AN responses (Carney *et al.*, 1999). Inclusion of this potentially important temporal response property into an AN model would be of interest for future studies.

Another feature of several AN models is the inclusion of high, medium, and low spontaneous-rate AN fibers (Liberman, 1978). Sachs and Abbas (1974) and Schoonhoven *et al.* (1998) investigated phenomenological models that explained rate-level functions for AN fibers with different spontaneous rates in terms of the relation between AN threshold and BM compression. Also, detailed models of the IHC-AN synapse (e.g., Schwid and Geisler, 1982; Meddis, 1986, 1988; Westerman and Smith, 1988; Geisler, 1990; Hewitt and Meddis,

1991; Lopez-Poveda *et al.*, 1998) provide descriptions of several features of AN rate-level functions. In the study presented here, results are limited to high spontaneous-rate AN fibers.

One benefit of the filterbank model developed here is that it allows the simulation of population responses of AN fibers to simple and complex sounds. Previous models have also been developed for this purpose, and these models share some of the properties of the present model. Jenison *et al.* (1991) developed a composite model for AN responses that included level-dependent peripheral filter bandwidths, which were based on filters derived from a database of AN rate-level functions. Deng and Geisler (1987) developed a composite model for AN responses based on a nonlinear cochlear model with longitudinal stiffness coupling. Giguere and Woodland (1994) proposed an analog/digital composite model that included the compressive nonlinearity. All of these models were tested primarily with speech stimuli, and showed several interesting features, such as synchrony capture by low-frequency formants. None of these models were tested closely using simple tones or pairs of tones to explore the details of their level-dependent phase properties or details of two-tone suppression.

The model described in the present study shares some general features with a recent model proposed by Robert and Eriksson (1999), which included nonlinear gain, bandwidth, and some aspects of two-tone suppression. However, the Robert and Eriksson (1999) model did not address several key response properties that are a focus of the present study. For example, their study did not include the temporal response properties of AN fibers, such as the dependence of synchrony on level and frequency, and the level-dependence of the phase of phase-locked responses. In addition, they did not address the asymmetry in suppression growth above and below CF. Their model involved a feedback control mechanism that combined control signals from neighboring fibers with different CFs to achieve an effectively wider-band control path, and thus wide-band two-tone suppression.

In the present study, the strategy is instead to use a wide-band feedforward path, which allows the properties of two-tone suppression to be included in a model of a single CF fiber (without having to simulate responses of the neighboring fibers). Furthermore, the level-dependent gain and bandwidth of the feedforward control path in the present model allow the asymmetry of two-tone suppression to be included. Because studies of information coding in the auditory system are still investigating the roles of temporal information and/or average discharge rate, this model was designed to simulate both aspects of the AN discharges with as much accuracy as possible, over wide ranges of CF and SPL.

This report presents the overall design and implementation of the model and shows responses of the model to a number of stimuli that have been used in physiological studies of AN fibers. The next section describes each stage of the model and its parameter values (either in equations or in a table), justification for parameter choices, and an explanation of the major effects of each parameter. Following the model description, response properties of the model are shown and discussed. Model responses are compared to several examples of AN responses from the literature; note that the parameters of the model were the same for all simulations. However, the levels of the stimuli tested were sometimes adjusted to accommodate differences in threshold between the model and a particular AN fiber. The model parameters are primarily based on the responses of AN fibers in cat; however, data from other species were used when necessary. The model presented here is focused on nonlinear tuning properties and is limited to high-spontaneous-rate AN fibers. Interactions between nonlinear aspects of basilar-membrane tuning and the properties of the IHC-AN synapse create different response properties for low- and medium-spontaneous-rate AN fibers (Sachs and Abbas, 1974). Future study in this series will focus on inclusion of more detail in the IHC-AN synapse, as well as the interaction of the properties of the synapse with the other nonlinear features of the present model, and will thus extend this model to include the other spontaneous-rate groups.

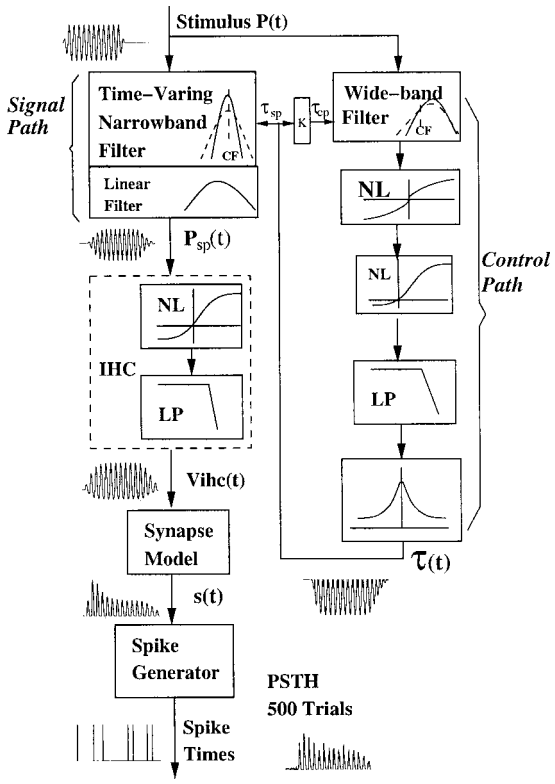


FIG. 1. Block diagram of the AN model. The waveforms that illustrate the output of each stage are the responses of a 500 Hz CF fiber to a 50 dB SPL pure tone at CF.

II. MODEL DESCRIPTION

A. Overview

The general scheme of the AN model implementation[1] is illustrated in Fig. 1. The input to the model is the instantaneous pressure waveform of the stimulus in Pascals. The effects of the external and middle ears are not considered here. The model includes properties described in recent physiological studies of the auditory system; however, it is a phenomenological model and the main effort is to simulate realistic level-dependent average-rate and temporal re-

sponses of AN fibers with the simplest possible model. To determine the model parameters, experimental data available in the literature were used whenever possible and reasonable assumptions were made when there were no data to support a particular parameter value. The selection of parameters and equations in the model was guided mainly by the response properties of the model rather than by the mechanisms of the actual physiological system. The two major parts of the nonlinear filtering section of the model are the *signal path* and the feedforward *control path* (Fig. 1).

The tuning of the signal path (corresponding roughly to tuning on the basilar membrane) is modeled by a cascade of a time-varying filter and a linear filter. The control path acts to regulate the tuning of the time-varying signal-path filter and is responsible for the compression and suppression effects observed in model AN responses. The bandpass filter in the control path has a broader bandwidth than the signal path to achieve two-tone suppression over a wide frequency range. A saturating nonlinearity followed by a low-pass filter in the control path determines the dynamic range and the dynamics of the compression and suppression. The control signal is then shifted and scaled to adjust the threshold and range of compression (each fiber is scaled based on its CF). The output of the signal-path filter is then passed through models for the IHC and IHC-AN synapse that represent corresponding processing stages in the cochlea. A nonhomogeneous Poisson-process model with refractory effects is used to generate the discharge times of the AN fiber. Brief descriptions and values for all model parameters are provided in Table I.

B. Signal-path filter

The signal-path filter represents the tuning properties of a specific location on the basilar membrane; the output of the signal-path filter provides the input to the IHC model. The signal path consists of a time-varying filter followed by a linear filter. The gain and bandwidth of the time-varying narrow-band filter are changed continuously as the control signal fluctuates, varying on a cycle-by-cycle basis with stimulus fluctuations below 800 Hz (limited by the low-pass filter in the control path). The signal-path time-varying filter is the source of the level-dependent-phase and the two-tone-suppression response nonlinearities illustrated in the figures below. The nonlinear filter also introduces asymmetry in the output signal, resulting in a dc component that varies across stimuli and sound levels. This dc component (which may or may not be biophysically appropriate) is difficult to accommodate using the simplified model stages that follow the nonlinear signal-path filter. Therefore, the final stage in the signal-path filter is a linear bandpass filter that eliminates the dc component of the response.

Both the time-varying nonlinear filter and the linear filter in the signal path were based on gammatone filters, which have been used in several studies to represent the impulse responses of AN fibers (Johannesma, 1972; de Boer, 1975; de Boer and de Jongh, 1978; de Boer and Kruidenier, 1990; Carney and Yin, 1988; Carney, 1993). The impulse response of the gamma-tone filter is given by

$$g(t) = u(t-\alpha)(t-\alpha)^{\gamma-1} e^{-(t-\alpha)/\tau} \cos[\omega_{\mathrm{CF}}(t-\alpha)] \quad (1)$$

where $u(\cdot)$ is the unit-step function, α is a delay added to the gammatone response, τ is the time constant, ω_{CF} is the radian frequency corresponding to the characteristic frequency (CF) of the model fiber, and γ is the order of the filter. This function has a simple expression in the frequency domain (Patterson *et al.*, 1988):

$$\begin{aligned} G(\omega) &= 0.5\tau^{\gamma}(\gamma-1)!\, e^{-j\omega\alpha}\left[\frac{1}{[1+j\tau(\omega-\omega_{\mathrm{CF}})]^{\gamma}}\right. \\ &\quad \left. + \frac{1}{[1+j\tau(\omega+\omega_{\mathrm{CF}})]^{\gamma}}\right] \\ &\cong \frac{\tau^{\gamma}(\gamma-1)!\, e^{-j\omega\alpha}}{2[1+j\tau\cdot(\omega-\omega_{\mathrm{CF}})]^{\gamma}}, \quad \omega_{\mathrm{CF}} \gg 0, \end{aligned} \quad (2)$$

where j is $\sqrt{-1}$.

The signal path consists of a nonlinear third-order gammatone filter followed by a linear first-order gammatone filter. The nonlinear filter is implemented by frequency shifting the input signal downward by CF, then using a cascade of three first-order low-pass filters, based on the strategy of Patterson *et al.* (1988). The low-pass filters were implemented digitally using the IIR bilinear transformation (Oppenheim and Schafer, 1975). The time delay α in the nonlinear filter is the additional delay that is required for a gammatone filter to represent the AN impulse response, including traveling-wave, acoustical, and synaptic delays (Carney and Yin, 1988; Shera and Guinan, 2000). The delay is a function of CF, estimated from fits of gammatone functions to measured reverse-correlation functions (Carney and Yin, 1988; Carney, 1993):

$$\alpha(\mathrm{CF}) = A_D e^{-x_{\mathrm{CF}}/A_L} - 2\pi/\omega_{\mathrm{CF}}, \quad (3)$$

where A_D and A_L are from Carney (1993) and x_{CF} is the distance (mm) from the apex of the basilar membrane from Liberman's (1982) frequency map for cat.

From Eq. (2) it is clear that both the gain and bandwidth of the filter are controlled by the time-varying time constant $\tau(t)$. The output of the control path specifies the time constant $\tau_{sp}(t)$ for each of the three first-order gammatone filters in the cascade that comprises the time-varying third-order filter in the signal path. The time constant $\tau_{sp}(t)$ varies over a range determined by τ_{narrow} (for sharp tuning at low SPL) and τ_{wide} (for broad tuning at high SPL), where τ_{narrow} is greater than τ_{wide}. The time-invariant time constant for the first-order linear gammatone filter in the signal path is set to τ_{wide} and the gain of this filter is set to 0 dB at CF. The values of τ_{narrow} and τ_{wide} are determined by the tuning properties of AN fibers. A linear fit of measured values of Q_{10} (the ratio between CF and bandwidth measured 10 dB above the fiber threshold) for cat AN fibers (Miller *et al.*, 1997) determines the value of τ_{narrow}:

$$\tau_{\mathrm{narrow}} = \frac{2Q_{10}}{2\pi\mathrm{CF}}, \quad (4)$$

where the $Q10$ data is fit by

$$\log_{10}(Q_{10}) = 0.4708\log_{10}(\mathrm{CF}/1000) + 0.4664. \quad (5)$$

TABLE I. Description of the parameters used in the AN model. The desired values of PST histogram characteristics are used to derive parameter values for Westerman and Smith's (1988) three-store diffusion model (see the Appendix). The resulting values of these characteristics for the model response are not the same as these input parameter values due to the effects of refractoriness. For example, the spontaneous rate of the model fiber response is approximately 45 spikes/s, rather than the 50 spikes/s indicated in the table below (see Fig. 7).

Parameters	Description	Values
Basilar membrane tuning filter and control path		
ω_{CF}	characteristic frequency of the fiber (rad/s)	
α	delay of the onset tone responses for cat (s)	See Eq. (3)
A_D	coefficient for traveling wave delay (ms)	8.13
A_L	length constant for traveling wave delay (nm)	6.49
x_{CF}	distance from apex of basilar membrane (mm)	
$\tau(t)$	output of the control path	
τ_{narrow}	estimated time constant at low sound level	See Eq. (4)
τ_{wide}	estimated time constant at high sound level	See Eq. (6)
γ_{cp}	order of the wide bandpass filter in control path	3
ω_{cp}	center frequency of the wide bandpass filter	1.2 mm basal to fiber CF
K	ratio of time constant in control path to that in signal path	$0.2+0.8\ \tau_{\mathrm{wide}}/\tau_{\mathrm{narrow}}$
A_{cp}	parameter in logarithmic nonlinearity	970
B_{cp}	parameter in logarithmic nonlinearity	2.75
C_{cp}	parameter in logarithmic nonlinearity	0.69
$x0_{cp}$	parameter in Boltzman function	7.6
$s0_{cp}$	parameter in Boltzman function	12
$x1_{cp}$	parameter in Boltzman function	5
$s1_{cp}$	parameter in Boltzman function	5
shift_{cp}	parameter in Boltzman function	0.125
cut_{cp}	cutoff frequency of control-path low-pass filter (Hz)	800
k_{cp}	order of control-path low-pass filter	3
dc	estimated dc shift of CP low-pass filter output at high level	0.37
R_0	ratio of τ_{LB} (lower bound of τ_{SP}) to τ_{narrow} [see Eq. (13)]	0.05
Inner hair cell model		
A_{ihc0}	scalar in IHC nonlinear function [see Eq. (16)]	0.1
B_{ihc}	parameter in IHC nonlinear function [see Eq. (15)]	2000
C_{ihc}	parameter in IHC nonlinear function [see Eq. (16)]	1.74
D_{ihc}	parameter in IHC nonlinear function [see Eq. (16)]	6.87*e*-9
cut_{ihc}	cutoff frequency of IHC low-pass filter (Hz)	3800
k_{ihc}	order of IHC low-pass filter	7
p_1	parameter in V_{ihc} rectifying function	0.0143
p_2	parameter in V_{ihc} rectifying function	See Eq. (18)
Synapse		
spont	spontaneous rate of fiber (spikes/s)	50
A_{SS}	steady state rate (spikes/s)	350
τ_{ST}	short-term time constant (ms)	60
τ_R	rapid time constant (ms)	2
$A_{R/ST}$	rapid response amplitude to short-term response amplitude ratio	6
PTS	peak to steady state ratio	8.6
$P_{I\ \max}$	permeability at high sound level	0.6
Spike generator and refractoriness		
$c0$	parameter for relative refractoriness	0.5
$c1$	parameter for relative refractoriness	0.5
$s0$	parameter for relative refractoriness (ms)	1.0
$s1$	parameter for relative refractoriness (ms)	12.5
R_A	absolute refractory period (ms)	0.75

While τ_{narrow} is based on physiological data in the literature and is the main parameter for the sharp low-level tuning of the signal-path filter, the actual tuning properties of the complete model output are affected by the compressive nonlinearity of the model. The parameter τ_{wide} is chosen based on the desired filter gain at high levels. The difference between τ_{wide} and τ_{narrow} is directly related to the gain of the cochlear amplifier at a given CF:

$$\tau_{\mathrm{wide}} = \tau_{\mathrm{narrow}} 10^{-\mathrm{gain(CF)}/60}, \tag{6}$$

based on the third-order nonlinear filter.

The gain of the cochlear amplifier, or equivalently the amount of compression in the model, is a simple function of CF and is limited between 15 dB (at low CFs) and 70 dB (at high CFs) as follows:

$$\text{gain(CF)} = \max\{15, \min[70,\ 20 + 42\log_{10}(\text{CF}/1000)]\}. \quad (7)$$

The gain of the cochlear amplifier, or the amount of compression, has not been well characterized at many CFs in cat. The function above was chosen based on a maximum gain of 70 dB for high-CFs (Ruggero *et al.*, 1997; Nuttall and Dolan, 1996) and a minimum gain of around 15 dB at low CFs (Cooper and Rhode, 1996, 1997) observed in other mammalian species, with a smooth transition between low and high CFs, as observed psychophysically in humans (Hicks and Bacon, 1999). This gain function could be easily modified when data are obtained from cat; however, the present implementation represents the generally accepted concept that there is stronger compression at high CFs than at low CFs. For model responses to pure tones at stimulus frequencies more than an octave away from CF, there was not much change in gain as a function of stimulus level [see Eq. (2)], which is consistent with the data of Ruggero *et al.* (1997).

C. Wide-band feedforward control path

The function of the control path is to provide a time-varying signal $\tau_{sp}(t)$ to the signal-path filter such that several level-dependent response properties can be replicated by the signal-path filter. The control path is designed to reflect the active process corresponding to the local CF place as well as to the neighboring CFs. The control path consists of (a) a time-varying bandpass filter with a broader bandwidth than the signal-path filter; (b) a symmetrical nonlinear function to compress the dynamic range of the control signal; (c) a nonlinear function followed by a low-pass filter to control the dynamic range and dynamics of compression; and (d) a nonlinear function to adjust the total strength of compression.

The wide-band control filter is a third-order gammatone filter with its center frequency shifted 1.2 mm basal to the fiber CF along the basilar membrane (i.e., higher in frequency than CF). The size of the shift of the wide-band filter and the order of the filter were based on the shape of AN suppression tuning curves in the literature (e.g., Sachs and Kiang, 1968; Arthur *et al.*, 1971; Delgutte, 1990). The bandwidth of the nonlinear wide-band control-path filter is varied by $\tau_{cp}(t)$, which is a scaled version of $\tau_{sp}(t)$, the signal that controls the bandwidth of the nonlinear filter in the signal path. The scaling of $\tau_{sp}(t)$ by the factor K (<1, see Table I to create $\tau_{cp}(t)$ (Fig. 1) guarantees that the control-path filter has a wider bandwidth than the time-varying signal-path filter, and that the bandwidth ratio is constant. The gain of the wide-band control-path filter is normalized to 0 dB at the signal-path CF instead of at the center frequency of the control-path filter. As a result, the level-dependence of the gain of the control-path filter differs for frequencies above and below CF. This asymmetry is the key to producing the different properties of low-side and high-side suppression in the present model.

The analytical description of the control-path filter is given by the equation:

$$G_{cp}(\omega) \cong \{\text{gain}_{cp}(t)/[1 + jK\tau_{sp}(t)(\omega - \omega_{cp})]\}^{\gamma_{cp}} e^{-j\omega\alpha}, \quad (8)$$

where

$$\text{gain}_{cp}(t) = \sqrt{1 + [K\tau_{sp}(\omega_{\text{CF}} - \omega_{cp})]^2}, \quad (9)$$

j is $\sqrt{-1}$, and ω_{CF} is the radian frequency corresponding to the fiber's CF. The center frequency of the wideband control-path filter, ω_{cp} was computed using Liberman's (1982) frequency map based on the 1.2 mm basal shift from CF. The parameter $\text{gain}_{cp}(t)$ is calculated for every time step of the simulation to normalize the gain of the control-path filter to 0 dB at CF.

Experimental data show that the cochlear response is linear at low sound levels and becomes compressive at medium and high levels (Ruggero *et al.*, 1997). The slope of the compression has been shown to be as low as 0.2 dB/dB in the range of 40–80 dB SPL. Two different saturating nonlinear functions are used in the control path to implement this compression. A symmetrical logarithmic function

$$V[x(t)] = \text{sgn}[x(t)] B_{cp} \log(1 + A_{cp}|x(t)|^{C_{cp}}), \quad (10)$$

where $x(t)$ represents the output signal of the control-path filter; A_{cp}, B_{cp}, and C_{cp} are parameters which determine the (compressed) dynamic range of the signal before the second nonlinearity. The second function, an asymmetrical saturating nonlinearity, is a second-order Boltzmann function with an asymmetry of 7:1 (Mountain and Hubbard, 1996), given as

$$\text{out}(V) = \frac{1}{1 - \text{shift}_{cp}} \times \left\{ \frac{1}{1 + e^{-(V - x0_{cp})/s0_{cp}}\left(1 + e^{-(V - x1_{cp})/s1_{cp}}\right)} - \text{shift}_{cp} \right\}, \quad (11)$$

where $x0_{cp}$, $s0_{cp}$, $x1_{cp}$, and $s1_{cp}$ are parameters, and

$$\text{shift}_{cp} = \frac{1}{1 + e^{x0_{cp}/s0_{cp}}\left(1 + e^{x1_{cp}/s1_{cp}}\right)}. \quad (12)$$

The parameters in the two nonlinear functions above were adjusted by comparing compression-versus-level curves for the signal-path filter to physiological BM responses described in the literature (e.g., Ruggero *et al.*, 1997). The parameter A_{cp} determines the level at which the signal-path filter became nonlinear. Together with the parameters in the Boltzmann function, B_{cp} determines the level at which the signal-path filter became less compressive again. The parameter shift_{cp} guarantees that the nonlinear function passes through the origin. The parameter values are reported in Table I; these parameter values are invariant as a function of CF.

The two nonlinear functions are followed by a third-order low-pass filter. The cutoff frequency of the low-pass filter in the control path is set to 800 Hz. This cutoff frequency was chosen to produce an approximately 0.2 ms time constant for the onset of the compressive nonlinearity, consistent with the time course of compression estimated from click responses of the basilar membrane (see Fig. 8 in Recio *et al.*, 1998).[2]

The last stage of the control path is a nonlinear function that converts the output of the low-pass filter, $V_{LP}(t)$, to the time-varying time constant of the signal-path filter (Fig. 1):

$$\tau_{sp}(t)=\tau_{\text{narrow}}\left[R_0+(1-R_0)\left(\frac{\tau_{\text{wide}}/\tau_{\text{narrow}}-R_0}{1-R_0}\right)^{|V_{LP}(t)|/dc}\right]. \quad (13)$$

$\tau_{sp}(t)$ varies continuously between a maximal value of τ_{narrow} (which corresponds to a long time-constant, for a narrowly tuned filter) and an asymptotic lower bound, τ_{LB}. The value of R_0 is determined by the ratio $\tau_{LB}/\tau_{\text{narrow}}$. The parameter dc is an estimate of the dc component of the control-path output at high levels (i.e., a measure of the asymmetry of the control-path nonlinearities). The nonlinear function in Eq. (13) varies the dc value of $\tau_{sp}(t)$ from τ_{narrow} at low levels to τ_{wide} at high levels [see Eq. (6) for τ_{wide}]. The even function described in Eq. (13) causes the signal-path filter to be compressive for both the positive and negative parts of the instantaneous stimulus pressure for low frequency stimuli (Cooper, 1996). Note that the nonlinear filters in both the signal path and the control path are controlled by varying the filter time constants and associated gains: thus these filters are inherently stable (as long as positive time-constants and finite gains are specified). The time-varying time constant for the control path is simply a scaled version of the time constant for the signal path:

$$\tau_{cp}(t)=K\tau_{sp}(t) \quad \text{where} \quad K=0.2+0.8\frac{\tau_{\text{wide}}(\text{CF})}{\tau_{\text{narrow}}(\text{CF})}. \quad (14)$$

D. IHC-AN synapse

Physiological studies have shown that the IHCs transduce the mechanical responses of the basilar membrane to an electrical potential that results in the release of neurotransmitter at the synapse between the IHC and the AN fiber to generate action potentials in the AN fiber. Many studies have explored IHC potential changes in response to different stimuli. It is widely agreed that the synchrony coefficient of fibers responding to tones is affected by the ratio of the ac and dc components of the IHC response (e.g., Dallos, 1985; Palmer and Russell, 1986). This finding is the guideline for the analytical description of the present IHC model. The nonlinear function in the IHC model is a logarithmic compressive function

$$V_{\text{ihc}}(t)=A_{\text{ihc}}[P_{sp}(t)]\log(1+B_{\text{ihc}}|P_{sp}(t)|), \quad (15)$$

where $P_{sp}(t)$ is the output of the signal-path filter (Fig. 1). The function $A_{\text{ihc}}[P_{sp}(t)]$ and the parameter B_{ihc} were adjusted to achieve the appropriate IHC response properties, as follows: The asymmetry of the nonlinear function for $A_{\text{ihc}}[P_{sp}(t)]$, given by

$$A_{\text{ihc}}[P_{sp}(t)]=\begin{cases}A_{\text{ihc0}} & \text{for } P_{sp}(t)>0\\ -\dfrac{|P_{sp}(t)|^{C_{\text{ihc}}}+D_{\text{ihc}}}{3^*|P_{sp}(t)|^{C_{\text{ihc}}}+D_{\text{ihc}}}A_{\text{ihc0}}, & \text{for } P_{sp}(t)<0\end{cases} \quad (16)$$

changes smoothly as a function of the level of its input, $P_{sp}(t)$, from 1:1 to 3:1, such that at low sound levels, the dc response increases with a slope of 2 dB/dB compared with the 1 dB/dB slope of the ac response (Dallos, 1985). A_{ihc0} is a scalar, and the values of B_{ihc}, C_{ihc}, and D_{ihc} are constants that determine the SPLs of the inflection in the ac and dc components of the nonlinear function (see Table I). Note that these parameters are invariant with CF. This function for $A_{\text{ihc}}[P_{sp}(t)]$ guarantees the appropriate relationship between ac and dc response components, as opposed to a hyperbolic tangent (e.g., Carney, 1993) or Boltzman function, and was thus critical for obtaining realistic synchrony-versus-level responses for pure tones across a wide range of CFs (especially between 1 and 4 kHz).

The low-pass filter in the inner hair cell is a seventh-order filter with a cutoff frequency of 3800 Hz. This cutoff frequency was chosen to match the maximum sync coefficient versus CF for the model to data from cat (Johnson, 1980). The high order of the low-pass filter likely represents not only the low-pass filtering properties of the IHC membrane but other low-pass mechanisms such as the calcium-related synaptic processes (Weiss and Rose, 1988).

The nonlinear IHC-AN synapse also affects the characteristics of the AN fiber discharge patterns. A simplified implementation of a previous time-varying three-store diffusion model (Westerman and Smith, 1988; Carney, 1993) was used in the present model. The parameters in the current model were determined according to the equations in the Appendix of Westerman and Smith (1988) based on desired characteristics of post-stimulus-time (PST) histograms for tones (see the Appendix). The values of parameters used in the model are provided in Table I. A detailed description and discussion of the synapse parameters and their effects on the PST histograms as a function of spontaneous rate will be discussed in another paper. The results presented here are limited to high-spontaneous-rate model fibers, and thus depend less strongly on the details of the synapse model.

The immediate permeability $P_I(t)$ is a soft rectifying function of the model inner-hair-cell response, V_{ihc} (Fig. 1), described as

$$P_I(t)=p_1\log(1+e^{p_2V_{\text{ihc}}(t)}), \quad (17)$$

where p_1 determines the immediate permeability at rest and the spontaneous rate of the model fiber. The parameter p_2 is given by

$$p_2=\begin{cases}1165 & \text{for CF}<685\ \text{Hz}\\ -5430+1010\log(\text{CF}) & \text{for CF}>685\ \text{Hz}\end{cases} \quad (18)$$

and determines the slope of the relationship between P_I and V_{ihc}. Therefore p_2 affects the threshold of the model fiber. The CF-dependence of p_2 in Eq. (18) adjusts the thresholds of model fibers at CF to be approximately 0 dB for all CFs. The numerical expressions in Eqs. (17) and (18) (and Table I for p_1) were derived from basic parameters related to the IHC-Synapse model (see the Appendix for details). The effects of the external and middle ears on the threshold of AN fibers can be included in future models by appropriately attenuating the input to the model as a function of frequency, referenced to the 0 dB baseline threshold in the present model. The output of the diffusion model is the time-varying discharge rate $s(t)$ prior to the inclusion of refractory effects,

where $s(t)=P_I(t)C_I(t)$. For further details of the IHC-AN synapse model, see the Appendix; also see the appendices of Westerman and Smith (1988) and Carney (1993).

E. Discharge generator

The model discharge times are produced by a renewal process that simulates a nonhomogeneous Poisson process driven by the synapse output $s(t)$ and modified to include refractory effects (Carney, 1993). The time-varying arrival rate of the Poisson process is described as

$$R(t)=s(t)[1-H(t)]. \tag{19}$$

The discharge-history effect, $H(t)$, was determined by a sum of two exponentials (Westerman and Smith, 1985):

$$H(t)=\begin{cases} c_0 e^{-(t-t_1-R_A)/s_0}+c_1 e^{-(t-t_1-R_A)/s_1} \\ \quad \text{for } (t-t_1)\geq R_A \\ 1.0 \quad \text{for } (t-t_1)<R_A \end{cases}, \tag{20}$$

where t_1 is the time of the preceding discharge, and c_0, c_1, s_0, s_1 are parameters (Table I). Discharges are not allowed to occur during the absolute refractory time R_A, and $H(t)$ varies continuously from 1 to 0 as the interval from the previous discharge increases beyond R_A. The statistics of the discharge are affected by refractoriness (Teich and Lachs, 1979), and the parameters in the expression for $H(t)$ (Table I) were adjusted to match the statistical properties of the responses (mean and variance of discharge rate) to published data (Young and Barta, 1986; Winter and Palmer, 1991).

III. RESULTS

A. Responses to pure tones

Figure 2 illustrates an overview of the responses of several model stages for three fibers tuned to different frequencies in response to 60 dB SPL pure tones at CF. The magnitude of the signal-path filter output $P_{sp}(t)$ decreases as the fiber's CF increases due to the larger compression at high CFs. The control signal $\tau(t)$, which varies the gain and bandwidth of the signal-path filter, varies on a cycle-by-cycle basis with the stimulus at low frequencies. At higher frequencies, the control signal becomes increasingly dominated by dc energy. Also, the IHC voltage $V_{\text{ihc}}(t)$ changes on a cycle-by-cycle basis at low frequencies, but is dominated by a dc bias for responses to pure tones at high frequencies. The adaptation in the diffusion model shapes the onset response of the fiber in the synapse output $s(t)$.

Threshold tuning curves are shown in Fig. 3(a) using the paradigm of Liberman (1978). The thresholds at CF are around 0 dB; these thresholds can be adjusted by adding a middle-ear model to simulate the changes in threshold as a function of frequency that contribute to the audiogram. The model's Q_{10} bandwidths [Fig. 3(b)] are comparable to Q_{10} data for normal cats from Miller *et al.* (1997). The high-CF fibers respond at low frequencies, despite the lack of an explicit ''tail'' (Kiang and Moxon, 1974; Kiang, 1975; Liberman, 1978; Narayan *et al.*, 1998) mechanism included in this model. This low-frequency response is due to distortion introduced by the nonlinear filter in the signal path. The low-frequency portion of the tuning curve for high-CF model fibers is strongly influenced by the cutoff frequency of the low-pass filter in the control path, which influences the dynamics of the nonlinear variation in tuning and thus the degree of distortion. The details of the low-frequency response (tail) of high-CF fibers were not a focus of this study.

The response to a CF tone grows at a rate less than 1 dB/1 dB due to basilar membrane compression. It has been reported by several authors (e.g., Ruggero *et al.*, 1997; Rhode and Cooper, 1996) that the basilar membrane responds linearly at low levels and is most compressive between 40 and 80 dB SPL. The growth rate for high-CF fibers can be as low as 0.2 dB/dB (Ruggero *et al.*, 1997). The compression gain (the gain difference between low levels and high levels) varies from approximately 10 dB to 70 dB as the CF of the fibers increases (Nuttall and Dolan, 1996; Cooper and Rhode, 1996, 1997; Ruggero *et al.*, 1997). The response magnitude (ac RMS) of the signal-path filter output as a function of CF-tone level is illustrated in Fig. 4(a). As the stimulus level increases, the average value of the control signal is decreases as the system changes from linear to compressive. At high levels, the model response is less compressive due to the saturation of the control path. The results of the model are similar to the nonlinear response properties described by Rhode and Cooper (1996) and Ruggero *et al.* (1997), except perhaps at levels above 80 dB SPL, where the most sensitive cochleae appear to have basilar-membrane responses that remain compressive at very high levels (Ruggero *et al.*, 1997). The amount of compression in the model as a function of CF is illustrated in Fig. 4(b).

As the level of the input tone is changed, the fiber response properties, such as average rate, synchrony, and PST-histogram shape, also change. The phases of the temporal AN responses are affected by continuous changes in the phase versus frequency properties of the signal-path filter; the phase properties of the signal-path filter are time-variant, as they are affected by the control-path signal. As the bandwidth of the signal-path filter changes, its phase properties also change. The temporal response properties are also affected by the low-pass filter in the IHC model, which limits the phase-locking of the response at high frequencies (and introduces a time-invariant phase shift). The diffusion model of the IHC-AN synapse also has some influence on temporal response properties, for example, interval statistics are influenced by the adaptation and refractory properties of the synapse model.

It is believed that depolarizing voltage responses of the IHC determine the detailed firing patterns in AN fibers. The neural synchrony of AN fibers especially depends on the ratio between the ac and dc components of the IHC receptor potential (Dallos, 1985; Palmer and Russell, 1986; Cheatham and Dallos, 1993). The ac and dc responses for model fibers at two CFs are shown in Fig. 5. For the fiber with a CF of 1 kHz, the ac potential always dominates the output of the IHC, resulting in a high synchronization coefficient. For high-CF fibers, the dc potential dominates the IHC output, and the synchronization coefficient is lower. At low levels, the dc response in the model increases at a rate of 2 dB/dB, and the ac response increases at a rate of 1 dB/dB, as re-

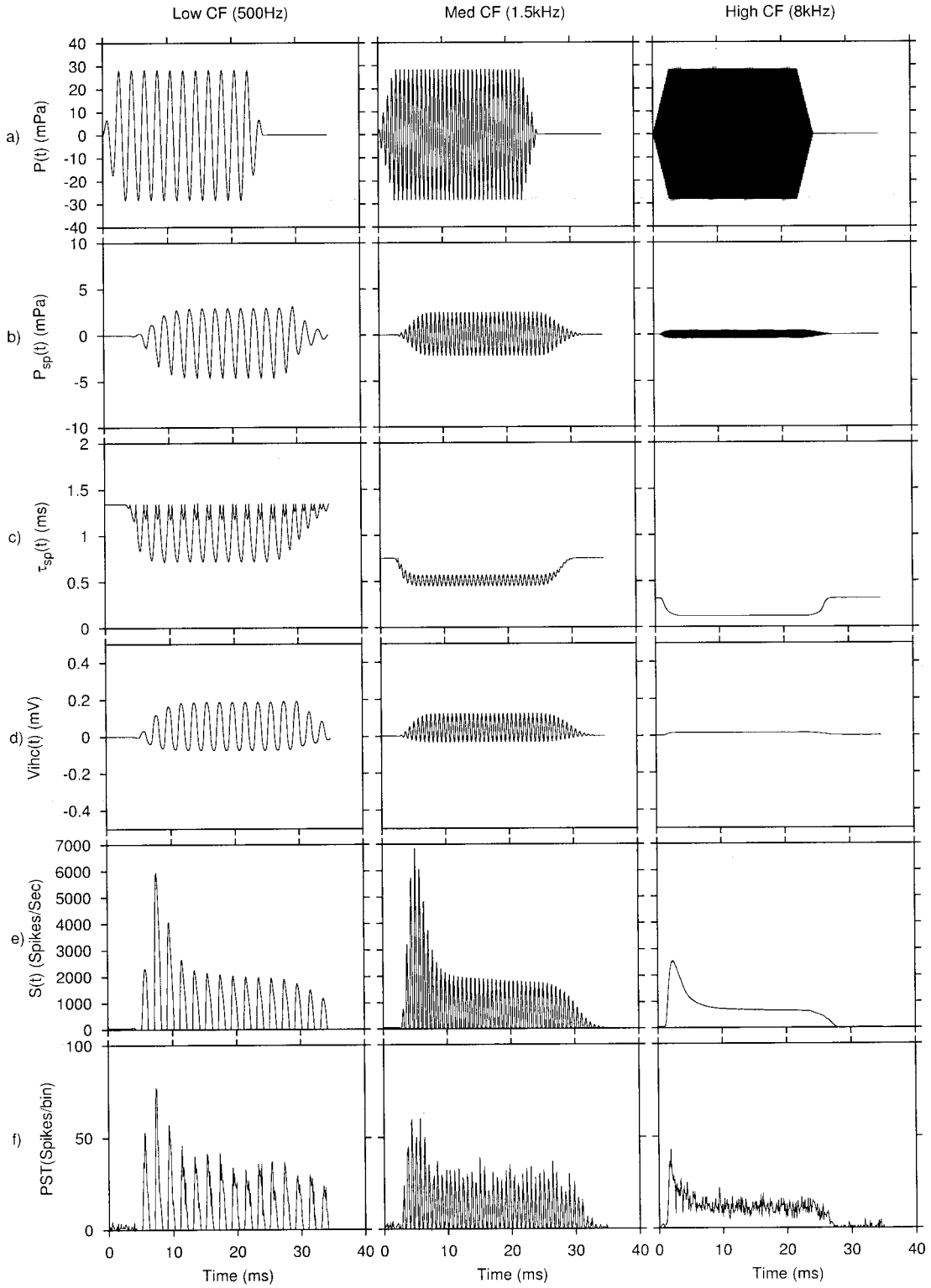


FIG. 2. Responses of several model stages to a CF tone for different model-fiber CFs. The stimulus was a 25 ms duration tone burst with a 2.5 ms rise/fall time and was presented at 60 dB SPL. The waveforms shown (see Fig. 1) are (a) $P(t)$, stimulus, (b) $P_{SP}(t)$, signal-path filter output, (c) $\tau(t)$ [or $\tau_{sp}(t)$], control signal, (d) $V_{IHC}(t)$, IHC response, (e) $S(t)$, synapse output, and (f) PST histogram based on 500 presentations and 0.1 ms bin size.

ported in the literature (Dallos, 1985). If a transition exists from an ac-dominated output to a dc-dominated output within the fiber's dynamic range, then experimental data would be expected to show dramatic nonmonotonic changes in the neural synchrony coefficient as a function of sound level, especially for mid-frequency CFs (1–4 kHz). This phenomenon is not found in experimental responses (Joris, 1999), and the model for the IHC was designed to avoid strongly nonmonotonic synchrony-level functions by using the asymmetrical function in Eq. (14).

Figure 6(a) shows the ac and dc responses of the IHC stage for an 800 Hz fiber in response to frequencies below,

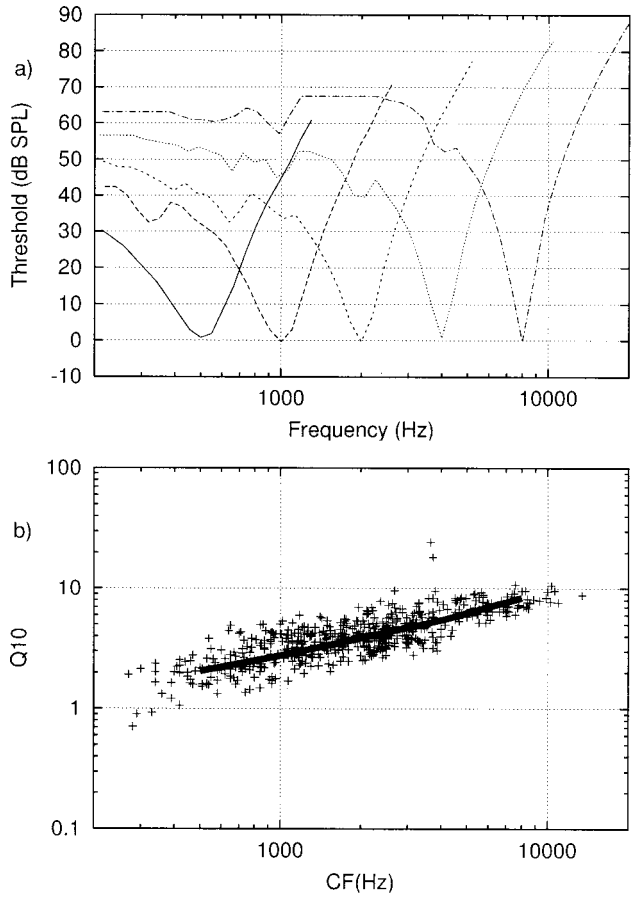


FIG. 3. (a) Threshold tuning curves for fibers with different CFs. The tuning threshold was defined as the level that results in a response 10 spikes/s greater than the spontaneous rate (200 repetitions of each stimulus). The stimulus is a 50 ms tone burst with a 2.5 ms rise/fall time (Liberman, 1978). Differences in discharge rates were counted between the 50 ms tone burst interval with 1.25 ms delay and the subsequent 50 ms silent interval. (b) Q_{10} measured from the tuning curve of model fibers (solid line) compared with the data (crosses) from Miller *et al.* (1997, Fig. 3). A linear fit of these data was used to determine the bandwidth of the signal-path filter at low SPLs. Q_{10} was then recomputed from the simulated tuning curves of the complete nonlinear model across a range of CFs to validate the model tuning.

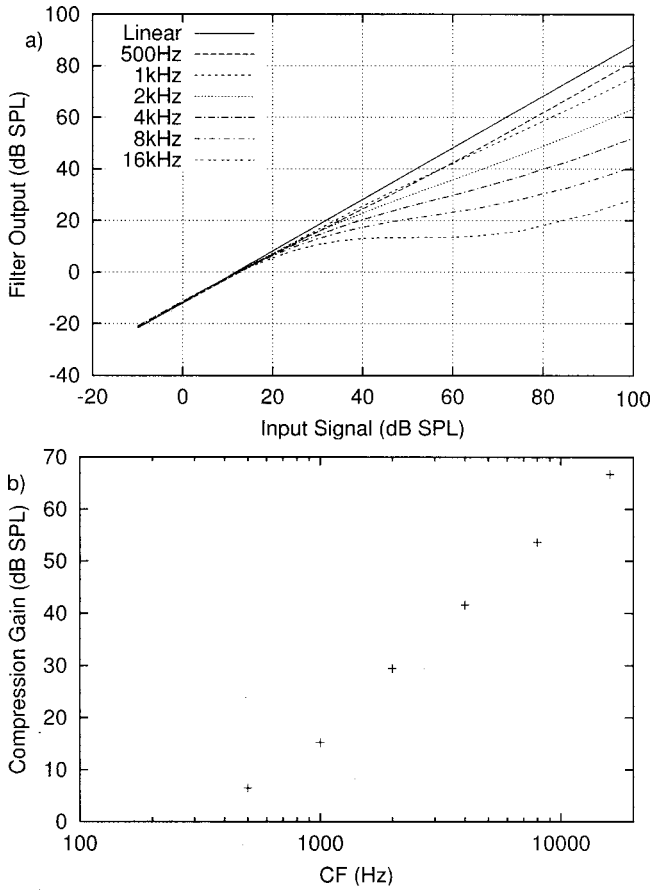


FIG. 4. Compressive nonlinearity of the signal-path filter output in response to CF tones for different CFs. (a) rms of the ac component (based on the peak-to-peak amplitude of the sinusoidal steady-state response) of the filter output as a function of stimulus intensity. The rms was computed from 10 cycles of the stimulus starting 40 ms after the onset of the stimulus. The solid line is the output of a linear filter. (b) Compression gain (the reduction in gain due to the compressive nonlinearity) for model fibers at different CFs. The compression gain was calculated as the difference of the gain of the responses to CF tones at 0 dB SPL and 120 dB SPL (Rhode and Cooper, 1996; Ruggero *et al.*, 1997).

at, and above CF. For comparison, Fig. 6(b) shows an example of ac and dc components of responses recorded from an IHC (Dallos, 1985). The model responses [Fig. 6(a)] to tones at frequencies below, at, and above CF are generally similar to the data, however, they do not saturate as completely as those shown in the examples from Dallos [Fig. 6(b)]. Other reports of IHC responses indicate that these cells may not completely saturate (e.g., Russell *et al.*, 1986), and recent models derived to fit IHC responses have incomplete saturation (e.g., Zagaeski *et al.*, 1994; Mountain and Hubbard, 1996). For example, the dotted curve in Fig. 6(a) shows the fit used by Zagaeski *et al.* (1994) to the dc component of IHC responses to tones well below CF. The relatively large dynamic range of the IHC response in the model is important in order to support the wide dynamic range of low-spontaneous-rate fibers, which will be pursued in future extensions of this model.

Figure 7 illustrates average discharge rate versus level and synchronization coefficient versus level functions in response to pure-tone stimuli at CF for fibers with CFs of 1 kHz and 4 kHz. The simulations presented in this study are limited to high-spontaneous-rate fibers; both fibers in Fig. 7 have spontaneous rates of about 45 sp/se. The sustained rate has a dynamic range of 40 dB and the onset rate has a wider dynamic range (Smith, 1988). The synchronization coefficient reaches its maximum at about 10 dB above threshold and then drops slightly as level increases, similar to AN-fiber responses (Johnson, 1980).

The increased dynamic range of the onset rate is a result of the adaptation included in the IHC-AN synapse model. This adaptation can also be illustrated by PST histograms of responses to tones (Fig. 8). The shape of the PST histogram changes as SPL is increased. The peak-to-sustained discharge rate increases with SPL, and the latency of the response decreases by integral multiples of 1/CF (Kiang *et al.*, 1965; Carney, 1993).

Rate and phase responses to pure tones at frequencies away from CF provide more information about the nonlinear

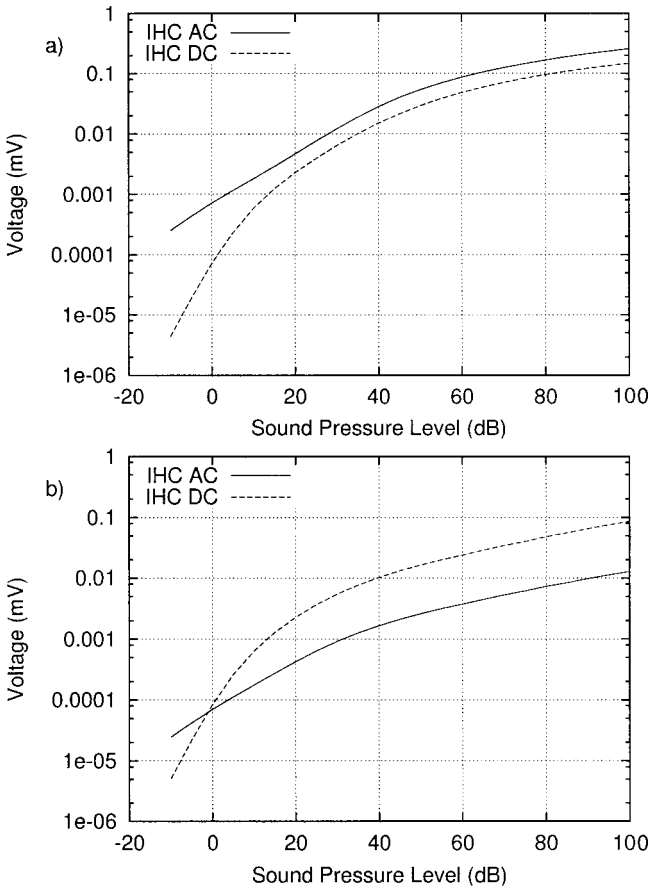


FIG. 5. IHC responses to pure tones at CF versus stimulus intensity for model fibers with CFs of (a) 1 kHz and (b) 4 kHz. The responses were measured over 10 cycles of the stimulus beginning 40 ms after the onset of the stimulus. The ac and dc components of the responses increase at different rates at low levels, and the dc component dominates the output as the CF of the fibers increases (Dallos, 1985).

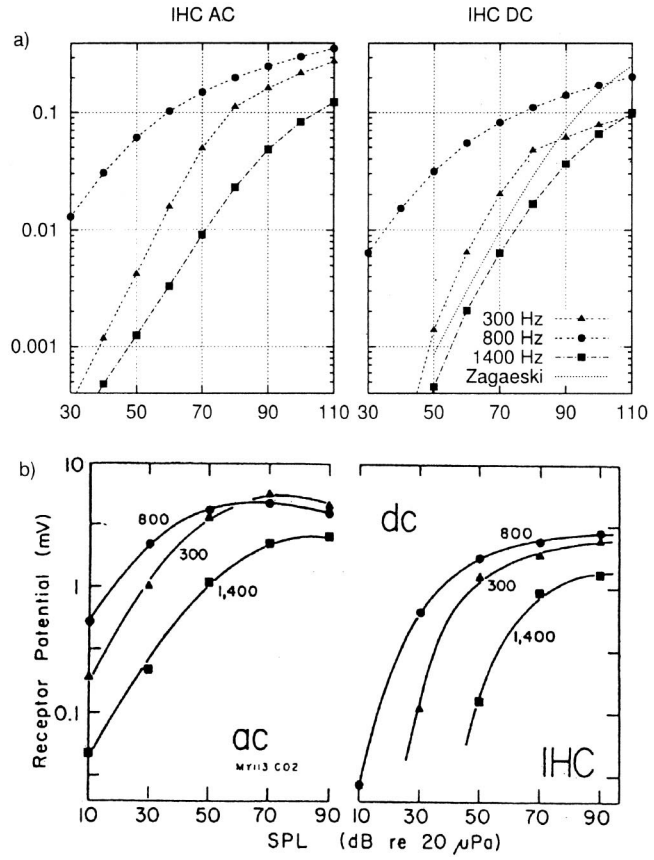


FIG. 6. ac and dc response components of (a) model IHC and (b) actual IHC responses (Dallos, 1985; with permission). Both model and actual IHCs have a CF of 800 Hz, responses to tones at 300, 800, and 1400 Hz are shown. The dynamic range of the model and actual responses are similar, although the threshold of the actual IHC in this example is lower than that of the model. The dotted line in the model dc response plot (upper right) is the fit of Zagaeski *et al.* (1994) to IHC responses to tones well below CF. Zagaeski *et al.*'s fit was shifted up to the normalized units of the IHC model (by setting Zagaeski *et al.*'s parameter $RP_{\max}$ to 0.6).

tuning of the model. Figure 9 illustrates the response area (responses to tones across a range of frequencies and levels) for an AN model fiber with CF of 2300 Hz, chosen for the purpose of comparison with the example shown from Anderson *et al.* (1971). Changes in rate as a function of level for frequencies above and below CF are illustrated in the upper panels. The response area (iso-level contours of discharge rate) spreads as the input level increases. A level-dependent shift in the peak frequency of the response area was not observed in the results due to the symmetry of the gammatone filter, which is a limitation of the model. Changes in the phase of phase-locked responses as a function of level, above and below CF, are illustrated in the lower panels. The phases are referenced to the response at 90 dB SPL, following Anderson *et al.*'s (1971) convention. Systematic level-dependent phase changes have been observed in basilar membrane motion (Ruggero *et al.*, 1997), in IHC responses (Cheatham and Dallos, 1993), and in AN fibers (Anderson *et al.*, 1971). The level-dependent change in phase of the phase-locked responses is consistent with the phase change expected due to broadening of the signal-path filter as level increases. The size of the phase change for low levels was comparable to the data from Anderson *et al.* (1971); maximum phase changes were approximately 0.5π.

Figure 10 illustrates level-dependent discharge rates and phases, across a range of stimulus frequencies, for a high-CF fiber. In this case, the phases were computed from the output of the signal-path filter and are compared to measurements of basilar-membrane phase from Ruggero *et al.* (1997). The phase change for high-CF model fibers was stronger than for the low-CF fiber because of the stronger compression at high-CFs [Fig. 4(b)]. Due to the rolloff in synchrony at high CFs (see below), it is difficult to compare the level-dependent phase at high CFs to AN responses. Illustration of the level-dependent phase based on the filter output also makes it clear that this phenomenon is a result of the nonlinear filter in the signal path. The IHC, synapse, and discharge generator models that follow the signal path do not introduce level-dependent phase shifts.

The maximum synchronization of responses to pure tones at CF as a function of CF is an important description of the temporal response of AN fibers. The synchronization coefficient was strongly influenced by the low-pass properties of the IHC model. Other factors, such as the low-pass properties of the synapse, also affect the synchronization of the AN fiber (Weiss and Rose, 1988). Figure 11 shows the maxi-

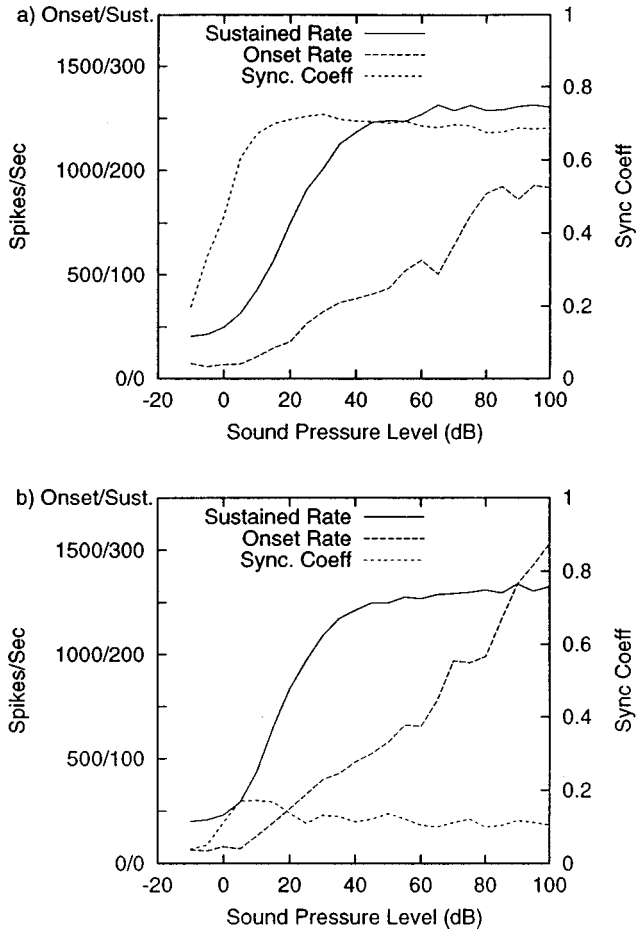


FIG. 7. Rate-Level and Sync-Level functions for model fibers with CFs of (a) 1 kHz and (b) 4 kHz. Sustained rate is calculated in the 10–45 ms response time window for 400 repetitions. Onset rate is the maximum discharge rate during the first 10 ms and was calculated using 0.5 ms bins (bin size is one stimulus cycle if the stimulus cycle is greater than 0.5 ms). The onset response in the figure is scaled (divided by 5.0) for better illustration. The discharges in the 10–45 ms time window were analyzed using a 32 bin period histogram from which the synchronization coefficient was then calculated (scale on right) (Johnson, 1980).

mum synchronization coefficient for the model and for data from cat (Johnson, 1980). The model synchronization coefficient is within the scattered data, but is slightly lower than the mean of the data. The parameters of the seventh-order low-pass filter and of the IHC-Synapse model were adjusted to achieve this representation of the rolloff in synchrony as a function of CF over the wide range of CFs modeled.

B. Responses to complex sounds

One goal of this study was to develop a model capable of simulating the responses of AN fibers to complex sounds. Interactions between different components of complex sounds have been studied for many years to gain a better understanding of these responses. Two-tone suppression is a nonlinear phenomenon of AN responses whereby the rate response to a pure tone at CF can be suppressed by a second tone (suppressor) (Nomoto *et al.*, 1964; Sachs and Kiang, 1968; Delgutte, 1990). More recent studies have reported two-tone suppression in the responses of IHCs and in cochlear mechanics (Cheatham and Dallos, 1989, 1990, 1992; Rhode, 1977; Ruggero *et al.*, 1992; Cooper and Rhode, 1996; Cooper, 1996). These observations have shown that the two-tone suppression nonlinearity originates in the mechanics of the cochlear partition (Ruggero *et al.*, 1992). Two-tone suppression is asymmetrical for low- and high-side suppressors, both in terms of the shape of the suppression threshold tuning curve and in the suppression growth rate for tones above and below CF (e.g., Arthur *et al.*, 1971; Prijs, 1989; Delgutte, 1990).

Figure 12 shows suppression tuning curves of two low-frequency model fibers with CFs at 1 kHz and 2 kHz. The asymmetry of the suppression-tuning threshold is due to the shift of the center frequency of the wide-band control-path filter with respect to the model fiber's CF. The asymmetry of the model suppression tuning curve is comparable to that reported in the literature across a range of CFs (e.g., Sachs and Kiang, 1968; Arthur *et al.*, 1971; Delgutte, 1990). Suppression was found even when the level of the suppressor was in the excitatory region of the tuning curve. This is because the CF tone level is set at a level which results in 2/3 of the maximum driven discharge rate (e.g., Delgutte, 1990), and the suppressor acts to suppress the response to the CF tone at this level. The relatively weak suppression for low-CF fibers, as compared to high-CF fibers (see below), especially for frequencies below CF, is due to the relatively

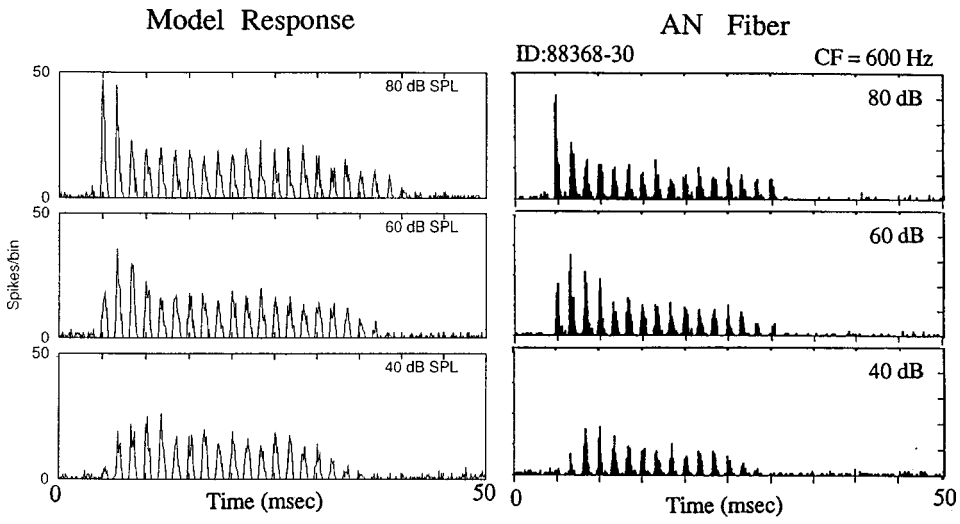


FIG. 8. PST histograms for (a) model and (b) actual AN fibers with CF equal to 600 Hz. Responses are to pure tones at CF at three different SPLs. The stimulus was 25 ms in duration and presented every 50 ms. The rise/fall time was 3.9 ms and the bin size was 0.1 ms. The PST histograms contain responses to 200 presentations at each SPL. The AN response was recorded as part of another study (Carney, 1990; see also Fig. 7, Carney, 1993).

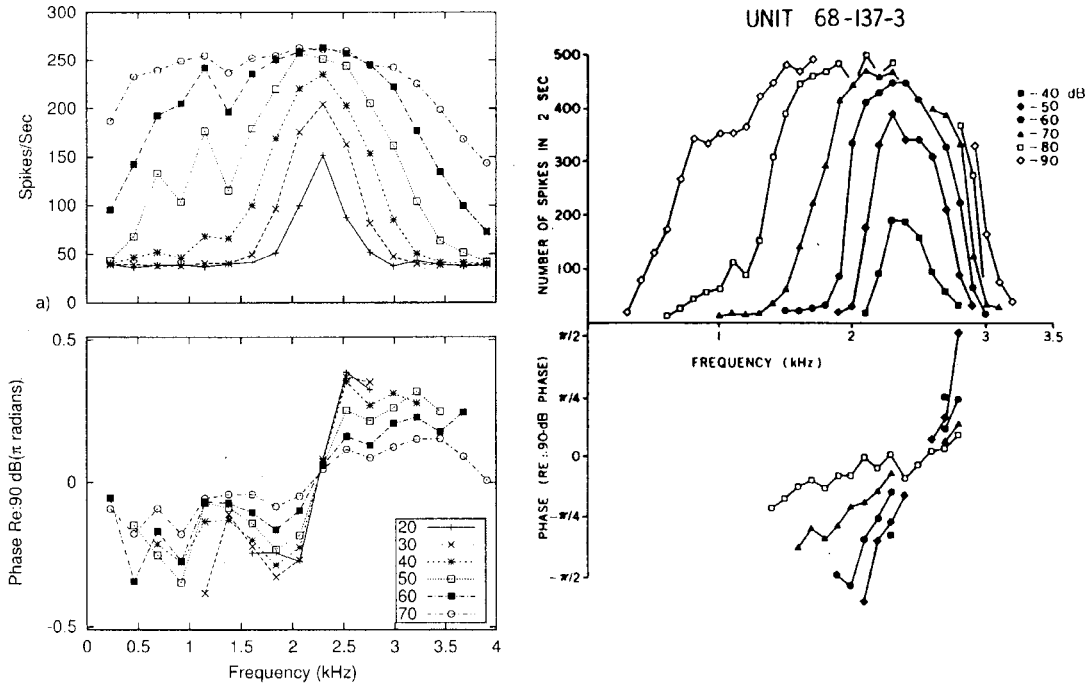


FIG. 9. Comparison of model and actual response areas. Left: Intensity-dependent discharge rate and phase responses for a model fiber with CF of 2300 kHz. Rates were computed from the sustained responses to 100 repetitions of a 50 ms tone burst. The phase for each response was referenced to the phase in response to that frequency at 90 dB SPL (e.g., Anderson *et al.*, 1971). Right: Rates and phases for the response area of an actual AN fiber with the same CF from Anderson *et al.* (1971, Fig. 8; with permission). The levels chosen for the model simulation (20–70 dB SPL) were selected to approximately match the levels, with respect to threshold, of Anderson *et al.*'s data set (i.e., note that the response of the model at 20 dB SPL is similar to the AN fiber's response at 40 dB).

low gain of the cochlear amplifier for low CFs [see Fig. 4(b)]. This pattern of suppression at low CFs is consistent with physiological data (Delgutte, 1990).

The upper panels of Fig. 13 show an example of a two-tone suppression tuning curve for a high-CF model fiber and an actual suppression tuning curve with the same CF (Delgutte, 1990). The general shape of the AN-model suppression tuning curve is similar to those reported for AN fibers: suppression thresholds are lowest for frequencies just above CF, and there is a broad low-frequency region of suppression. These two properties were introduced into the model by shifting the center frequency of the control-path filter to a frequency slightly higher than the CF of the signal-path filter (see above). In addition to the asymmetry in thresholds above and below CF, there is an asymmetrical growth of two-tone suppression for suppressors above and below CF. This property is illustrated in the lower two panels of Fig. 13, which show the level of a tone at CF that is required to maintain a constant discharge rate (i.e., to offset the suppression) as a function of suppressor level. The suppression growth curves for suppressor frequencies above CF are shallower than for tones below CF in both model and actual AN responses (Delgutte, 1990). The asymmetry in the model responses is qualitatively correct; however, it is not as strong as in the actual responses, especially for high CFs for which the asymmetry in growth is stronger than for low-CF fibers (Delgutte, 1990). More realistic filter shapes, as opposed to the symmetrical gammatone filter, would be expected to enhance this asymmetry. Also, the model suppression at very low frequencies does not grow as high as in the actual data; this limitation may be related to the lack of a tail mechanism at low frequencies in the model.

Another illustration of the differences between two-tone suppression properties for suppressors below and above CF is provided in Fig. 14. These plots show rate-level functions for a CF-tone alone, and for a two-tone stimulus with $F1 = \mathrm{CF}$, and $F2$ either above CF [Figs. 14(a), (c)] or below CF [Figs. 14(b), (c)]. Model responses are shown for two fibers with CFs chosen to match examples from Sachs and Abbas (1976). The difference in level between the suppressor tone and the CF tone are fixed. The range of SPLs shown, and the difference in levels between tones, were chosen to yield rate-level functions comparable to the examples (note that the model fibers have lower thresholds).

The functions in Fig. 14 show several of the trends as seen in the data (Sachs and Abbas, 1976). For a suppressor above CF [Figs. 14(a), (c)], the response to the two-tone complex begins to diverge from the CF-tone response at low levels, and the growth of suppression (as both the CF tone and the suppressor are increased in level) is gradual. For a suppressor below CF [Figs. 14(b), (d)], the threshold of suppression is high with respect to that for the CF tone alone. The growth of suppression as level increases above suppression threshold is steeper for the below-CF suppressor than for the above-CF suppressor, although suppression by below-CF frequencies, especially at high levels, is not as strong in the model as in the data (as seen above, Fig. 13). The lack of low-frequency tail mechanisms may impose a limitation on the behavior of below-CF suppressors in this model. The differences between the properties of the above-

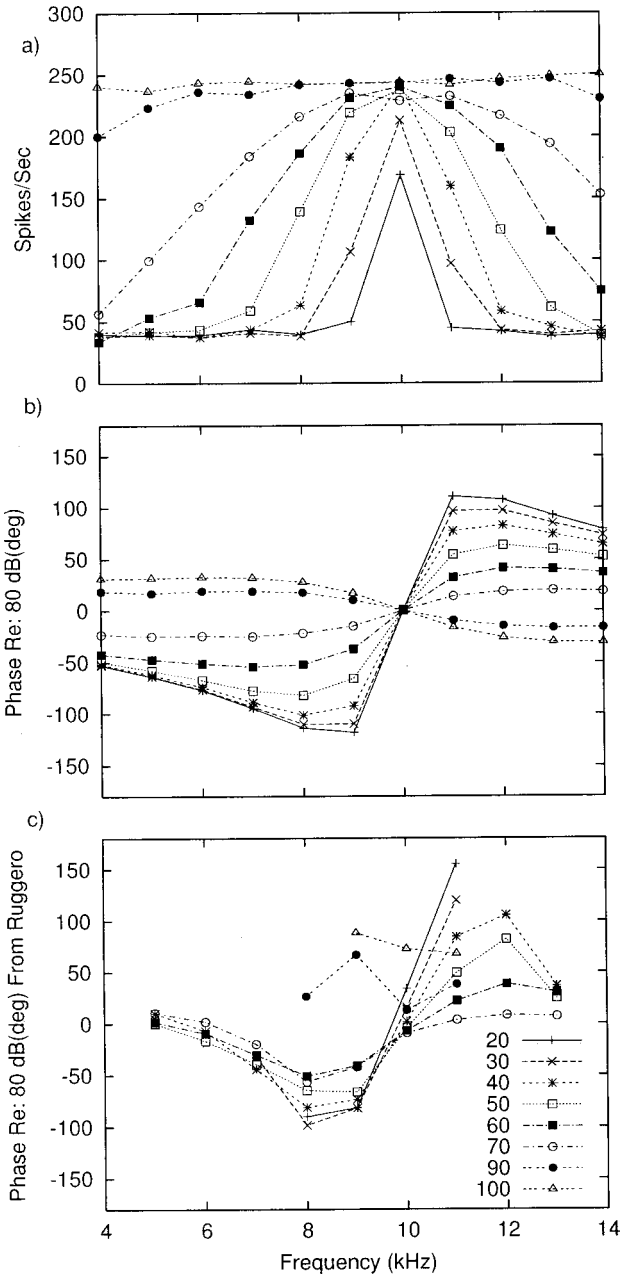


FIG. 10. Response area rates (a) and phases (b) for a model fiber with CF equal to 10 kHz. The phases are measured from the output of the signal-path filter; rolloff of phase-locking precludes using phase-locked discharge times to illustrated the level-dependent phase properties of the high-CF model fiber. Phases are referenced to the phase of the response at each frequency at 80 dB SPL, as in Ruggero *et al.* (1997). (c) For comparison, phases responses from the 10 kHz place of the basilar membrane of chinchilla (see Fig. 14 in Ruggero *et al.*, 1997; data are replotted here using the same convention as Fig. 9).

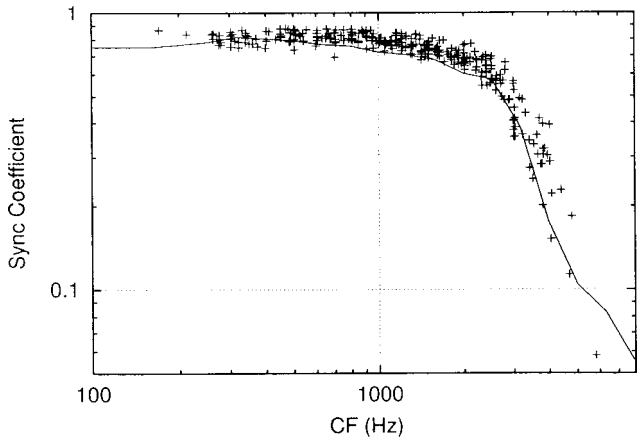


FIG. 11. Maximum synchronization coefficient of model AN fibers to CF tones as a function of CF (solid curve). Each coefficient was computed as the peak of the sync-level function (see Fig. 7) for a model fiber at each CF. The stimulus was 100 ms in duration with 3.9 ms rise/fall times. The synchronization coefficients were computed from the 10–100 ms time window referenced to the stimulus onset. Maximum synchronization coefficients for a population of AN fibers in cat (Johnson, 1980) are plotted (crosses) for comparison.

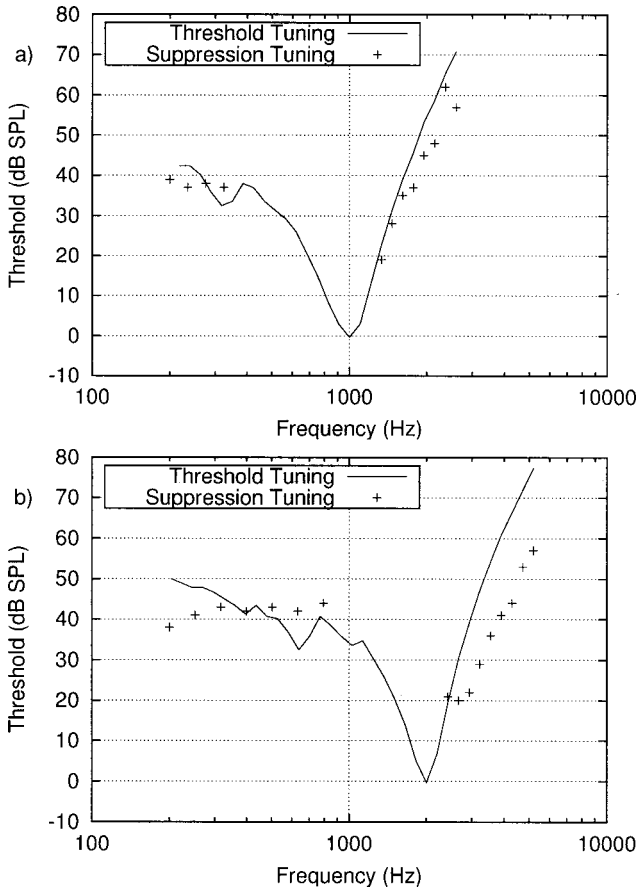


FIG. 12. Excitatory-tuning (line) and suppression-tuning (plusses) curves for model fibers with a CF at (a) 1 kHz and (b) 2 kHz. Suppression thresholds were measured using the algorithm of Delgutte (1990). Two 50 ms stimuli (combination of suppressor tone and CF tone and CF tone alone) were presented in alternation and the average discharge rate was calculated over 200 repetitions. The level of CF tone was set to produce 2/3 of the saturated rate response in the model AN fiber (typically at 20 to 25 dB SPL). The suppressor threshold was then found as the level of the suppressor that resulted in a 10 sp/s decrease in the response to the combined tones as compared to the response to the CF tone alone. The bandwidth of the suppression-tuning curve is mainly determined by the wide-band filter in the control path.

and below-CF suppressors are further illustrated by the fractional response curves (dotted lines, Fig. 14) which show the ratio between the response to the CF tone alone and the two-tone complex (Sachs and Abbas, 1976).

Addition of a suppressor tone affects not only the rate of the response to a tone at CF, but also its phase (e.g., Deng and Geisler, 1985). Figure 15(a) illustrates rate-level functions in response to $F1$ (a tone at CF, 2900 Hz) and to the two-tone complex, $F1+F2$, where $F2$ is a suppressor tone at 4700 Hz with a fixed level of 50 dB SPL. Figure 15(b)

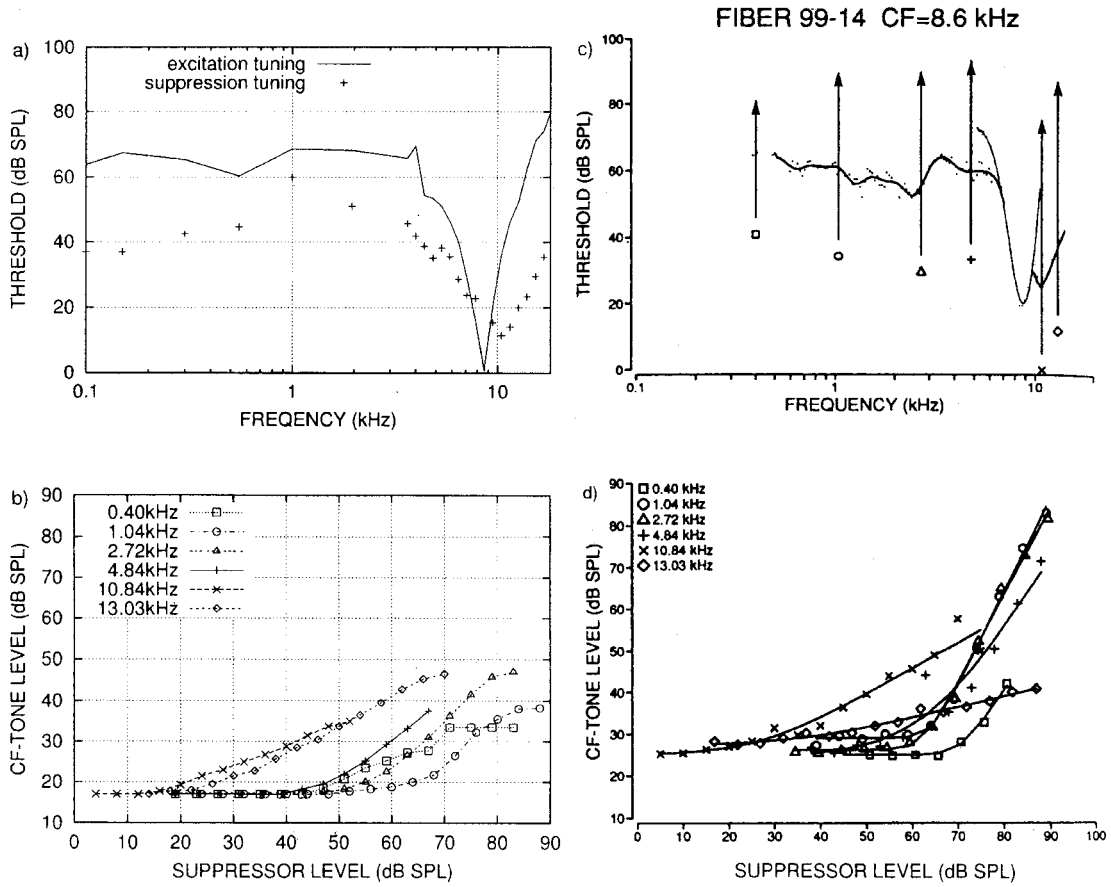


FIG. 13. Two-tone suppression tuning curve for a model fiber with CF equal to 8.6 kHz (left) in comparison to and AN fiber with the same CF (right, Delgutte, 1990; with permission). The upper panels show excitatory and suppression tuning curves (see Fig. 12 caption). The lower panels show growth of suppression for suppressors at several frequencies. The level of the CF tone required to maintain a criterion rate (2/3 of saturated rate), in the presence of the suppressor tone, is plotted as a function of the suppressor level (Delgutte, 1990). The asymmetry in high-side and low-side growth rate is comparable with experimental results from Delgutte (1990) for low-CF AN fibers, but underestimates the asymmetry for high-CF AN fibers.

shows the phase of the phase-locked responses to $F1$ alone, and the phase of the response to the $F1$ component in the two-tone complex, $F1+F2$. The trends seen in the model phase are similar to those reported by Deng and Geisler [1985, Figs. 15(c), (d)] for the majority of AN fibers. The main effect that they reported, also seen in the model response, is that addition of the suppressor $F2$ causes an increase in phase lag (where phase lag is the phase delay of the response with respect to the stimulus phase). The change in phase is seen in the response to $F1$ at levels of $F1$ where rate-suppression is seen; at higher levels of $F1$, where the model rate-level functions in response to $F1$ and $F1+F2$ converge [Figs. 15(a), (c)] the model phase functions in response to $F1$ and $F1+F2$ also converge [Figs. 15(b), (d)]; similar trends were reported for AN fibers (Deng and Geisler, 1985). Note that the change in phase with level of the model response to $F1$ is much less than for the AN data; this discrepancy is consistent with the fact that CF is level-dependent in AN fibers in this CF range while the model CF is level-independent. A change in phase in response to the CF tone of the AN fiber is also consistent with the possibility that the CF of the actual AN fiber was not exactly equal to the frequency chosen for $F1$ (whereas in the model, the CF is known); a mismatch between CF and $F1$ would result in level-dependent phase (as illustrated in Fig. 9).

The final aspect of the nonlinear response of the model to two-tone complexes illustrated here are the combination tones, or distortion products (Fig. 16). In this figure, the magnitudes of the components in the signal-path filter in response to $F1$, $F2$, and the distortion products $2F1-F2$ and $2F2-F1$, are shown. Although this filter-bank model does not provide a medium for propagation of the distortion products to the place tuned to the distortion product frequency, it is still possible to examine the trends of the low-level distortion products created by a nonlinear filter model (e.g., Goldstein, 1995). Figure 16(a) shows the response of the model's signal-path filter (shown as the magnitude of the Fourier component at the frequency of interest) to a CF tone alone, to $2F2-F1$ with $F2/F1=1.1$ and 1.2, and to $2F1-F2$, with $F2/F1=1.1$. In each case, $F1$ and $F2$ have equal amplitudes, and their frequencies are chosen such that the frequency of the combination tone is equal to the model CF of 8.5 kHz, chosen to match the CF of the example in Robles *et al.* [1991, Fig. 16(b)]. The trends in the level-dependence of both of these distortion products as a function of level, and for different $F2/F1$ ratios, are consistent with measurements from the basilar membrane. The amplitudes of the distortion products in the model are lower than those reported by Robles *et al.* (1991), which might be expected due to the fact that propagation of the distortion products and

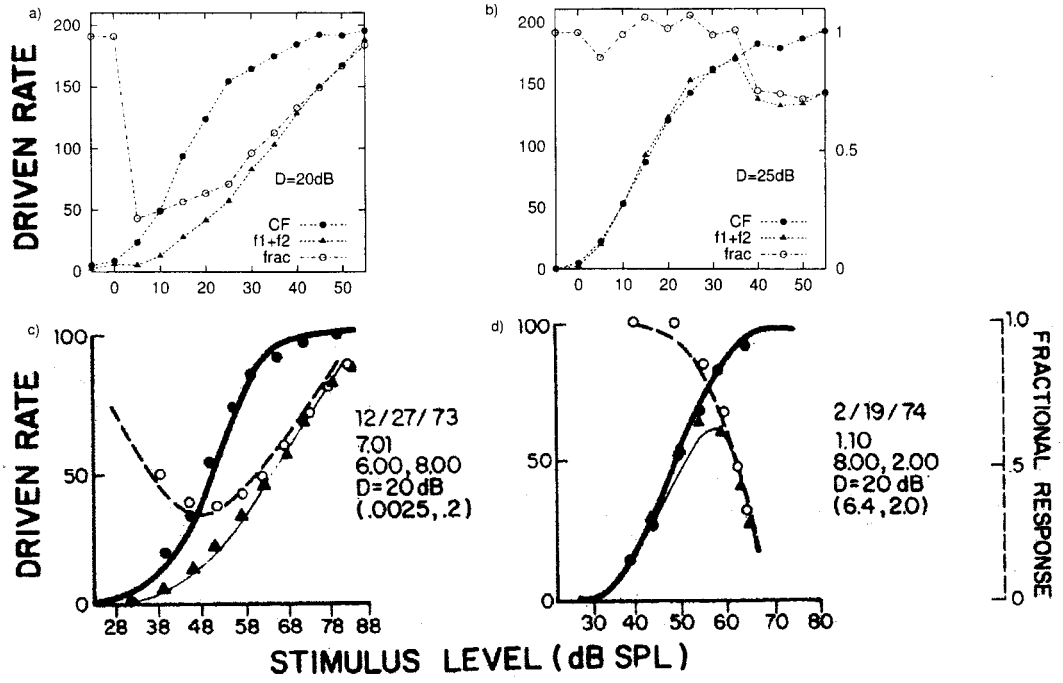


FIG. 14. Different properties of suppression for above- and below-CF suppressors. Responses of two model fibers and two AN fibers to tones at CF and two-tone complexes. (a) Model fiber with CF=6 kHz. Suppressor frequency is 8 kHz, and suppressor level is 25 dB SPL above that of the CF tone. (b) Model fiber with CF=8 kHz. Suppressor frequency is 2 kHz, and suppressor level is 20 dB SPL above that of the CF tone. (c) AN fiber from Sachs and Abbas (1976, their Fig. 4) with the same CF and suppressor parameters as the model fiber in panel a. (d) AN fiber from Sachs and Abbas (1976; with permission) with the same CF and suppressor parameters as the model fiber in panel b. Driven rate is the sustained rate (computed over the 25–45 ms window of the 60 ms duration stimulus; 500 repetitions were used) minus spontaneous rate. The fractional response (open circles) is the ratio of the driven rates in response to the CF tone and to the two-tone complex.

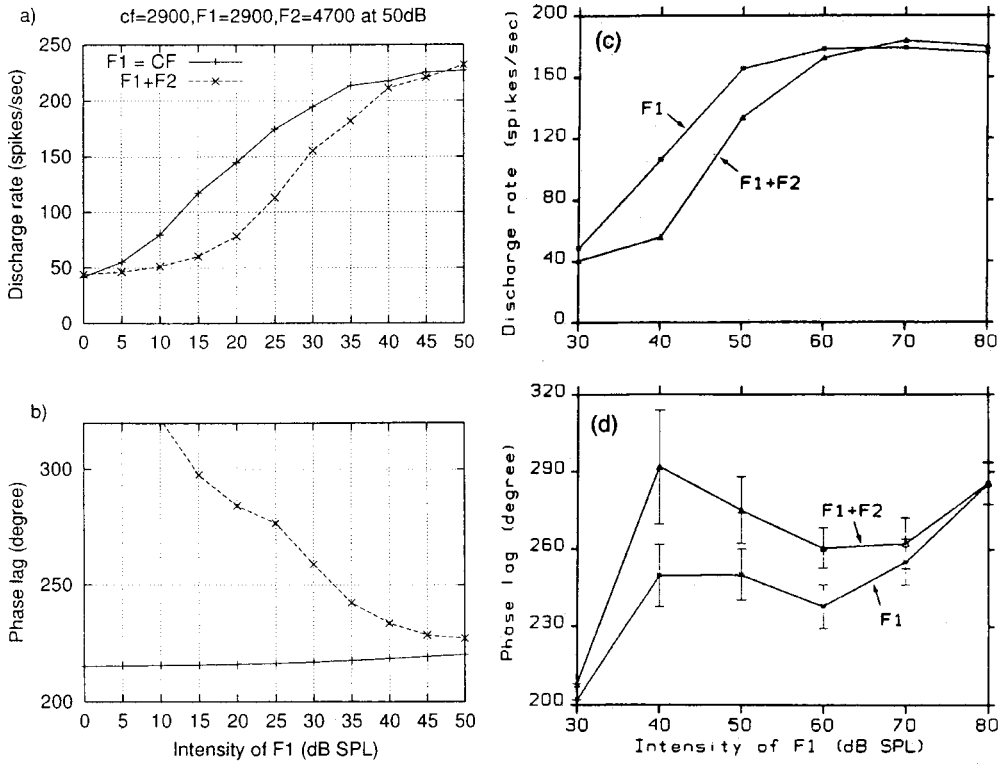


FIG. 15. Effect of a suppressor on the rate and phase of the response to a tone at CF. (a) Rate-level functions for response of model fiber with CF equal to 2900 Hz. $F1$ is at the CF, and $F2$ is 4700 Hz. The level of the tone at $F2$ is 50 dB SPL. (b) The phase of the phase-locked response to $F1$ is shown as a function of the level of the $F1$ tone. Responses are shown for responses to $F1$ alone, and for the phase of the $F1$ component of the response to $F1+F2$. (c) Rate, and (d) phase for a comparable example from Deng and Geisler (1985; with permission). The level of the suppressor in their figure was 90 dB.

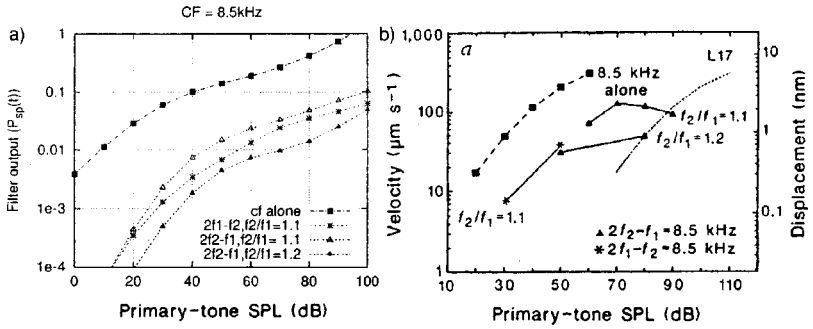


FIG. 16. (a) Amplitude of the distortion products (or combination tones) in the model response as a function of level for equal-intensity two-tone complexes. The frequencies of the primary tones were chosen so that the frequency of the distortion product (i.e., $2f1-f2$, or $2f2-f1$) was matched to the model CF (8.5 kHz). The level-dependent trends of the distortion products are similar to those measured on the basilar membrane (b) (Robles *et al.*, 1991; with permission). The amplitudes of the model distortion products are not as large as in the data, especially for the $2f1-f2$ distortion product. The low-amplitude model responses are consistent with the lack of a propagation medium in this filterbank model; the gain that would be experienced by a component as it travelled to the best place along the basilar membrane is not included here (see Goldstein, 1995).

associated amplification are not included in the model (cf. Goldstein, 1995).

Arbitrary input waveforms can be used as inputs to the model, which allows the study of the detailed temporal and statistical response properties for complex sounds. Figure 17 illustrates the temporal responses of a model fiber with a CF of 1 kHz to two broadband stimuli, clicks and wide-band noises. Both responses show periodicities in the PST histograms that are dominated by the CF of the model fiber. The representation in the model response of the envelope of the complex sound is influenced by the peripheral filtering. Representation of envelope properties in the model responses are qualitatively appropriate. A quantitative analysis of the envelope response properties of the model is not addressed here.

The empirical statistics of AN-fiber discharge counts (e.g., mean and variance) are important for evaluating potential coding schemes in the auditory nerve. Several reports (e.g., Young and Barta, 1986; Winter and Palmer, 1991) show that the empirical discharge-count statistics differ from the behavior of a Poisson process. Differences from Poisson statistics are partly due to the effects of refractoriness (Teich and Lachs, 1979). Figure 18(a) shows the mean and standard deviation of sustained-rate responses of a 1 kHz CF model

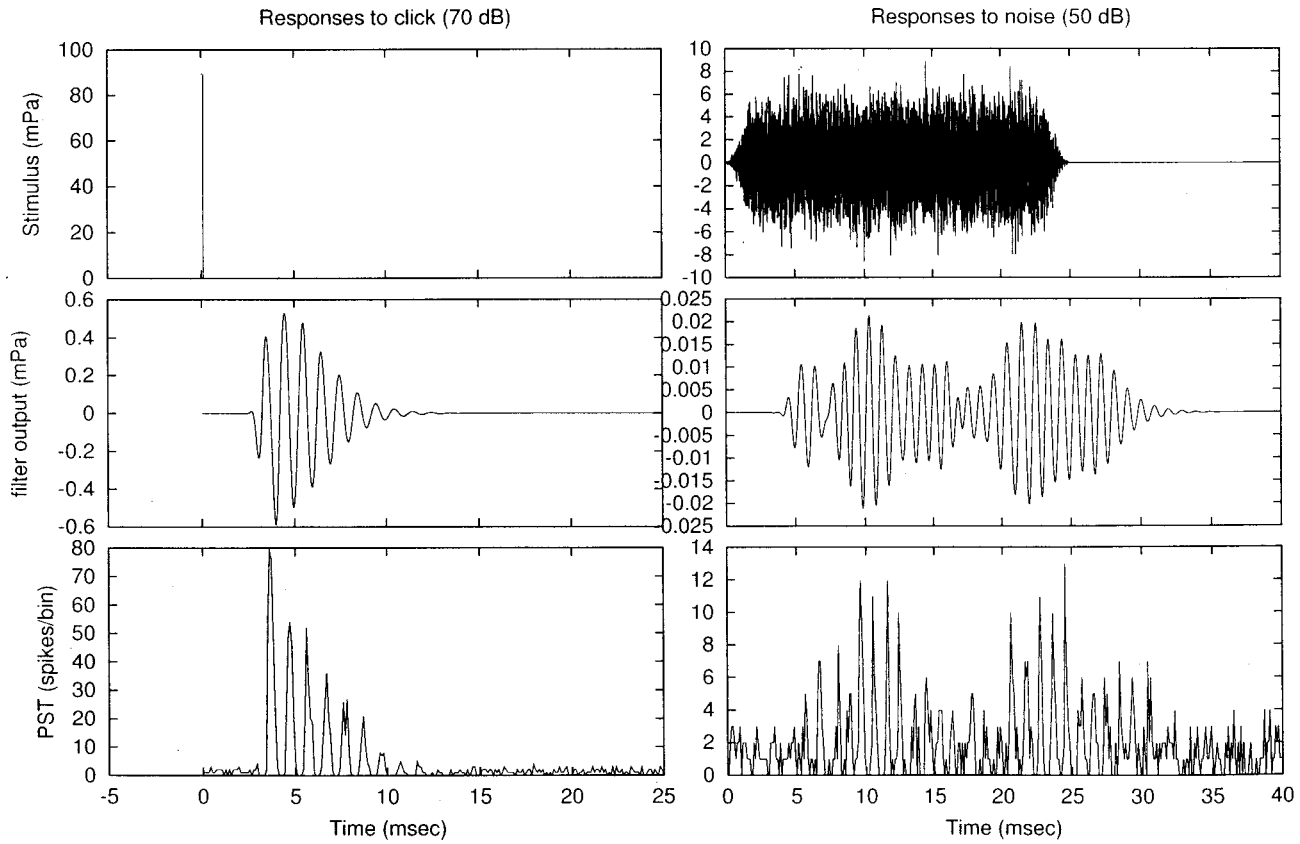


FIG. 17. Model fiber responses for CF of 1 kHz to click and noise waveforms. The top panels show the stimulus waveforms, middle panels show the response of the signal-path filter, and lower panels show PST histograms of response to 500 repetitions. The click stimulus is 200 μs long and 70 dB SPL (peak equivalent). The Gaussian noise stimulus has a bandwidth of 10 kHz and was presented at 50 dB SPL rms.

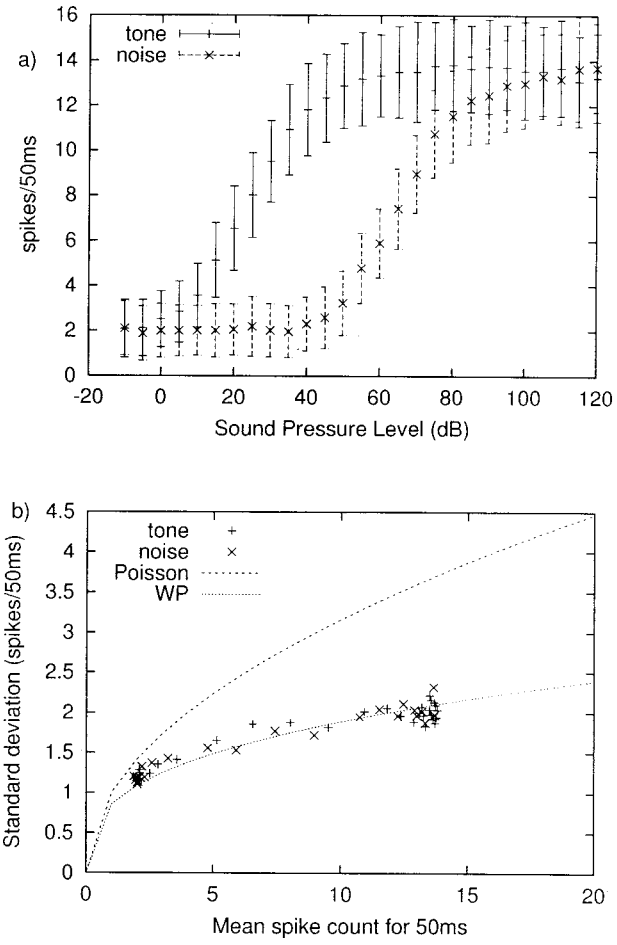


FIG. 18. (a) Rate-level functions (mean and standard deviation) for responses to tones at CF and to wide-band noises for a model fiber with a CF of 1 kHz. Discharges in the 1.25–51.25 ms time window were counted for 200 repetitions of each stimulus. The Gaussian noise stimulus has a bandwidth of 10 kHz. The responses to the noise are shifted to the right (with respect to the tone responses) because they are plotted as a function of dB SPL rms. (b) Standard deviation versus mean discharge counts for tone and noise responses. The dotted line is the fit to Young and Barta's (1986) data used by Winter and Palmer (1991), and the dashed line is the standard deviation for a Poisson process.

fiber to a CF tone and to a wide-band noise as a function of level. The model shows a wider dynamic range for noises than for tones (Schalk and Sachs, 1980; May and Sachs, 1992). Figure 18(b) shows the standard deviation of the model-response rate as a function of the mean discharge count for both CF tones (crosses) and noise (plusses). The dotted line is a fit (from Winter and Palmer, 1991) to the experimental data of Young and Barta (1986), and the dashed line is the standard deviation expected for a Poisson process. The statistical responses of the model fiber are consistent with physiological responses as summarized by the dotted line.

The revcor function is a useful estimate of the AN fiber's impulse response based on responses to wide-band noise (de Boer and de Jongh, 1978). The revcor filters estimated from model noise responses at different levels are illustrated in Fig. 19(a) and show changes in bandwidth as a function of the input level that are comparable to actual AN revcor filters [Fig. 19(b)] (Carney and Yin, 1988). The changes in bandwidth of the revcor filters are relatively small for low-CF AN fibers (cf. Evans, 1977; de Boer and de Jongh, 1978; Carney and Yin, 1988) due to the relatively small gain of the cochlear amplifier at low CFs. However, the level-dependent changes in phase near the CF (Fig. 19) associated with these changes in bandwidth represent significant level-dependent temporal shifts of the AN discharge times at low frequencies that have been hypothesized to be important for level encoding (Carney, 1994). Recent studies have quantified the information contained in these nonlinear temporal response properties and have shown that they contribute significant information for level encoding that is not included in the average-rate responses of the majority of AN fibers (Heinz *et al.*, 1999, 2001; Colburn *et al.*, 2001).

IV. DISCUSSION

This report describes a phenomenological model for AN responses that includes the major nonlinearities associated with the cochlear amplifier. The model has compressive gain in response to tones at CF, nonlinear bandwidth and phase, and two-tone suppression. The model has asymmetric growth of two-tone suppression, i.e., suppression-growth is greater for tones below CF than for tones above CF. This property has been difficult to incorporate into previous models that predict responses to arbitrary stimuli (Delgutte, 1990). Other aspects of the model discharge properties for high-spontaneous-rate fibers are also appropriate, including saturating rate-level functions, the drop in synchrony as a function of CF, and the statistical nature of discharge counts across a range of discharge rates. The responses of this model to a wide range of simple and complex stimuli provide reasonably accurate representations of the discharge rates and discharge times of AN fibers across a wide range of CFs and levels.

Physiological data show that the active process of the basilar membrane is affected by a wide region of the basilar membrane (Delgutte, 1990; Zhao and Santos-Sacchi, 1999). It has also been shown that tones at frequencies above and below CF have different effects on nonlinear tuning (Delgutte, 1990; Ruggero *et al.*, 1992). Experimental data on IHCs and AN suppression tuning curves also suggest that the OHCs responsible for enhancing vibration of a group of IHCs are located basal to the IHCs (Patuzzi, 1996). The current model's time-varying wide-band filter in the control path, which is shifted slightly basal to the AN-fiber CF, is used to represent the effect of the local and neighboring OHCs on responses of the signal-path filter.

Several AN response properties were not included in this model and are challenges for future modeling studies. For example, the model does not include tails of tuning curves (Kiang and Moxon, 1974; Liberman and Kiang, 1978), the glide in instantaneous frequency in the AN impulse response that is related to the level-dependent shift in best frequency (Carney *et al.*, 1999), the dramatic changes in rate and phase in response to tones (i.e., simple-tone interference) and clicks at high sound levels (Kiang, 1990; Ruggero *et al.*, 1996; Lin and Guinan, 2000), the effects of middle- and external-ear acoustics (Rosowski, 1996), the effects of efferents on the rate and timing of AN discharges (Wiederhold, 1986; Guinan, 1996), and low and medium

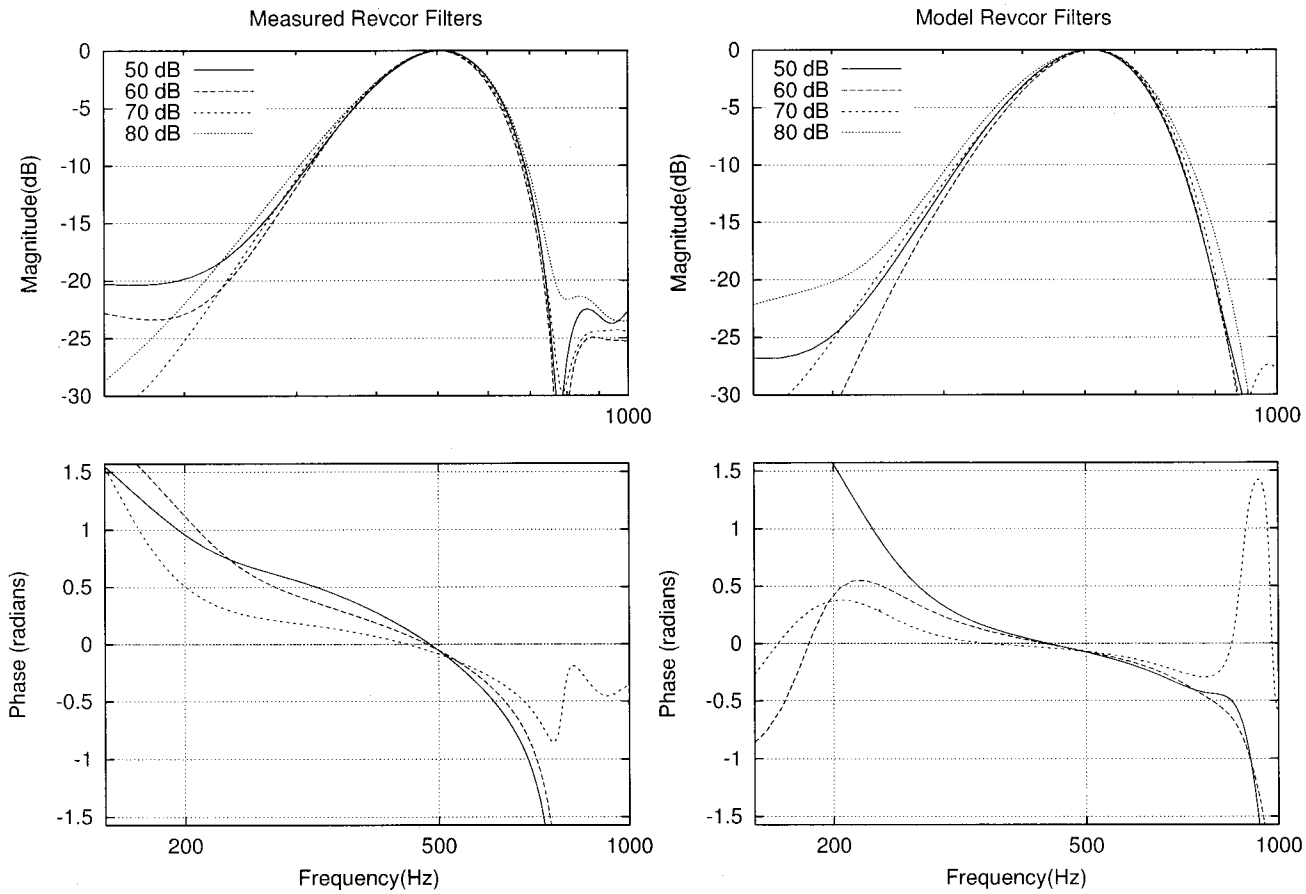


FIG. 19. (a) Spectral magnitude and phase of the revcor filters of a model fiber with CF of 510 Hz measured from responses to wide-band noise presented at several levels. The functions were calculated by averaging 15 ms of the stimulus prior to each discharge time that resulted from a 10 s noise stimulus presented for five repetitions. A Hamming window was used to calculate the frequency responses. The amplitudes of the filter responses were normalized to 0 dB at CF (Carney, 1993; Carney and Yin, 1988). Noise levels are given in dB SPL rms. (b) Examples of revcor functions at several levels for actual cat AN fiber with CF equal to 510 Hz (Unit ID: U86166-39, recorded as part of a previous study; Carney and Yin, 1988).

spontaneous rates (Liberman, 1978). In addition, as discussed above, the filterbank nature of this phenomenological model precludes explanation of phenomena that depend upon propagation of signals along the basilar membrane, such as the overall phase properties and propagation of combination tones to the position along the basilar membrane tuned to the distortion tones (see Goldstein, 1990, for a description of this limitation).

Inclusion of the tails of the tuning curves may play an important role in understanding population responses to moderate- to high-level broadband stimuli, such as speech sounds, in which significant energy at low frequencies may influence responses of neurons over a wide range of CFs. Also, the glide in the instantaneous frequency of the impulse responses will influence both the rate and timing of discharges for most AN fibers over a wide range of stimulus levels. Because this property interacts with the nonlinear tuning to produce shifts in frequency tuning with sound level (Recio *et al.*, 1996; Carney *et al.*, 1999), this property is important to explore in future models that include nonlinear tuning.

The model described here represents a significant extension of our previous modeling efforts (Carney, 1993) in that the nonlinear tuning is much more accurate without a significant increase in the complexity of the model, and the model has been extended to high CFs. The gain and bandwidth of each AN fiber were influenced by a broad frequency range surrounding the fiber's CF, which resulted in more accurate two-tone suppression. One aspect of the two-tone suppression property that was not accurately captured in this model was the degree of asymmetry in the growth of suppression with suppressor level. The asymmetry of the model is similar to that described for low-CF fibers, but not as strong as reported for high-CF fibers (Delgutte, 1990). A future model that includes CF-dependent shifts in frequency with level, in both the signal path and the control path, might be able to accurately simulate this aspect of two-tone suppression. Inclusion of the CF-dependent glide in the instantaneous frequency of the impulse responses (Carney *et al.*, 1999) may be sufficient to make this aspect of two-tone suppression more accurate. In addition, interactions between mechanisms for two-tone interaction and tail mechanisms in future models would be expected to influence the asymmetry of growth of suppression.

Extension of this model to include low- and medium-spontaneous-rate fibers is motivated by the need for accurate descriptions of these responses to examine their potential role in level encoding of pure tones, as well as complex sounds. It is also important to explore the different spontaneous-rate groups in the context of a nonlinear model,

since the slopes and dynamic ranges of their rate-level functions are strongly influenced by the cochlear amplifier (e.g., Yates *et al.*, 1992). In addition, the ability of this model to simulate responses over a wide range of CFs is important to the further study of models for AN fibers with different spontaneous rates. The change in the amount of compression (or equivalently, the change in the gain of the cochlear amplifier) as a function of CF has strong implications for the roles of the different spontaneous-rate fibers in encoding stimulus features at different CFs (Heinz *et al.*, 1999, 2001). A nonlinear model for AN responses over a wide range of CFs and spontaneous rates will allow quantification of the information encoded by these fibers (e.g., Colburn *et al.*, 2001).

Another direction for extension of this model is to combine the nonlinear filterbank approach used here with some aspects of the parallel signal-processing pathways of the MBPNL approach. Such an effort might allow explanation of a wider range of phenomena than can be explained by either model independently. The tails of tuning curves (Goldstein, 1990), the multiple modes of excitation that are suggested by Lin and Guinan's (2000), and the level-dependent phase relationship between tip and tail phenomena suggested by these and other AN responses (e.g., Goblick and Pfeiffer, 1969; Lin and Goldstein, 1995) are examples of phenomena that could be approached by combining aspects of these two modeling approaches. The present modeling approach provides a useful description of both rate and temporal (including nonlinear phase) response properties from excitation modes associated with the active cochlear process, while the MBPNL approach provides a valuable framework for modeling response properties associated with the combination of multiple modes (either active or passive).

ACKNOWLEDGMENTS

We acknowledge the helpful comments of Susan Moscynski on earlier versions of this manuscript, and of Susan Early on later versions. Dr. Roger Miller generously provided the Q10 data in Fig. 3. Dr. Don Johnson generously provided the synchronization coefficients illustrated in Fig. 12. Dr. Mario Ruggero generously provided the basilar membrane phase data plotted in Fig. 11. Dr. Philip Joris provided unpublished plots of synchronization coefficient versus SPL for a number of mid-frequency AN fibers, which were critical for modification of the IHC model. Dr. Ray Meddis provided helpful comments in a review of the manuscript, and Dr. Julius Goldstein generously provided detailed information on the MBPNL model. This work was supported by the Office of Naval Research Grant Agmt No. Z883402 (XZ, MGH, LHC), NIH T32DC00038 (MGH), NIDCD DC00109 (ICB), and NIDCD DC00023(ICB).

APPENDIX

The IHC-AN synapse model is a time-varying implementation of Westerman and Smith's (1988) three-store diffusion model that allows simulations of responses to arbitrary stimuli. The three stores are the immediate (I), local (L), and global (G) stores. The concentration (C) at each store varies as the ''substance'' to be released diffuses; the concentration differences and the permeability (P) at each boundary regulate the rate of diffusion.

The local and global permeabilities and the three volumes are constant values; these values are specified in terms of time constants for diffusion and response amplitudes that can be estimated from the literature and adjusted based on the shape of the model's PST histogram. The desired response properties, based on PST histograms in response to tones, are used to set the parameter values for the steady-state discharge rate, A_{SS}; spontaneous rate, *spont*; rapid time constant for adaptation, τ_R; short-term time constant for adaptation, τ_{ST}; the ratio of rapid response amplitude to short-term response amplitude, $A_{R/ST}$; the maximum immediate permeability, $P_{I\,\max}$; and the peak-to-sustained response amplitude ratio, PTS.

Several intermediate parameters can then be determined from those above. The onset rate, rapid and short-term response amplitudes, resting permeability and global concentration are computed as follows:

$$A_{\mathrm{ON}}=\mathrm{PTS}\,A_{SS}, \tag{A1}$$

$$A_R=(A_{\mathrm{ON}}-A_{SS})\frac{A_{R/ST}}{(1+A_{R/ST})}, \tag{A2}$$

$$A_{ST}=A_{\mathrm{ON}}-A_{SS}-A_R, \tag{A3}$$

$$P_{\mathrm{rest}}=P_{I\,\max}\frac{spont}{A_{\mathrm{ON}}}, \tag{A4}$$

$$C_G=\frac{spont(A_{\mathrm{ON}}-spont)}{A_{\mathrm{ON}}P_{\mathrm{rest}}(1-spont/A_{SS})}. \tag{A5}$$

The volumes of the three stores are then computed; first several intermediate parameters are defined as (see Westerman and Smith, 1988):

$$\gamma_1=C_G/spont,\quad \gamma_2=C_G/A_{SS},$$
$$\kappa_1=-1/\tau_R,\quad \kappa_2=-1/\tau_{ST}. \tag{A6}$$

Following the strategy of Westerman and Smith (1988), the immediate volume is computed as:

$$V_{I0}=\frac{1-P_{I\,\max}/P_{\mathrm{rest}}}{\gamma_1\{[A_R(\kappa_1-\kappa_2)/(C_GP_{I\,\max})]+\kappa_2/(P_{\mathrm{rest}}\gamma_1)-\kappa_2/(P_{I\,\max}\gamma_2)\}}, \tag{A7}$$

$$V_{I1}=\frac{1-P_{I\,\max}/P_{\mathrm{rest}}}{\gamma_1\{[A_{ST}(\kappa_2-\kappa_1)/(C_GP_{I\,\max})]+\kappa_1/(P_{\mathrm{rest}}\gamma_1)-\kappa_1/(P_{I\,\max}\gamma_2)\}}, \tag{A8}$$

$$V_I=\frac{V_{I0}+V_{I1}}{2}. \tag{A9}$$

Several other intermediate parameters are useful for the derivation of the other volumes and permeabilities:

$$\alpha = \frac{\gamma_2}{\kappa_1 \kappa_2}, \quad \beta = -(\kappa_1 + \kappa_2)\alpha, \quad \theta_1 = \alpha P_{I\,\max}/V_I,$$
$$\theta_2 = V_I/P_{I\,\max}, \quad \theta_3 = \gamma_2 - 1/P_{I\,\max}. \tag{A10}$$

Then, the local and global permeabilities can be specified as:

$$P_L = \left(\frac{\beta - \theta_2\theta_3}{\theta_1} - 1\right) P_{I\,\max}, \tag{A11}$$

$$P_G = \frac{1}{\theta_3 - 1/P_L}. \tag{A12}$$

The local volume and concentration, and the resting value of the immediate concentration, are set as follows:

$$V_L = \theta_1 P_L P_G, \tag{A13}$$

$$C_{I\,\mathrm{rest}} = spont/P_{\mathrm{rest}}, \tag{A14}$$

$$C_{L\,\mathrm{rest}} = C_{I\,\mathrm{rest}} \frac{(P_{\mathrm{rest}} + P_L)}{P_L}. \tag{A15}$$

The voltage-dependent immediate permeability, P_I varies with the IHC-model voltage, V_{IHC}, through a soft-rectifier specified as:

$$P_I(t) = \frac{P_{SL}}{P_{ST}} \log(1 + e^{P_{ST} V_{\mathrm{IHC}}(t)})$$
$$\text{where} \quad P_{SL} = P_{\mathrm{rest}} P_{ST}/\log(2)$$
$$\text{and} \quad P_{ST} = \log(2^{V_{\mathrm{sat}}/P_{\mathrm{rest}}} - 1), \tag{A16}$$

where P_{rest} is the resting value of the immediate permeability, and V_{sat} sets the saturation voltage for the soft rectifier. The value of V_{sat} varies with CF, in order to maintain the threshold of the model fibers at about 0 dB SPL at CF:

$$V_{\mathrm{sat}}(\mathrm{CF}) = 18.54\, P_{I\,\max} K_{\mathrm{CF}}$$
$$\text{where} \quad K_{\mathrm{CF}} = \max[1.5, 2 + 3\log_{10}(\mathrm{CF}/1000)]. \tag{A17}$$

Note that Eqs. (17) and (18) in the text provide simple numerical descriptions for the IHC-synapse model expressions shown above; p_1 in the text is equivalent to P_{SL}/P_{ST} and p_2 is equal to P_{ST}.

The difference equations for the concentrations of the stores in the diffusion model are as follows:

$$C_I(kT+T) = C_I(kT) + \frac{T}{V_I}\{[-P_I(kT)C_I(kT)] + P_L[C_L(kT) - C_I(kT)]\}, \tag{A18}$$

$$C_L(kT+T) = C_L(kT) + \frac{T}{V_L}\{-P_L[C_L(kT) - C_I(kT)] + P_G[C_G - C_L(kT)]\}, \tag{A19}$$

where T is the sampling time used in the simulation. The initial values are set as follows: $C_I(0) = C_{I\,\mathrm{rest}}$, $C_L(0) = C_{L\,\mathrm{rest}}$, and $P_I(0) = P_{\mathrm{rest}}$. Finally, the output of the synapse model is:

$$S(kT) = C_I(kT)\, P_I(kT). \tag{A20}$$

[1]The code will be made available on the Earlab website at Boston University: http://earlab.bu.edu

[2]The order of this low-pass filter influences the strength of harmonic distortions in the model responses. For filter orders less than three, the harmonic distortions seen in the tuning curves and response areas are much stronger than those observed in reports of these measures in the literature of AN responses (e.g., Kiang *et al.*, 1965; Anderson *et al.*, 1971; Rose *et al.*, 1971; Liberman, 1978).

Anderson, D. J., Rose, J. E., Hind, J. E., and Brugge, J. F. (**1971**). ''Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: Frequency and intensity effects,'' J. Acoust. Soc. Am. **49**, 1131–1139.

Arthur, R. M., Pfeiffer, R. R., and Suga, N. (**1971**). ''Properties of two-tone inhibition in primary auditory neurones,'' J. Physiol. **212**, 593–609.

Carney, L. H. (**1990**). ''Sensitivities of cells in the anteroventral cochlear nucleus of cat to spatio-temporal discharge patterns across primary afferents,'' J. Neurophysiol. **64**, 437–456.

Carney, L. H. (**1993**). ''A model for the responses of low-frequency auditory-nerve fibers in cat,'' J. Acoust. Soc. Am. **93**, 401–417.

Carney, L. H. (**1994**). ''Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness,'' Hear. Res. **76**, 31–44.

Carney, L. H., and Yin, T. C. T. (**1988**). ''Temporal coding of resonances by low-frequency auditory nerve fibers: Single-fiber responses and a population model,'' J. Neurophysiol. **60**, 1653–1677.

Carney, L. H., McDuffy, M. J., and Shekhter, I. (**1999**). ''Frequency glides in the impulse responses of auditory-nerve fibers,'' J. Acoust. Soc. Am. **105**, 2384–2391.

Cheatham, M. A., and Dallos, P. (**1989**). ''Two-tone suppression in inner hair cell responses,'' Hear. Res. **40**, 187–196.

Cheatham, M. A., and Dallos, P. (**1990**). ''Comparisons of low- and high-side two-tone suppression in inner hair cell and organ of corti responses,'' Hear. Res. **50**, 193–210.

Cheatham, M. A., and Dallos, P. (**1992**). ''Two-tone suppression in inner hair cell responses: Correlates of rate suppression in the auditory nerve,'' Hear. Res. **60**, 1–12.

Cheatham, M. A., and Dallos, P. (**1993**). ''Longitudinal comparisons of IHC ac and dc receptor potentials recorded from the guinea pig cochlea,'' Hear. Res. **68**, 107–114.

Cheatham, M. A., and Dallos, P. (**1998**). ''The level dependence of response phase: Observations from cochlear hair cells,'' J. Acoust. Soc. Am. **104**, 356–369.

Cheatham, M. A., and Dallos, P. (**1999**). ''Response phase: A view from the inner hair cell,'' J. Acoust. Soc. Am. **105**, 799–810.

Colburn, H. S., Carney, L. H., and Heinz, M. G. (**2001**). ''Quantifying the information in auditory-nerve responses for level discrimination,'' J. Assoc. Res. Otolaryngol. (submitted).

Cooper, N. P. (**1996**). ''Two-tone suppression in cochlear mechanics,'' J. Acoust. Soc. Am. **99**, 3087–3098.

Cooper, N. P., and Rhode, W. S. (**1996**). ''Two-tone suppression in apical cochlear mechanics,'' Aud. Neurosci. **3**, 123–134.

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

Costalupes, J. A., Rich, N. C., and Ruggero, M. A. (**1987**). ''Effects of excitatory and non-excitatory suppressor tones on two-tone rate suppression in auditory nerve fibers,'' Hear. Res. **26**, 155–164.

Dallos, P. (**1985**). ''Response characteristics of mammalian cochlear hair cells,'' J. Neurosci. **5**, 1591–1608.

de Boer, E. (**1975**). ''Synthetic whole-nerve action potentials for the cat,'' J. Acoust. Soc. Am. **58**, 1030–1045.

de Boer, E., and Krudenier, C. (**1990**). ''On ringing limits of the auditory periphery,'' Biol. Cybern. **63**, 433–442.

de Boer, E., and de Jongh, H. R. (**1978**). ''On cochlear encoding: Potentialities and limitations of the reverse correlation technique,'' J. Acoust. Soc. Am. **63**, 115–135.

de Boer, E., and Nuttall, A. L. (**1997**). ''The mechanical waveform of the basilar membrane. I. Frequency modulations (''glides'') in impulse responses and cross-correlation functions,'' J. Acoust. Soc. Am. **101**, 3583–3592.

Delgutte, B. (**1990**). ''Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level,'' Hear. Res. **49**, 225–246.

Deng, L., and Geisler, C. D. (**1985**). ''Changes in phase of excitor-tone

responses in cat auditory-nerve fibers by suppressor tones and fatigue,'' J. Acoust. Soc. Am. **78**, 1633–1643.
Deng, L., and Geisler, C. D. (**1987**). ''A composite model for processing speech sounds,'' J. Acoust. Soc. Am. **82**, 2001–2012.
Duifhuis, H. (**1976**). ''Cochlear nonlinearity and second filter: Possible mechanism and implications,'' J. Acoust. Soc. Am. **67**, 914–927.
Evans, E. F. (**1977**). ''Frequency selectivity at high signal levels of single units in cochlear nerve and nucleus,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, New York), pp. 185–192.
Geisler, C. D. (**1990**). ''Evidence for expansive power functions in the generation of the discharges of 'low- and medium-spontaneous' auditory-nerve fibers,'' Hear. Res. **44**, 1–12.
Geisler, C. D., and Rhode, W. D. (**1982**). ''The phases of basilar membrane vibrations,'' J. Acoust. Soc. Am. **71**, 1201–1203.
Geisler, C. D., and Sinex, D. G. (**1980**). ''Responses of primary auditory fibers to combined noise and tonal stimuli,'' Hear. Res. **3**, 317–334.
Giguere, C., and Woodland, P. C. (**1994**). ''A computational model of the auditory periphery for speech and hearing research. I. Ascending path,'' J. Acoust. Soc. Am. **95**, 331–342.
Goblick, Jr., T. J., and Pfeiffer, R. R. (**1969**). ''Time-domain measurements of cochlear nonlinearities using combination click stimuli,'' J. Acoust. Soc. Am. **46**, 924–938.
Goldstein, J. L. (**1990**). ''Modeling rapid waveform compression on the basilar membrane as multiple-bandpass-nonlinearity filtering,'' Hear. Res. **49**, 39–60.
Goldstein, J. L. (**1995**). ''Relations among compression, suppression, and combination tones in mechanical responses of the basilar membrane: Data and MBPNL model,'' Hear. Res. **89**, 52–68.
Guinan, Jr., J. J. (**1996**). ''Physiology of olivocochlear efferents,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 435–502.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**1999**). ''Monaural, cross-frequency coincidence detection as a mechanism for decoding perceptual cues provided by the cochlear amplifier,'' (Abstract) J. Acoust. Soc. Am. **105**, 1023.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001**). ''Monaural cross-frequency coincidence detection for level discrimination: Decoding rate and timing cues associated with the cochlear amplifier,'' J. Acoust. Soc. Am. (submitted).
Hewitt, M. J., and Meddis, R. (**1991**). ''An evaluation of eight computer models of mammalian inner hair-cell function,'' J. Acoust. Soc. Am. **90**, 904–917.
Hicks, M. L., and Bacon, S. P. (**1999**). ''Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing,'' J. Acoust. Soc. Am. **105**, 326–338.
Holley, M. C. (**1996**). ''Outer hair cell motility,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 386–434.
Irino, T., and Patterson, R. D. (**1997**). ''A time-domain level-dependent auditory filter: The gammachirp,'' J. Acoust. Soc. Am. **101**, 412–419.
Javel, E., Geisler, C. D., and Ravindran, A. (**1978**). ''Two-tone suppression in auditory nerve of the cat: Rate-intensity and temporal analyses,'' J. Acoust. Soc. Am. **63**, 1093–1104.
Javel, E., McGee, J., Walsh, E. J., Farley, G. R., and Gorga, M. P. (**1983**). ''Suppression of auditory nerve responses. II. Suppression threshold and growth, iso-suppression contours,'' J. Acoust. Soc. Am. **74**, 801–813.
Jenison, R. L., Greenberg, S., Kluender, K. R., and Rhode, W. S. (**1991**). ''A composite model of the auditory periphery for the processing of speech based on the filter response functions of single auditory-nerve fibers,'' J. Acoust. Soc. Am. **90**, 773–786.
Johannesma, P. I. M. (**1972**). The pre-response stimulus ensemble of neurons in the cochlear nucleus,'' in *Proceedings of the Symposium on Hearing Theory* (IPO, Eindhoven, The Netherlands), pp. 58–69.
Johnson, D. H. (**1980**). ''The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones,'' J. Acoust. Soc. Am. **68**, 1115–1122.
Joris, P. (**1999**) (personal communication).
Kiang, N. Y. S. (**1975**). ''Stimulus representation in the discharge patterns of auditory neurons,'' in *The Nervous System, Vol. 3: Human Communication and Its Disorders*, edited by D. B. Tower (Raven, New York), pp. 81–95.
Kiang, N. Y. S. (**1990**). ''Curious oddments of auditory-nerve studies,'' Hear. Res. **49**, 1–16.
Kiang, N. Y. S., and Moxon, E. C. (**1974**). ''Tails of tuning curves of auditory-nerve fibers,'' J. Acoust. Soc. Am. **55**, 620–630.
Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve*, M.I.T. Research Monograph 35 (MIT Press, Cambridge).
Liberman, M. C. (**1978**). ''Auditory-nerve responses from cats raised in a low-noise chamber,'' J. Acoust. Soc. Am. **63**, 442–455.
Liberman, M. C. (**1982**). ''The cochlear frequency map for the cat: Labelling auditory-nerve fibers of known characteristic frequency,'' J. Acoust. Soc. Am. **72**, 1441–1449.
Liberman, M. C., and Kiang, N. Y. S. (**1978**). ''Acoustic trauma in cats: Cochlear pathology and auditory-nerve activity,'' Acta Oto-Laryngol., Suppl. **358**, 1–63.
Lin, T., and Goldstein, J. L. (**1995**). ''Quantifying 2-factor phase relations in nonlinear responses from low characteristic-frequency auditory-nerve fibers,'' Hear. Res. **90**, 126–138.
Lin, T., and Guinan, Jr., J. J. (**2000**). ''Auditory-nerve-fiber responses to high-level clicks: Interference patterns indicate that excitation is due to the combination of multiple drives,'' J. Acoust. Soc. Am. **107**, 2615–2630.
Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**1998**). ''A revised computational inner hair cell model,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 112–121.
May, B. J., and Sachs, M. B. (**1992**). ''Dynamic range of neural rate responses in the ventral cochlear nucleus of awake cats,'' J. Neurophysiol. **68**, 1589–1602.
Meddis, R. (**1986**). ''Simulation of mechanical to neural transduction in the auditory receptor,'' J. Acoust. Soc. Am. **79**, 702–711.
Meddis, R. (**1988**). ''Simulation of auditory-neural transduction: Further studies,'' J. Acoust. Soc. Am. **83**, 1056–1063.
Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (**1997**). ''Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers,'' J. Acoust. Soc. Am. **101**, 3602–3616.
Mountain, D. C., and Hubbard, A. E. (**1996**). ''Hair cell and auditory nerve,'' in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 121–156.
Narayan, S. S., Temchin, A. N., Recio, A., and Ruggero, M. A. (**1998**). ''Frequency tuning of basilar membrane and auditory nerve fibers in the same cochleae,'' Science **282**, 1882–1884.
Nomoto, M., Suga, N., and Katsuki, Y. (**1964**). ''Discharge pattern and inhibition of primary auditory nerve fibers in the monkey,'' J. Neurophysiol. **27**, 768–787.
Nuttall, A. L., and Dolan, D. F. (**1993**). ''Two-tone suppression of inner hair cell and basilar membrane responses in the guinea-pig,'' J. Acoust. Soc. Am. **93**, 390–400.
Nuttall, A. L., and Dolan, D. F. (**1996**). ''Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig,'' J. Acoust. Soc. Am. **99**, 1556–1564.
Oppenheim, A. V., and Schafer, R. W. (**1975**). *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).
Palmer, A. R., and Russell, I. J. (**1986**). ''Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair cell,'' Hear. Res. **24**, 1–15.
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1988**). ''Implementing a gammatone filter bank,'' SVOS Final Report: The Auditory Filter Bank.
Patuzzi, R. (**1996**). ''Cochlear micromechanics and macromechanics,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 186–257.
Patuzzi, R., and Robertson, D. (**1988**). ''Tuning in the mammalian cochlea,'' Physiol. Rev. **68**, 1009–1082.
Pfeiffer, R. R. (**1970**). ''A model for two-tone inhibition of single cochlear nerve fibers,'' J. Acoust. Soc. Am. **48**, 1373–1378.
Prijs, V. F. (**1989**). ''Lower boundaries of two-tone suppression regions in the guinea pig,'' Hear. Res. **42**, 73–82.
Recio, A., Narayan, S. S., and Ruggero, M. A. (**1996**). ''Wiener-kernel analysis of basilar membrane responses to noise,'' in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 325–331.
Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). ''Basilar-membrane responses to clicks at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **103**, 1972–1989.

Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.
Rhode, W. S. (**1977**). ''Some observations on two-tone interaction measured with the Mossbauer effect,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 27–38.
Rhode, W. S., and Cooper, N. P. (**1993**). ''Two-tone suppression and distortion production on the basilar membrane in the hook region of cat and guinea pig cochleae,'' Hear. Res. **66**, 31–45.
Rhode, W. S., and Cooper, N. P. (**1996**). ''Nonlinear mechanics in the apical turn of the chinchilla,'' Aud. Neurosci. **3**, 101–120.
Robert, A., and Eriksson, J. L. (**1999**). ''A composite model of the auditory periphery for simulating responses to complex sounds,'' J. Acoust. Soc. Am. **106**, 1852–1864.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1991**). ''Two-tone distortion in the basilar membrane of the cochlea,'' Nature (London) **349**, 413–414.
Rose, J. E., Hind, J. E., Anderson, D. J., and Brugge, J. F. (**1971**). ''Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey,'' J. Neurophysiol. **34**, 685–699.
Rosowski, J. J. (**1996**). ''Models of External- and Middle-Ear Function,'' in *Auditory Computation (Springer Handbook of Auditory Research, Vol. 6)*, edited by H. L. Hawkins, T. A. McMullen, and A. N. Popper (Springer, New York), pp. 15–61.
Ruggero, M. A., and Rich, N. C. (**1987**). ''Timing of spike initiation in cochlear afferents: Dependence on site of innervation,'' J. Neurophysiol. **58**, 379–403.
Ruggero, M. A., Rich, N. C., Recio, A. (**1996**). ''The effects of intense acoustic stimulation on basilar-membrane vibrations,'' Aud. Neurosci. **2**, 329–346.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Ruggero, M. A., Robles, L., and Rich, N. C. (**1992**). ''Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression,'' J. Neurophysiol. **68**, 1087–1099.
Russell, I. J., Cody, A. R., and Richardson, G. P. (**1986**). ''The responses of inner and outer hair cells in the basal turn of the guinea-pig cochlea and in the mouse cochlea grown in vitro,'' Hear. Res. **22**, 199–216.
Sachs, M. B., and Abbas, P. J. (**1974**). ''Rate versus level functions for auditory-nerve fibers in cats: Tone-burst stimuli,'' J. Acoust. Soc. Am. **56**, 1835–1847.
Sachs, M. B., and Abbas, P. J. (**1976**). ''Phenomenological model for two-tone suppression,'' J. Acoust. Soc. Am. **60**, 1157–1163.
Sachs, M. B., and Kiang, N. Y. S. (**1968**). ''Two-tone inhibition in auditory-nerve fibers,'' J. Acoust. Soc. Am. **43**, 1120–1128.
Schalk, T. B., and Sachs, M. B. (**1980**). ''Nonlinearities in auditory-nerve fiber responses to bandlimited noise,'' J. Acoust. Soc. Am. **67**, 903–913.
Schoonhoven, R., Prijs, V. F., and Frijns, J. H. M. (**1998**). ''Transmitter release in inner hair cell synapses: A model analysis of spontaneous and driven properties of cochlear nerve fibers,'' Hear. Res. **113**, 247–260.
Schwid, H. A., and Geisler, C. D. (**1982**). ''Multiple reservoir model of neurotransmitter release by a cochlea inner hair cell,'' J. Acoust. Soc. Am. **72**, 1435–1440.
Shera, C. A., and Guinan, Jr., J. J. (**2000**). ''Reflection-emission phase: A test of coherent reflection filtering and a window on cochlear-tuning,'' Abstracts of the 23rd Midwinter Meeting of the Association for Research in Otolaryngology, 157.
Smith, R. L. (**1988**). ''Encoding of sound intensity by auditory neurons,'' in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 243–274.
Teich, M. C., and Lachs, G. (**1979**). ''A neural-counting model incorporating refractoriness and spread of excitation. I. Application to intensity discrimination,'' J. Acoust. Soc. Am. **66**, 1738–1749.
Temchin, A. N., Rich, N. C., and Ruggero, M. A. (**1997**). ''Low-frequency suppression of auditory nerve responses to characteristic frequency tones,'' Hear. Res. **113**, 29–56.
Weiss, T. F., and Rose, C. (**1988**). ''A comparison of synchronization filters in different auditory receptor organs,'' Hear. Res. **33**, 175–180.
Westerman, L. A., and Smith, R. L. (**1985**). ''Rapid adaptation depends on the characteristic frequency of auditory nerve fibers,'' Hear. Res. **17**, 197–198.
Westerman, L. A., and Smith, R. L. (**1988**). ''A diffusion model of the transient response of the cochlear inner hair cell synapse,'' J. Acoust. Soc. Am. **83**, 2266–2276.
Wiederhold, M. L. (**1986**). ''Physiology of the olivocochlear system,'' in *Neurobiology of Hearing: The Cochlea*, edited by R. A. Altschuler, R. P. Robbin, and D. W. Hoffman (Raven, New York), pp. 349–370.
Winter, I. M., and Palmer, A. R. (**1991**). ''Intensity coding in low-frequency auditory-nerve fibers of the guinea pig,'' J. Acoust. Soc. Am. **90**, 1958–1967.
Yates, G. K., Johnstone, B. M., Patuzzi, R. B., and Robertson, D. (**1992**). ''Mechanical preprocessing in the mammalian cochlea,'' Trends Neurosci. **15**, 57–61.
Young, E. D., and Barta, P. E. (**1986**). ''Rate responses of auditory nerve fibers to tones in noise near masked threshold,'' J. Acoust. Soc. Am. **79**, 426–442.
Zagaeski, M., Cody, A. R., Russell, I. J., and Mountain, D. C. (**1994**). ''Transfer characteristic of the inner hair cell synapse: Steady-state analysis,'' J. Acoust. Soc. Am. **95**, 3430–3434.
Zhao, H. B., and Santos-Sacchi, J. (**1999**). ''Auditory collusion and a coupled couple of outer hair cells,'' Nature (London) **399**, 359–362.

Distinguishing cochlear pathophysiology in 4-aminopyridine and furosemide treated ears using a nonlinear systems identification technique

Lin Bian and Mark E. Chertoff[a)]
Hearing and Speech Department, University of Kansas Medical Center, Kansas City, Kansas 66160

(Received 26 January 2000; revised 12 July 2000; accepted 20 November 2000)

To test the adequacy of physiologic indices derived from a third-order polynomial model quantifying cochlear mechano-electric transduction (MET), 24 Mongolian gerbils were exposed to either 250-mM glucose (control), 150-mM 4-aminopyridine (4-AP), or 30-mM furosemide solutions applied to the round window (RW) membrane. The cochlear microphonic (CM) was recorded from the RW in response to 68- and 88-dB SPL Gaussian noise. A nonlinear systems identification technique (NLID) provided the frequency-domain parameters and physiologic indices of the polynomial model of MET. The control group showed no change in both compound action potential (CAP) thresholds and CM. Exposure to 4-AP and furosemide resulted in a similar elevation in CAP thresholds and a reduction in CM. However, the polynomial model of MET showed different changes. The operating point, slope, and symmetry of the MET function, the polynomial model parameters, and related nonlinear coherences differed between the experimental groups. It is concluded that the NLID technique is sensitive and specific to alterations in the cochlear physiology.
 [DOI: 10.1121/1.1340644]

PACS numbers: 43.64.Nf, 43.64.Kc, 43.64.Gz [LHC]

I. INTRODUCTION

The major function of the inner ear is to convert the acoustic energy into electric signal to be sent to the central auditory system. This is accomplished by mechano-electric transduction (MET), in which cochlear partition motion is transformed into an electric signal via deflection of the hair cell stereocilia. Cochlear MET is highly nonlinear (Dallos, 1973) and many hearing disorders in the inner ear can alter the characteristics of MET and its nonlinearity. These alterations may occur either due to ultrastructural changes in the cochlear partition (Saunders *et al.*, 1991), or via modifications of the electric properties of the hair cells (Salt and Konishi, 1986). An electric potential which reflects cochlear mechanics is the cochlear microphonic (CM). The CM is an ac potential that can be measured as a spatial summation of receptor currents passing through outer hair cells (OHCs). These electric currents, carried predominantly by K^+ ions (Corey and Hudspeth, 1979; Zucca *et al.*, 1982), are modulated by the vibration of the cochlear partition, and thus can be used as an index of cochlear function (Patuzzi *et al.*, 1989).

Recently, we have applied a nonlinear systems identification (NLID) technique (Bendat and Piersol, 1986, 1993; Bendat, 1990, 1998) to the CM evoked by random noise to characterize cochlear MET in normal and hearing-impaired gerbils (Chertoff *et al.*, 1996, 1997; Bian and Chertoff, 1998a, 1998b). The NLID analysis procedure provided the parameters of a third-order polynomial model of cochlear MET in the frequency domain

$$y(f)=A_1(f)x+A_2(f)x^2+A_3(f)x^3+\mathrm{dc}, \quad (1)$$

where y is the CM output in volts, x is the acoustic input in pascals, and the ''A''s are the polynomial coefficients. A series of physiologic indices characterizing MET was derived from further mathematical explorations of the polynomial coefficients. These indices included slope, maxima, minima, and saturation sound-pressure levels at maxima (SPL_{max}) and minima (SPL_{min}), operating point (OP), dynamic range (DR), operating range (OR), and symmetry (Fig. 1). In a previous experiment (Bian and Chertoff, 1998b), exposure to an 8-kHz pure tone and round window (RW) application of salicylate resulted in different changes in the polynomial parameters and physiologic indices, even though the induced hearing losses were similar. These changes in the physiologic indices were in agreement with the results of Patuzzi and Moleirinho (1998), who demonstrated that a Boltzmann function of MET modified its parameters following either acoustic trauma or perfusion of salicylate through scala tympani (ST). The results of our previous studies indicated that the physiologic indices were both sensitive and specific to alterations in cochlear physiology.

To further test the adequacy of the physiologic indices in the differential diagnosis of cochlear hearing losses, it is necessary to apply them to other pathologic conditions. Since our previous results suggested that damaging the transduction channels in the OHC stereocilia by noise exposure reduced the CM and altered the physiologic indices of MET, could other pathologies that block the K^+ currents be distinguished by the polynomial model of MET? One candidate for the K^+-channel blockers is 4-aminopyridine (4-AP) (Hille, 1992), which has been widely used to study the K^+ currents of cochlear hair cells (Kros and Crawford, 1990; Mammano and Ashmore, 1996; Nenov *et al.*, 1997). Another agent that can reduce the ion currents passing through the OHCs is ototoxic drug furosemide, which blocks the K^+

a)Electronic mail: mchertof@kumc.edu

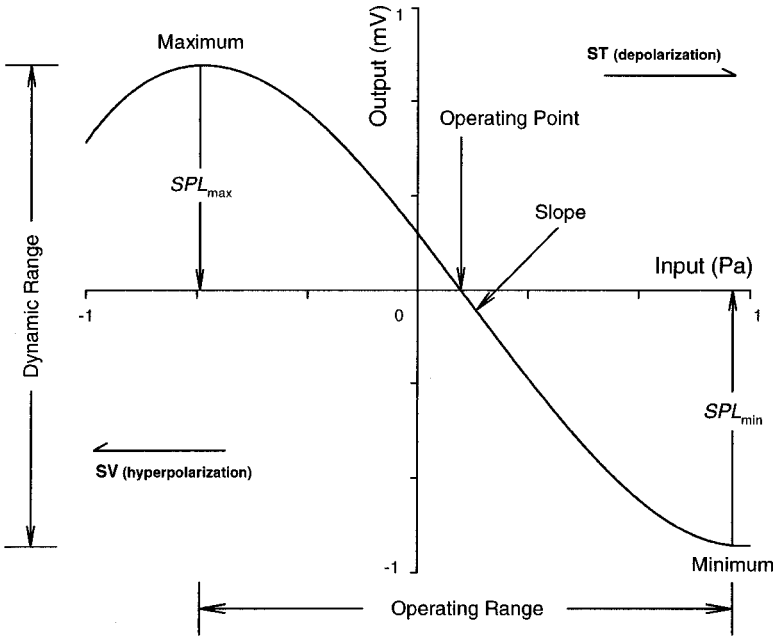


FIG. 1. Illustration of the MET transduction curve and the physiologic indices. SV and ST represent the directions of the input sound pressure towards scala vestibuli and scala tympani.

secretion from the stria vascularis (Brown and Feldman, 1978; Pike and Bosher, 1980). In our previous study (Bian and Chertoff, 1998b), the CM was increased by salicylate and decreased by pure-tone exposure. However, both furosemide and 4-AP should reduce the receptor currents; thus, distinguishing between the pathophysiologies induced by these two drugs should be a more challenging task. The purpose of this study was to determine if the parameters of the polynomial model of MET and associated physiologic indices could distinguish between 4-AP and furosemide ototoxicity.

II. METHODS

A. Experimental design

Mongolian gerbils (*Meriones unguiculatus*) were used as experimental subjects in this study. The ''Animal Welfare Act'' of the U.S. government, and the ''Guide for the Care and Use of Laboratory Animals'' of the National Institute of Health (NIH) were followed in the experiments. A total of 24 animals, equal number of male and female, weighing between 50 and 71 grams, was randomly assigned to a control group and two experimental groups. The number of subjects per group ($n=8$) was determined by performing a statistical power analysis on the normative data that have been accumulated from 96 gerbils in our laboratory (Chertoff *et al.*, 1996, 1997; Bian and Chertoff, 1998b). The experimental groups received either a 4-AP, or a furosemide solution applied on the RW. The control animals received a sham solution of D-glucose which has a formula weight (FW) between 4-AP and furosemide, but no ototoxic effect.

Since all the animals were to receive chemical solutions as treatments, we adopted a ''double-blind '' design in the present study to minimize the bias of the experimenter. Three solutions were made prior to each experiment, from which one solution was randomly selected and coded by a computer program. Then, the selected solution was given to the experimenter for the treatment of the animal. After completion of the data collection and NLID analysis on all the subjects, the codes were broken for further analysis. To increase the statistical power and reduce the variability in the data, a within-subject repeated-measure design was employed. For each animal, the CM responses to Gaussian noise were recorded three times, two recordings before and one after the treatment.

B. Drug preparation

A preliminary study showed that applying either 4-AP or furosemide onto the RW for 10 min produced an acute high-frequency compound action potential (CAP) threshold shift. Pharmacodynamic data demonstrated that both 150-mM 4-AP and 30-mM furosemide could achieve a similar high-frequency threshold shift. The audiometric configuration and the extent of the hearing loss were also similar to animals exposed to a 110-dB SPL 8-kHz pure tone for 10 min or a 100-mM salicylate solution (Bian and Chertoff, 1998b). No threshold shift at any frequency of interest was found after the treatment of glucose solution, even at a concentration as high as 600 mM. To equalize the osmolarities among the treatment solutions, a concentration of 250 mM glucose was adopted for the control group.

All the chemicals were dissolved in phosphate buffered saline (PBS) with concentrations predetermined in the pilot study, i.e., 30-mM furosemide (Sigma F-4381, FW=330.7), 150-mM 4-AP (Sigma A-0152, FW=94.12), and 250-mM D-glucose (Sigma G-8270, FW=180.2). In order for furosemide to dissolve fully, a minimal amount of 10 *N* NaOH was added in the solution (AHFS, 1999). Then, the *p*H values of all solutions were adjusted to 7.3 by adding 5 *N* HCl. The osmolarities (5500 vapor pressure osmometer, Vescor) of the solutions were carefully balanced. The mean osmolarity of the furosemide solutions was about 350 mOsmol/kg H_2O, whereas that of 4-AP was 491 ± 16 (s.d.) mOsmol/kg H_2O. To compensate for this difference, D-glucose was added to the furosemide solution to increase the osmolarity.

TABLE I. Chemical components and physical chemistry properties of the treatment solutions.

Group	Dosage (mM)	Glucose (mM)	Osmolarity[a] (mOsmol/kg H_2O)	pH[a]
Control	0	250	492.8±29.4	7.27±0.03
Furosemide	30	150	470.2±17.0	7.29±0.04
4-AP	150	0	491.2±16.4	7.26±0.07

[a]Mean ± s.d.

The osmolarity of 30-mM furosemide in 150-mM glucose was 470±17 (s.d.) mOsmol/kg H_2O, close to the 4-AP solution. The control solutions maintained an average osmolarity of 493±29 mOsmol/kg H_2O, matching the other two treatment solutions. The chemical components and the physical chemistry properties of the solutions for each group are listed in Table I.

C. Experimental procedure

Animal preparation and experimental procedure were described previously (Bian and Chertoff, 1998b). Briefly, the animals were anesthetized with pentobarbital (64 mg/kg, i.p.) and maintained with 1/3 of initial dose given (i.m.) every hour. Rectal temperature was monitored and maintained at 37 °C by a homeothermic blanket (Harvard). The right bulla was exposed and opened to place a ball-tip silver wire electrode on the RW along side with a cotton wick to absorb the fluid in the RW niche. The bulla was kept open throughout the experiment. Acoustical stimuli were delivered by a μ-metal shielded headphone (Etymotic ER-2) via a closed tube connected to the bony external ear canal. At the end of the tube, a calibrated probe microphone (ER-7C) was sealed about 5 mm from the umbo of the tympanic membrane to monitor the signal levels and record the stimuli. A needle electrode was inserted into the neck musculature to serve as a ground.

At the beginning of the experiment, a CAP audiogram was estimated for frequencies of 1, 2, 4, 8, and 16 kHz with visual detection of N_1 on a digital storage oscilloscope (Hitachi VC-6045A). The CAP threshold was determined by reducing the intensity of the tone burst in 2-dB steps until a 50% visually detectable level. Only animals with normal hearing, i.e., CAP thresholds no higher than 25 dB SPL across the whole frequency range, were used in the study. Then, the CM in response to Gaussian noise was recorded twice separated by a 10-min rest. Next, two drops of the treatment solution were dripped into the RW niche and kept for 10 min. Once the solution was dried with cotton wicks, a third recording of CM response to Gaussian noise initiated. Immediately after the third CM recording, a CAP audiogram was obtained. The CAP threshold at 16 kHz was monitored for 90 min from the beginning of the treatments to map the time course of the threshold shift.

D. Signal processing and data analysis

Gaussian noise of 250 ms duration, windowed with a 5-ms $\cos^2$ ramp, was used to evoke the CM. The noise was presented at 68 and 88 dB SPL in random order. The signal level of the Gaussian noise was maintained at the desired level by automatically adjusting the attenuator according to the measured input rms from the probe microphone. The noise spectrum was equalized *in situ* between 0.1 and 10.24 kHz with an inverse filter method (Chertoff and Chen, 1996). The noise was limited below 11 kHz to minimize the out-of-phase cancellation in the CM recorded from the RW. Tone bursts 2 ms in duration, with a 1-ms $\cos^2$ rise and fall shape, were used to elicit CAP. All the acoustic signals were generated in an array processor (Tucker-Davis Technology, TDT AP2), sampled at 65.536 kHz in a 16-bit digital-to-analog converter (TDT DA1), and sent to a headphone buffer (TDT HB6).

Electric responses from the cochlea were amplified 500 times and bandpass filtered between 0.03–30 000 Hz (12 dB/octave) by a low-noise preamplifier (Stanford SR560). The signals were subsequently low-pass filtered at 16 kHz (54 dB/octave), further amplified 10 times by a dual variable filter (Steward VBF 10M), and digitized at 65.536 kHz (TDT AD2). The Gaussian noise and the corresponding CM response were averaged ten times and saved as an input/output record. At each signal level, ten records were obtained from ten different sequences of Gaussian noise. The NLID procedures were followed to analyze these records (MATLAB 5.3, MathWorks). First, the time-domain signal was Hanning windowed with 50% overlapping to correct the energy loss in the spectrum (Bendat and Piersol, 1986). Second, the frequency-domain coefficients of the polynomial model of MET were obtained by the ratio of the cross-spectral density function and the conditioned input autospectral density function for each path in the model (Chertoff *et al.*, 1996; Bendat and Palo, 1990; Bendat and Piersol, 1986). Third, physiologic indices that characterize MET were derived in a manner identical to our previous work (Chertoff *et al.*, 1997). Figure 1 shows the physiologic indices on an MET curve plotted from the third-order polynomial model.

E. Statistical analysis

Two change scores of each dependent variable among the three consecutive recordings were obtained by subtracting the value of an earlier recording from the next one. The first change score represents the effect of the 10-min rest, whereas the second is the effect of the treatment. These difference scores for each group were submitted to a three-way (group, time, frequency) repeated-measure analysis of variance (ANOVA) (STATISTICA 5.0, StatSoft). If the general ANOVA was significant, a two-way (group, frequency) ANOVA was performed for rest and treatment conditions separately. If there was a significant group × frequency interaction, the two experimental groups were compared with each other using a between-group one-way ANOVA. When the homogeneity of variance and sphericity assumptions were violated, the Huyn–Feldt procedure was adopted to adjust the degrees of freedom (Keppel, 1991). The effect sizes (η^2) for the group, frequency, and frequency × group effects were estimated (Young, 1993), and categorized with Cohen 's criteria (1988). The same statistical procedure was applied to the CAP threshold shifts. If a group × frequency interaction was significant, a Newman–Keuls *post hoc* com-

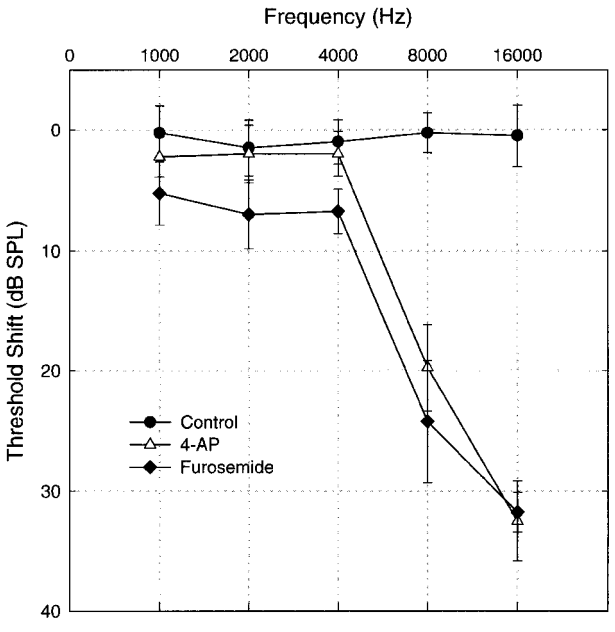


FIG. 2. The average CAP threshold shift ± s.d. for the three groups.

parison was made to determine the frequencies where the threshold shifts differed between two groups. A p value of less than 0.05 was considered statistically significant.

III. RESULTS

A. Hearing loss

1. Audiograms

The CAP audiogram was constructed by subtracting thresholds before treatment from thresholds after treatment for each frequency. The mean and standard deviation of threshold shift for the three groups are shown in Fig. 2. No considerable threshold change across the whole frequency range was detected after dripping the control solution of glucose. However, a similar high-frequency hearing loss was found in the ears that received either 4-AP or furosemide solution. The largest threshold shift, approximately 32 dB SPL, occurred at 16 kHz for both experimental groups. The average threshold shift at 8 kHz was 20 dB SPL for the 4-AP group and 24 dB SPL for the furosemide group. These threshold shifts were significantly different from the control group. At 4 kHz and below, the 4-AP group had an average of 2 dB SPL threshold change, which was not significantly different from the control, whereas the furosemide group had a threshold shift about 5–7 dB SPL that was significantly different from the control.

2. Kinetics of threshold shift

The CAP threshold at 16 kHz was monitored from the beginning of the treatment for up to 90 min for all the animals. The purpose of monitoring threshold was to ensure that the CM recording was made at the time when the animal had the largest threshold shift and no considerable recovery of the threshold. The time courses of the threshold shift in the three groups are illustrated in Fig. 3. It is apparent that glucose did not result in any threshold shift (top panel). However, the 4-AP and the furosemide exposure yielded a 30–40-dB SPL threshold shift after the 10-min exposures (lower two panels). Following the maximum threshold shifts, there were slow recoveries of different rates between the two experimental groups.

The time course of the threshold shift for each animal fit well (mean $r^2=0.95$) to a first-order kinetic model, which is a double exponential decay function (SIGMAPLOT 5.0, SPSS)

$$\theta = C_2 e^{-t/\tau_2} - C_1 e^{-t/\tau_1}, \tag{2}$$

where τ_1 is the time constant of the onset of the threshold shift, and τ_2 is the time constant for the recovery from the maximum threshold shift; C_1 and C_2 are the exponential coefficients, so that the sum of them is equal to the threshold shift at time zero, i.e., $\theta_0 = C_2 - C_1$.

The average τ_1 was about 5.5 ± 1.9 (s.d.) min for the 4-AP group and 5.0 ± 1.0 (s.d.) min for the furosemide group (middle and lower panels). As shown in the lower two panels, the recovery time constant (τ_2) of the furosemide group was 263.8 ± 75.9 (s.d.) min and the τ_2 of the 4-AP group was 86.7 ± 48.6 (s.d.) min. It took three times longer for the furosemide group to recover from the maximal threshold shift than the 4-AP group. The medians, 10th, and 90th percentiles of τ_2 for the two experimental groups are illustrated in the inset box chart in the bottom panel of Fig. 3 for a direct comparison.

B. Effects of cochlear damage

1. Input/output

The average input/output spectral density functions for all groups are shown in Fig. 4. The acoustic energy in the input spectra was evenly distributed across the frequency range (0.1–11 kHz). The input spectra of the three groups are almost identical [panels (A) and (B)]. The shapes of the output spectra of the CM were similar at both levels, though the CM output for the 68-dB SPL Gaussian noise was about 20 times smaller compared to the 88 dB SPL condition [panels (C) and (D)]. The output spectra showed different amounts of reduction in the two experimental groups. The CM reduction by furosemide was always greater than 4-AP regardless of the input signal levels. Compared to the control group, the energy in the CM decreased approximately 95% in the furosemide group at both signal levels. The 4-AP showed 75% and 89% reduction in CM spectral density at 68 and 88 dB SPL, respectively. Similarly, the CM gain was reduced by the two treatments with greater effect in the furosemide group [Figs. 4(E) and (F)]. Furosemide and 4-AP decreased the gain by 83% and 70% relative to the control at 88 dB SPL, 82% and 54% at 68 dB SPL, respectively. The 4-AP exposure produced a progressive phase delay towards high frequencies (>2.5 kHz) at both signal levels [Figs. 4(G) and (H)]. However, the furosemide group showed no phase delay in the same frequency range. Compared with the control group, both drugs caused a phase lead at frequencies below 1.5 kHz.

2. Polynomial coefficients

Figure 5 illustrates the treatment-induced changes in the magnitudes of the polynomial coefficients for the three

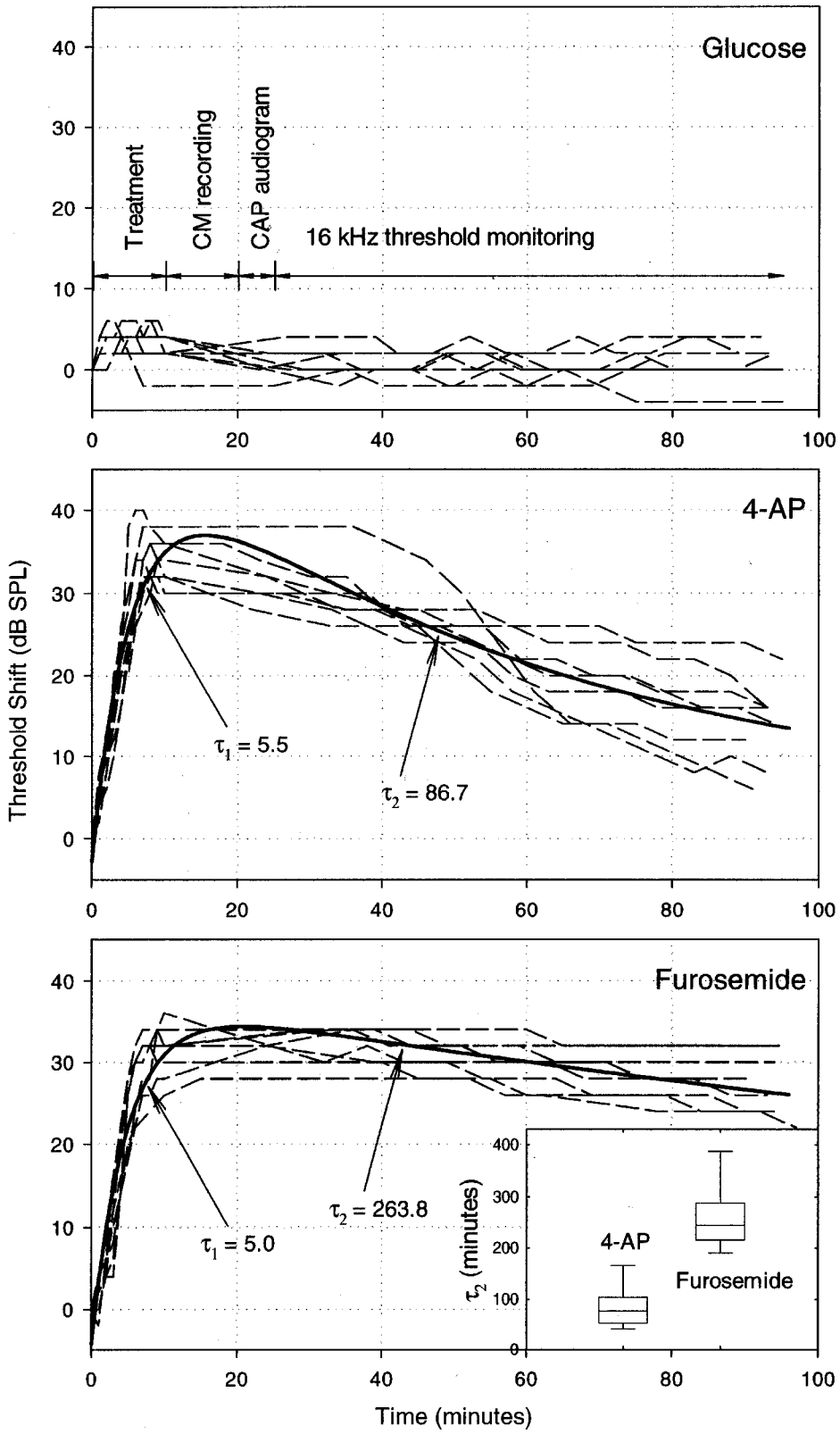


FIG. 3. Time courses of the CAP threshold shift at 16 kHz for each animal in the three groups (dashed lines). Treatments started at time zero and lasted 10 min. Solid lines indicate the best-fit kinetic model and associated time constants τ_1 and τ_2. Inset in the bottom panel shows the medians and the 10th and 90th percentiles of the recovery time constants (τ_2) for the two experiment groups. Also indicated in the top panel is the timing of the experiment procedures.

groups at the two signal levels. The change of a variable (in percentage), computed as the mean value across frequency for an experimental group relative to the control, is provided in parentheses. Both 4-AP and furosemide significantly reduced the amplitude of A_1 at both signal levels [panels (A) and (B)]. Furosemide showed a greater effect (75%) on A_1 than 4-AP (60%) at the 68 dB SPL condition, whereas both drugs displayed a similar effect (80%) at the higher signal level. The 4-AP group showed a significant increase (120%) in the magnitude of A_2 at frequencies between 1 and 5.6 kHz at 68 dB SPL, but not at the higher stimulus level [Figs. 6(C) and (D)]. Furosemide slightly reduced A_2 at both signal levels, but neither was statistically significant. Both drugs dramatically reduced (75%) the amplitude of A_3 at the 88 dB

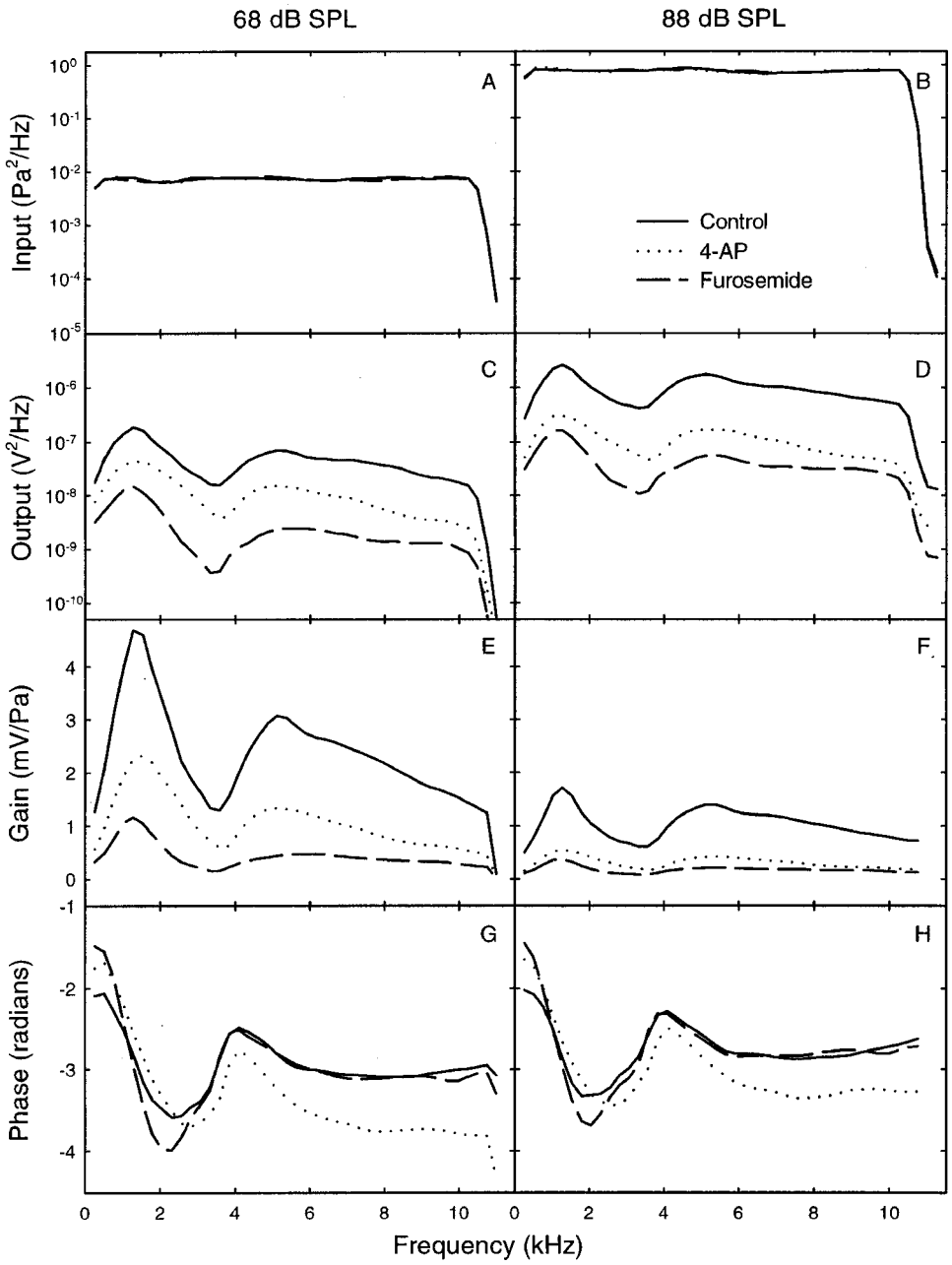


FIG. 4. Average input (1st row) and output (2nd row) spectral density functions of the three groups. The magnitude and phase of the input/output transfer function of the three groups are shown in the lower two rows.

SPL condition [Fig. 5(F)]. At the lower signal level, furosemide significantly decreased (80%) A_3 across the whole frequency range, but 4-AP showed no effect [Fig. 5(E)].

Furosemide and 4-AP exposures also altered the phase of the three polynomial coefficients (Fig. 6). The phase of A_1 showed a progressive delay towards high frequencies after the application of 4-AP at both signal levels [panels (A) and (B)]. In contrast, furosemide did not change A_1 phase for frequencies above 2.5 kHz at both signal levels. Moreover, the furosemide group showed a steep phase delay of A_1 at frequencies below 2 kHz. Both drugs did not alter the phase of A_2 with an exception at the lower signal level, where 4-AP group showed a phase lead in the high frequencies (>5 kHz) [Fig. 6(C)]. Furosemide and 4-AP significantly delayed the phase of A_3 towards high frequencies at the 68 and 88 dB SPL conditions, respectively [panels (E) and (F)]. The effect of furosemide on A_3 phase was more dramatic.

3. Coherences

The relative importance of a polynomial term in the model is measured by its coherence function, which is equivalent to the squared correlation coefficient of two variables in regression statistics. Each coherence function represents the unique contribution of the polynomial term to the spectrum of the response. The cumulative or total coherence, normally ranging from 0.92–0.98, is the sum of each coherence function and provides a measure of ''goodness of fit'' for the polynomial model of MET (Chertoff *et al.*, 1996). Both 4-AP and furosemide reduced the total coherence in frequencies below 4 kHz. This decrease in total coherence was mainly due to a significant reduction of the linear coherence [Figs. 7(A) and (B)]. At 68 dB SPL, furosemide resulted in a greater reduction than 4-AP with the maximal reduction of 0.3 (normal value 0.9) at 2.6 kHz. At 88 dB

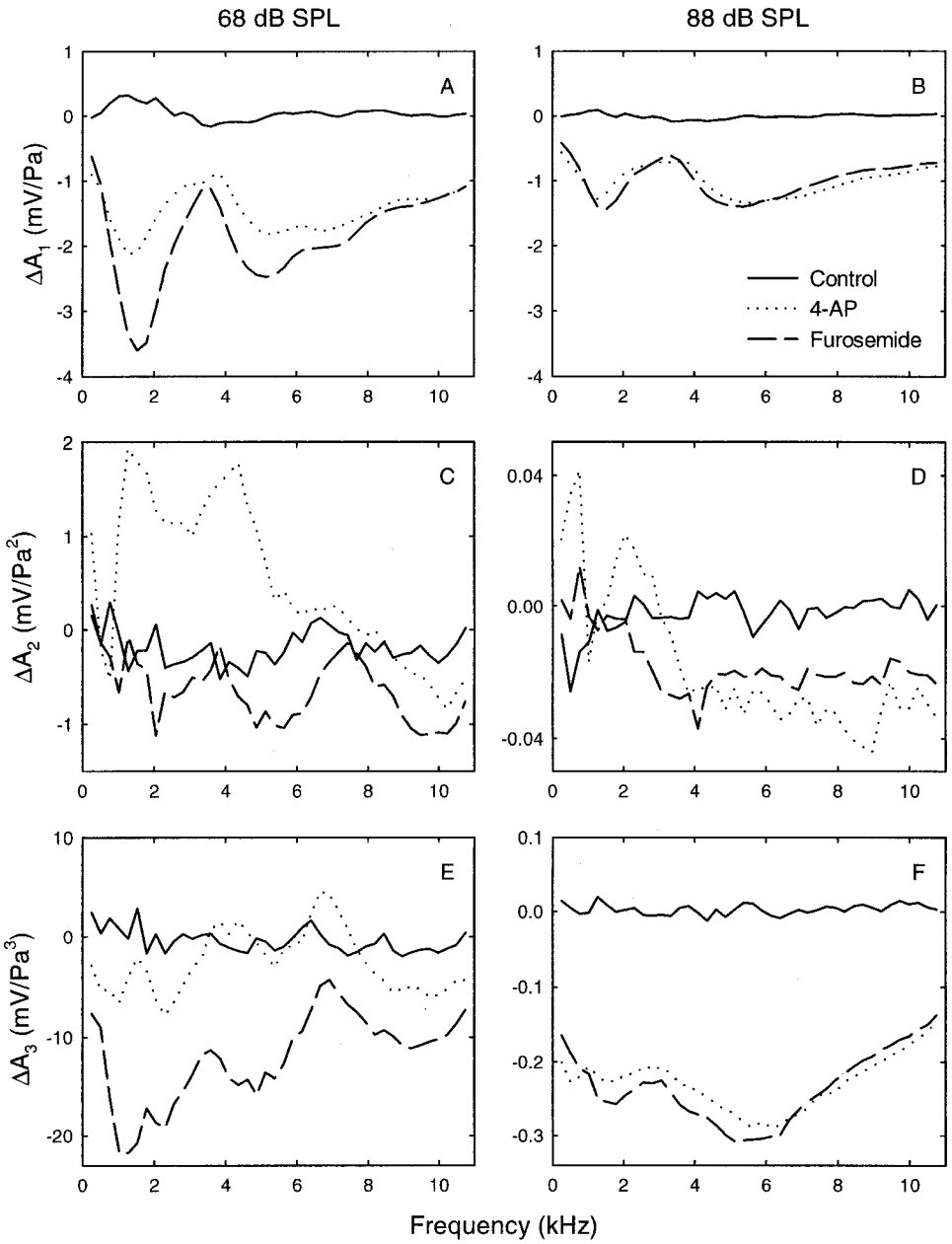


FIG. 5. Average changes in the polynomial coefficients for the three groups at 68 (left panels) and 88 dB SPL (right panels) conditions. Top panels, A_1; middle, A_2; bottom, A_3.

SPL, the 4-AP group showed a greater reduction in the linear coherence throughout the whole frequency region, with maximal effects of 0.2 (normal value 0.8) at 3.3 kHz. The furosemide increased the quadratic coherence in the 2–4-kHz region (normally 0.01) at both signal levels [panels (C) and (D)]; however, only at 88 dB SPL was the rise in quadratic coherence significant. The cubic coherence was increased by 4-AP and decreased by furosemide [panels (E) and (F)]. At 88 dB SPL, there was a significant difference between the effects of the two drugs in the high frequencies (>6 kHz). The increase in the cubic coherence by 4-AP was significant around 4 kHz (normally 0.003) at the lower signal level, whereas the effect of furosemide was not significant.

4. Physiologic indices

As shown in Fig. 1, to reduce the number of the physiologic indices, we combined the absolute values of maxima and minima, SPL_{max} and SPL_{min}, resulting in DR and OR, respectively (Bian and Chertoff, 1998a). A new index, symmetry, was adopted which is the absolute ratio of maxima/minima. We focus our report on these new indices and two other important ones: slope and OP. Table II summarizes the statistical results (F values, the adjusted degrees of freedom, and effect sizes) for the group, frequency, and group × frequency effects on the change scores of the physiologic indices between the two experimental groups. The relative importance of one effect on separating the two treatments can be determined by comparing the corresponding effect sizes (η^2).

The changes in the slope and the OP of MET due to the treatments are illustrated in Fig. 8. Since the slope is a negative value, positive change scores indicate a decrease, or shallower slopes of the MET curves. As shown in panels (A) and (B), both the furosemide and the 4-AP significantly decreased the slope of MET across the whole frequency range

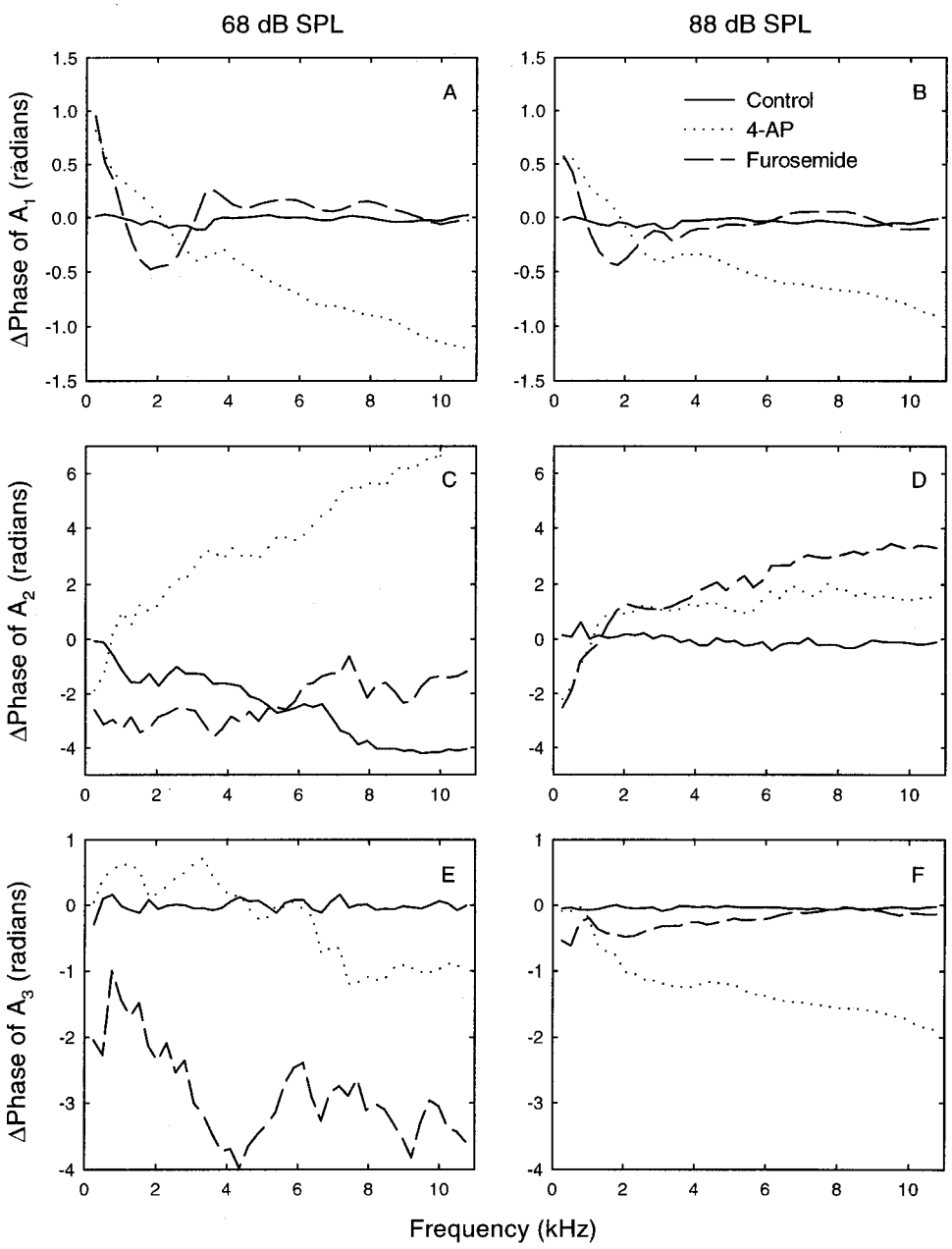


FIG. 6. Average changes in the phase of the polynomial coefficients. The left and right panels show the change scores for the phase of A_1 (upper panels), A_2 (middle panels), and A_3 (lower panels) under 68 and 88 dB SPL conditions, respectively.

at both signal levels with greater effects at 1.5, and 5.4 kHz. Furosemide showed greater reduction in slope than 4-AP at the lower signal level. By contrast, the OP of MET was changed in opposite ways by the two treatments [Figs. 8(C) and (D)]. Negative change score indicates that the OP moved towards negative sound pressures; positive value represents a shift towards positive sound pressures. At both stimulus levels, the 4-AP solution moved the OP in the negative sound-pressure direction. However, the furosemide treatment shifted the OP in the positive direction at the 88-dB SPL condition. The effect of 4-AP was more dramatic at the higher signal level. The total η^2 for the OP was the largest among the indices.

The symmetry values range between 0 and 1, with 1 representing a perfect symmetric MET curve. The normal value of symmetry was 0.6–0.8 at both signal levels, indicating that the magnitude of positive CM output was smaller than the negative CM. A negative change score means that the MET curve became more asymmetric than normal; a positive change indicates an MET curve that is more symmetric than normal. The symmetry was changed differently by the two treatments [Figs. 9(A) and (B)]. The 4-AP exposure significantly reduced the symmetry, i.e., the MET became more asymmetric at both signal levels. The furosemide solution decreased the symmetry in the frequencies below 6 and 4 kHz for 68- and 88-dB SPL conditions, respectively. Generally, 4-AP dramatically decreased the symmetry across all frequencies, whereas the effect of furosemide was smaller and frequency dependent.

The OR is the overall width of the MET curve. Under the 68-dB SPL condition, the 4-AP reduced the OR around 7 kHz, whereas the furosemide increased the OR in the whole frequency range [Fig. 9(C)]. At the higher signal level, both the 4-AP and the furosemide widened the OR in the low (<5

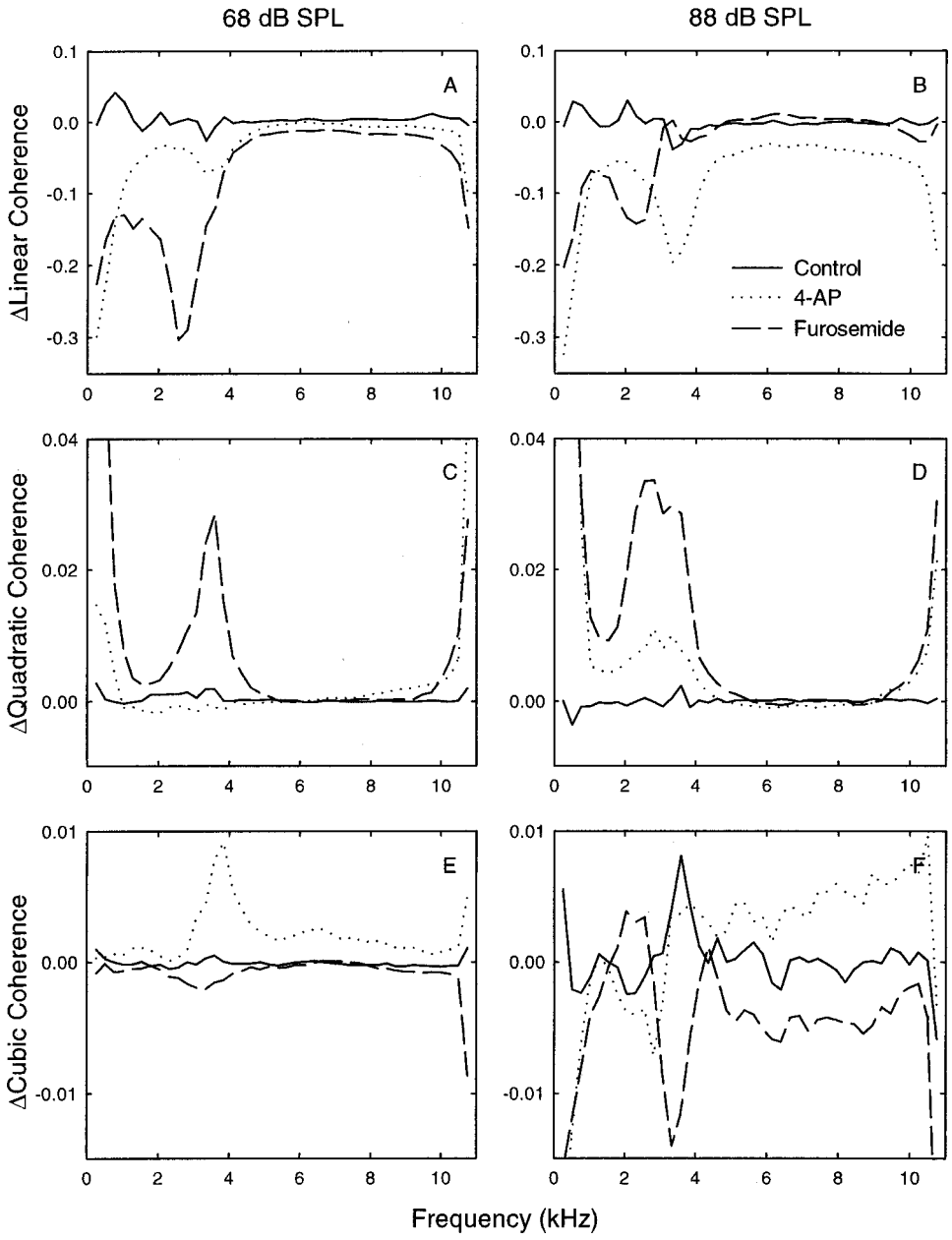


FIG. 7. Average changes in the multiple coherence functions of the three groups at the 68 (left panels) and 88 dB SPL (left panels) stimulus levels. Upper row: the linear coherence; middle row: the quadratic coherence; lower row: the cubic coherence.

kHz) and high (>3 kHz) frequencies, respectively [Fig. 9(D)]. The DR represents the overall peak-to-peak amplitude of the CM. Both of the groups demonstrated a similar reduction in DR [Figs. 9(E) and (F)]. This reduction was more dramatic at 2 and 6.5 kHz at 68 dB SPL, and 1.5 and 5 kHz at 88 dB SPL. The effect was greater at the higher signal level.

Since the physiologic indices are functions of frequency, the MET curves within the whole frequency range can be plotted as a 3D surface (Fig. 10). The shape of the surface provides a detailed description of the MET process in the cochlear base. At both 68- and 88-dB SPL conditions, the surface became flattened due to the effects of the ototoxic treatments. However, the surface of the 4-AP treated group is shifted downwards and to the left, whereas that of the furosemide group is less curvy and moved upward. These characteristics reflect the changes in OP, symmetry, slope, OR, and DR.

IV. DISCUSSION

A. Kinetics of hearing loss

The extent and configuration of hearing loss produced by 4-AP and furosemide were similar to that induced by salicylate and an 8-kHz pure-tone exposure from our previous study (Bian and Chertoff, 1998b). The CAP threshold shift started at 8 kHz and was most prominent at an octave above, i.e., 16 kHz. The recovery time constant (τ_2) of the 4-AP exposure (86.7 min) was close to that of salicylate (83.7 min). However, the τ_2 of the furosemide group (263.8 min) was much longer than either salicylate, 4-AP, or pure-tone exposure (147.8 min). The different kinetics of the hearing loss due to drug treatments may relate to the site of drug action and the size of the drug molecules. Both 4-AP and salicylate alter the membrane characteristics of the OHC body (Mammano and Ashmore, 1996; Shehata *et al.*, 1991),

TABLE II. Statistical results: effects and effect sizes (η^2s) on the physiologic indices. The F values, the adjusted degrees of freedom, and the η^2 were based on the ANOVA between the 4-AP and the furosemide group.

		68 dB SPL			88 dB SPL		
		F [a]	df [b]	η^{2c}	F	df	η^2
OP [d]	g×f [e]	4.65[h]	8, 109	0.06	5.73[h]	3, 47	0.05
	g[f]	13.79[h]	1, 14	0.34	30.74[h]	1, 14	0.56
	f [g]	7.78[h]	39, 546	0.10	1.83[h]	39, 546	0.02
Sym[h]	g×f	3.74[h]	18, 247	0.07	3.84[h]	3, 43	0.04
	g	33.70[h]	1, 14	0.37	17.55[h]	1, 14	0.40
	f	7.44[h]	39, 546	0.14	9.13[h]	39, 546	0.10
Slope	g×f	2.70	2, 26	0.08	2.52[h]	6, 89	0.03
	g	2.33	1, 14	0.04	0.01	1, 14	0.00
	f	9.44[h]	39, 546	0.27	40.13[h]	39, 546	0.53
DR[i]	g×f	1.03	39, 546	0.05	2.28[h]	7, 93	0.04
	g	0.86	1, 14	0.01	0.01	1, 14	0.00
	f	2.74[h]	39, 546	0.14	32.59[h]	39, 546	0.57
OR[j]	g×f	1.46	8, 111	0.04	1.52	3, 36	0.04
	g	30.51[h]	1, 14	0.32	0.02	1, 14	0.00
	f	9.29[h]	39, 546	0.08	2.70[h]	39, 546	0.07

[a]Each test involves the comparison between the two experimental groups.
[b]Degrees of freedom were adjusted for any violation of homogeneity and sphericity assumptions.
[c]Effect size was judged: $\eta^2>0.14$, large; $0.14>\eta^2>0.01$, medium; $\eta^2<0.01$, small (Cohen, 1988).
[d,h,i,j]OP: operating point; Sym: symmetry; DR: dynamic range; OR: operating range.
[e,f,g]g×f: group × frequency interaction; g: group effect; f: frequency effect.
[h]$p<0.05$.

whereas furosemide damages the stria vascularis. It could take a longer time to eliminate furosemide from stria vascularis due to its relatively larger molecule size (FW=330.7). Different recovery rate could also reflect the various types of damage produced by these agents. Recovery from OHC stereocilial damages induced by tone exposure could be slower than that of hair cell membrane alteration by 4-AP and salicylate. Furosemide causes both strial edema (Pike and Bosher, 1980) and stereocilial damages (Comis *et al.*, 1990; Forge and Brown, 1982); thus, it may prolong its ototoxicity.

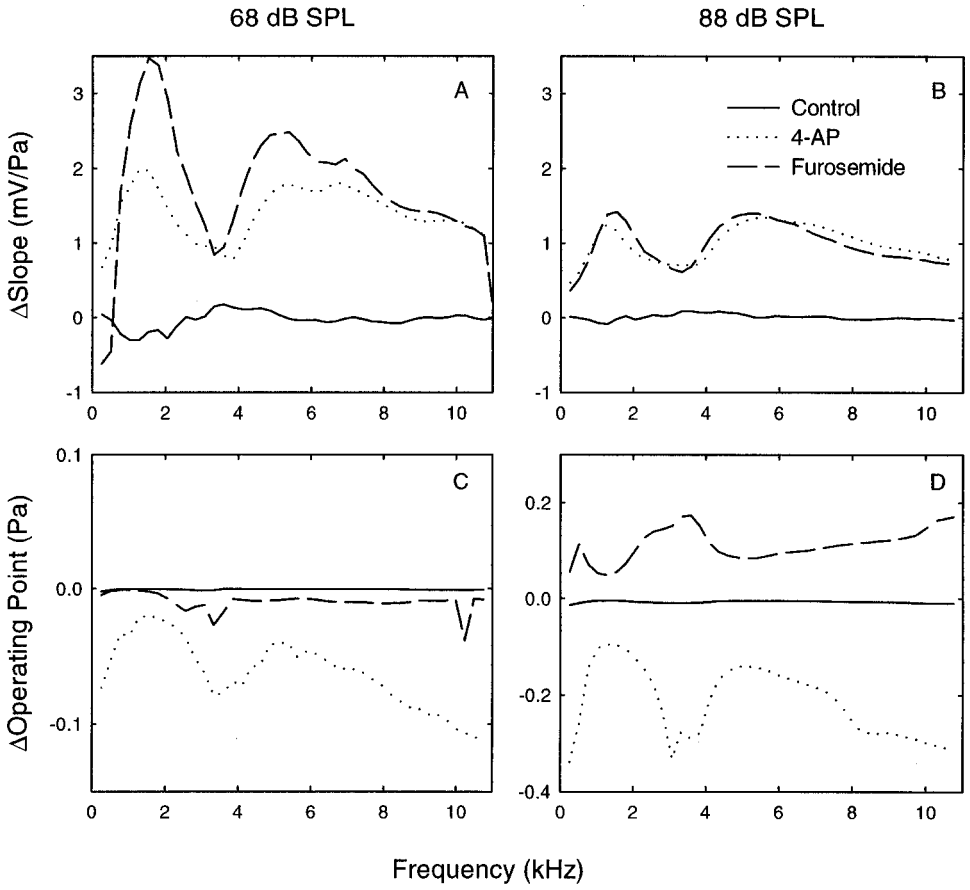


FIG. 8. Average changes in physiologic indices for the three groups (1). The change scores of the slope, and operating point of the MET are shown in the upper and lower panels for the 68 and 88 dB SPL conditions, respectively.

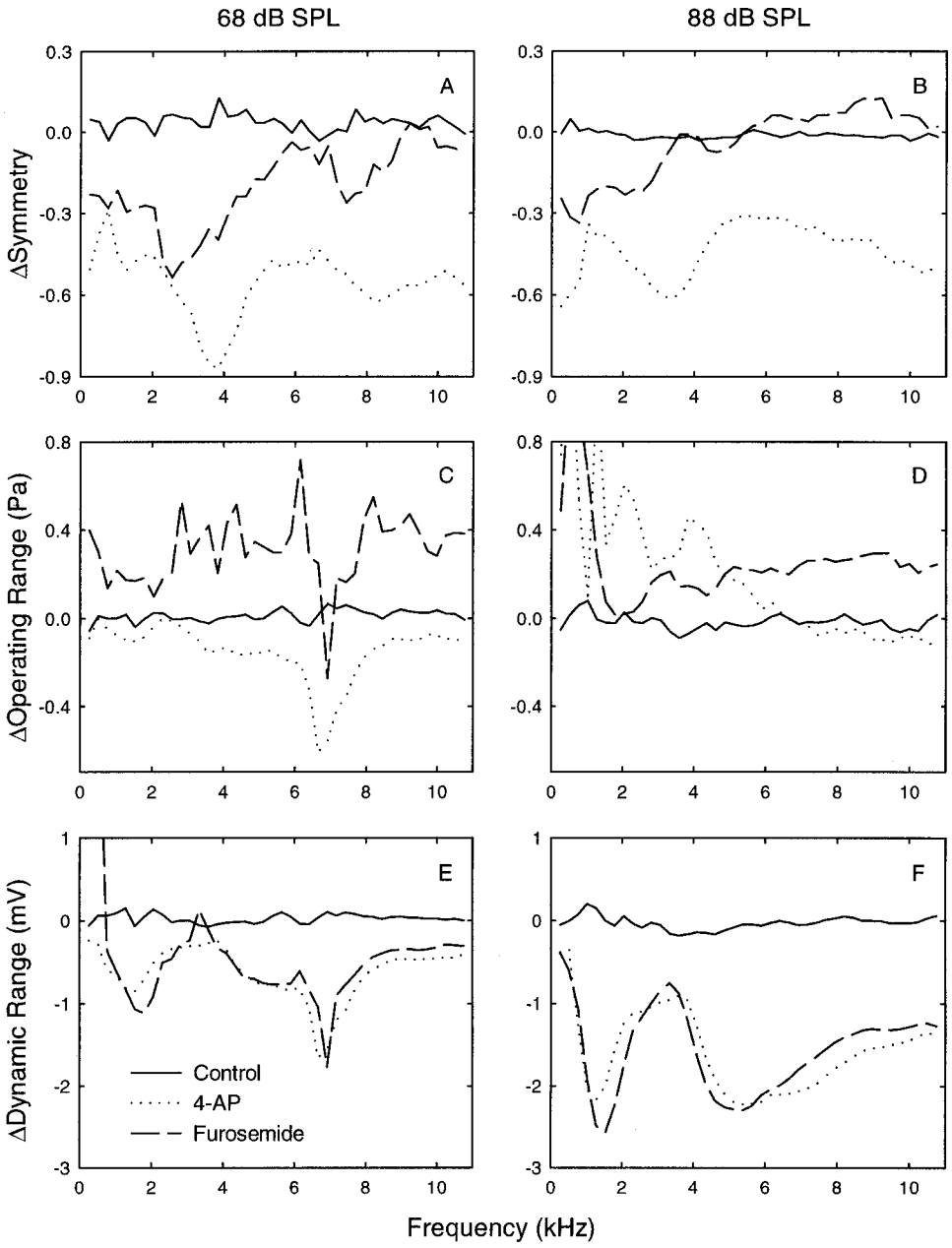


FIG. 9. Average changes in physiologic indices for the three groups (2). Left panels: 68 dB SPL; right panels: 88 dB SPL. Upper row: symmetry; middle row: operating range; lower row: dynamic range.

B. 4-AP exposure

The application of the K^+-channel blocker 4-AP to the RW induced a high-frequency CAP threshold shift and a reduction in the amplitude of the CM. This is consistent with the findings of other investigators (Wang *et al.*, 1993; Kirk and Yates, 1998). The subsequent changes in the polynomial parameters involved reductions in the CM gain, the A_1 and A_3 magnitudes, an increase in the amplitude of A_2, and a phase delay in the high frequencies. Moreover, the linear coherence was reduced while the cubic coherence increased at the high frequencies. As a result of these changes, the MET curve became more asymmetric with a shallower slope, reduced DR, and narrowed OR (Fig. 10).

The reduced CM magnitude, DR, and slope of MET can be explained by the blockade of the K^+ channels in the OHCs. One of the prominent outward K^+ currents through the OHC basolateral wall is a 4-AP sensitive channel (Lin *et al.*, 1995; Mammano and Ashmore, 1996; Nenov *et al.*, 1997). It is known that a fast K^+-channel (K_A) or A current (I_A), is very sensitive to the inhibition of 4-AP (Soria, 1998). Blocking of the K_A channels could suppress the receptor currents and result in a flatter slope and a smaller DR of MET. Since the CM was recorded at the RW, the effective blocking of the gross receptor currents by 4-AP suggests that the I_A could be the dominant CM generating pathway in the OHCs at the base of the gerbil cochlea. This is in agreement with the observation that shorter OHCs isolated from the basal turns of the guinea pig cochlea showed a rapidly activated K^+ current (Nenov *et al.*, 1997).

After the 4-AP exposure, the symmetry of MET was reduced. This is equivalent to a dc shift of the CM towards negative voltages. This result is comparable to the finding that 4-AP can increase the negative summating potential (SP) (Wang *et al.*, 1993; van Emst *et al.*, 1996). Blockade of the I_A near the cochlear base may cause the CM to be dominated by hair cells from more apical regions. Such a move of

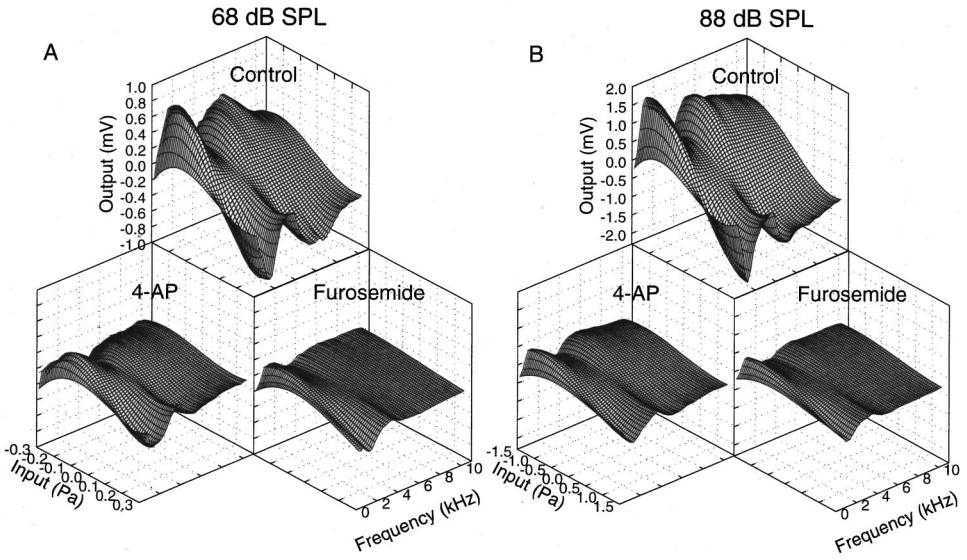


FIG. 10. 3D mesh plots of the MET curve as a function of frequency for the three groups. Panel (A): 68 dB SPL; panels (B): 88 dB SPL.

the CM source is consistent with the progressive phase delay at high frequencies [Figs. 4(G) and (H)]. More apical OHCs with more asymmetric input/output function than basal cells would produce a larger depolarizing dc component (Housley and Ashmore, 1992; Dallos, 1986). Another possibility is that the K^+ currents through the IHCs could increase, due to the inhibition of I_A in the OHCs. At a given sound-pressure level, IHCs produce greater dc receptor potentials in the depolarizing direction, and the IHC voltage response to sound is more asymmetric compared to OHCs (Russell *et al.*, 1986; Dallos, 1986). Since IHCs contribute significantly to the SP recorded at the RW (Durrant *et al.*, 1998), an increase in the K^+ currents through IHCs could elevate the negative SP. Therefore, selective blocking of I_A in the OHCs could account for the reduction in the symmetry of MET by 4-AP.

Application of 4-AP significantly shifted the OP in the negative sound-pressure direction, which corresponds to a displacement of the cochlear partition towards scala vestibuli (SV). Since 4-AP has no inhibitory effect on the secretion of K^+ from the stria vascularis (Hibino *et al.*, 1997; Takeuchi and Ando, 1999), there could be an imbalance of the K^+ circulation within scala media, i.e., higher K^+ concentration ($[K^+]$). This is consistent with Kirk and Yates (1998), who showed an increase in endocochlear potential (EP) following 4-AP application. Elevated extracellular $[K^+]$ tends to depolarize the OHCs and cause a shortening of the cells (Dulon *et al.*, 1988; Santos-Sacchi and Dilger, 1988). In an experiment on isolated OHCs of guinea pigs, the presence of 4-AP indeed allowed the cells to shorten in response to electrical stimulation (Ohnishi *et al.*, 1992). Such static contraction of the OHC could shift the basilar membrane (BM) towards SV (Zimmermann and Fermin, 1996).

C. Furosemide exposure

Dripping furosemide solution onto the RW caused a high-frequency hearing loss that was similar to the threshold shift induced by 4-AP. Furosemide significantly reduced the CM magnitude, which is in agreement with the previous findings (Comis *et al.*, 1990; van Emst *et al.*, 1997). The typical effects of furosemide also included a decrease in the polynomial coefficients, the MET slope, and an increase in the OR of MET. The OP shifted towards positive sound pressures. The linear and cubic coherences were reduced, whereas the quadratic coherence increased. Consequently, the MET curves became much flatter, and generally symmetric in the high frequencies (Fig. 10).

A possible explanation for the reduction in CM magnitude and the MET slope or sensitivity by furosemide could be attributed to the decrease in the driving voltage across hair cells due to the inhibition of the cochlear ''battery,'' stria vascularis. Furosemide inhibits the Na^+/K^+ ATPase and $Na^+/K^+/2Cl^-$ cotransporter in the stria (Rybak, 1986; Salt *et al.*, 1987), thus resulting in a decrease in $[K^+]$ within scala media (Rybak and Morizono, 1982), and a reduction in EP (Sewell, 1984). Low EP would result in a fall of the receptor potential and diminished ion currents passing through the OHCs. For a given variation in sound pressure, there would be less change in the receptor current, i.e., a shallower slope of MET. In addition, the alteration in the ionic composition of the cochlear fluids can indirectly damage the stereocilia of the OHCs in the basal turns, such as stretched cross links, coarseness of the cilial membrane (Comis *et al.*, 1990), and splaying of the cilial bundles (Forge and Brown, 1982). These ultrastructural changes induced by furosemide may hinder the normal functionality of the stereocilia, decrease the apical K^+ conductance, and further depress the receptor currents. Moreover, furosemide could decrease the voltage-dependent OHC electromotility (Ohnishi *et al.*, 1992), and impede the vibration of the BM by attenuating the mechanical feedback of the OHC to the cochlear partition at low stimulus intensities (Ruggero and Rich, 1991). Consequently, furosemide could reduce the slope of MET, especially at the lower signal level.

Furosemide moved the OP in the positive sound-pressure direction at the higher signal level. This represents either a shift of the cochlear partition towards ST, or a transit of the OHC stereocilia to an excitatory direction. This is in agreement with the observation that furosemide generated a large positive SP within 10 min after systemic injection (Syka and Melichar, 1985; van Emst *et al.*, 1997). Reduction

in scala media $[K^+]$ due to the inhibition of furosemide could hyperpolarize the OHCs (Gitter, 1993) and cause the cells to elongate. Fall of the $[K^+]$ in the cochlear fluids surrounding the OHC may induce a hypo-osmotic swelling (Dulon *et al.*, 1988; Chertoff and Brownell, 1994) *in vitro*, or an increase in cell length *in vivo*. Lengthening of the OHCs could either move the cochlear partition towards ST or deflect the stereocilia in the opening direction. Indeed, Comis *et al.* (1990) found that the upward-pointing cross links were stretched, a sign of the hair bundle being pushed towards the tallest stereocilia. Furthermore, furosemide can inhibit the shortening of the OHC induced by electrical stimulation (Ohnishi *et al.*, 1992). Shortening is a normal behavior of the OHC at high signal level and results in a shift of the OP to negative sound pressures (Bian and Chertoff, 1998b). Inhibition of such movement of the cochlear partition by furosemide could contrarily shift the OP of MET towards positive sound pressures, prominently at the higher stimulus level [Fig. 8(D)].

D. Comparison of treatments

1. Furosemide vs 4-AP

Application of 4-AP and furosemide caused different changes in the polynomial coefficients and coherences, and subsequently resulted in the differences in the physiologic indices. The most important index for distinguishing the pathophysiologic changes induced by 4-AP and furosemide was the OP, whose overall η^2 was the largest (Table II). The result that furosemide and 4-AP shifted the OP in opposite directions is similar to the finding of Russell and Kössl (1991). They injected positive and negative electric currents into OHCs *in vivo*, and shifted the OP of MET in the depolarization and hyperpolarization directions, respectively. Positive current injection could resemble the low basolateral K^+ conductance in the case of 4-AP application, while negative current injection could simulate the low EP caused by furosemide.

Another index that can be used to distinguish the effects of 4-AP from furosemide is the symmetry of MET. The 4-AP group showed a reduction in the symmetry across frequency, whereas the furosemide group showed no change in the symmetry at high frequencies. The decrease in OHC basolateral K^+ conductance caused by 4-AP could reduce the magnitude of OHC repolarization that generates positive CM, and increased apical extracellular $[K^+]$ may enhance the size of depolarization which produces negative CM. The moving of the standing current to IHCs could also contribute to the large reduction in the symmetry. However, lengthening of the OHC due to low $[K^+]$ caused by furosemide could push the stereocilia in the opening direction, thereby decreasing the magnitude of the OHC depolarization relative to hyperpolarization or a more symmetric MET. The gradient of the severity in furosemide damage from the RW could result in the frequency effect of the symmetry change.

The phase of the CM and A_1 showed different change following the two treatments [Figs. 4(G), (H), and Figs. 6(A), (B)]. Furosemide caused no phase change in frequencies above 2.5 kHz, whereas 4-AP caused a progressive phase delay across frequencies. This may suggest that the generation of the CM is shifted along the cochlear partition when the K^+ channels in the OHCs of the cochlear base are blocked. There was no such shift when the stria vascularis was inhibited by furosemide. Therefore, CM phase could be used to discern whether the sites of cochlear lesions that affect the generation of CM are hair-cell-related or not.

Nonlinear components in the polynomial model of MET were changed differently. Specifically, furosemide increased the quadratic coherence by more than 0.03 while it reduced the cubic coherence in the high frequencies at the lower signal level; 4-AP increased cubic coherence at both signal levels [Figs. 7(C), (D) and (E), (F)]. Rises in the weighting of the quadratic and fall in the cubic path of the MET are similar to the increase and decrease in the even- and odd-order distortion products in the CM following furosemide injection (van Emst *et al.*, 1997). Such differences in the odd and even components of the cochlear transduction induced by furosemide were also observed by others (Mills *et al.*, 1993; Mills and Rubel, 1994). They found that ''active'' or cubic distortion products otoacoustic emission evoked by low-level primary tones were diminished by furosemide, accompanied by a sharp phase lag. It is worth noting that the magnitude of A_3 was significantly decreased and its phase remarkably delayed at the lower signal level [panels (E) of Figs. 5 and 6].

2. Comparison with previous study of tone exposure

There are similarities and differences in the CM amplitude, polynomial coefficients, and physiologic indices among the treatments in the present study and our previous research on animals exposed to an 8-kHz pure tone (Bian and Chertoff, 1998b). The similarities include decrease in CM magnitude and slope of MET indicating reduction in hair cell receptor currents. However, different mechanisms mediate the effects, i.e., acoustic overstimulation collapses apical transduction channels in the stereocilia of the OHCs, 4-AP blocks K^+ channels in the basolateral wall of the OHCs, and furosemide diminishes the generation of the standing currents from the stria vascularis.

The most prominent difference between the 8-kHz tone and furosemide exposures is that the OP shifted in opposite directions. Structural alterations of the stereocilia, such as bending and disarray caused by noise exposure (Saunders *et al.*, 1991), could directly displace the cochlear partition towards SV. In contrast, reduction of endolymph K^+ and EP by furosemide could increase the OHC length via hyperpolarization or inhibition of the OHC shortening (Ohnishi *et al.*, 1992), thus moving the cochlear partition towards ST. Moreover, the OP shift caused by 4-AP was similar to the tone exposure. Like the tone exposure which blocks the apical transduction channels on the OHC, 4-AP blocked the basolateral K^+ channel, possibly causing a high extracellular $[K^+]$ (Johnstone *et al.*, 1989), and a subsequent shortening of the OHCs and an OP shift towards SV. However, the changes caused by 8-kHz tone exposure showed a large frequency effect, i.e., deep notches presented at certain frequencies.

Both exposures to pure tone and furosemide increased the OR. This could be attributed to the OHC stereocilial

damages induced by both treatments, i.e., more acoustic energy is needed to open the apical transduction channels when stereocilia are collapsed or stretched. In contrast, the 4-AP group showed a narrowed or unaffected OR [Figs. 9(C) and (D)]. Reduction in the basolateral K^+ conductance could saturate the receptor current at a smaller deflection of the stereocilia; in other words, less sound pressure is required to saturate the MET. Therefore, OR could be an indicator of the status of the OHC stereocilia. Expanded OR may suggest stereocilial malfunction of the OHCs; narrowed or unchanged OR may reflect normal stereocilia.

The reductions of the total and the linear coherences after the furosemide and 4-AP applications were not found in our previous study (Bian and Chertoff, 1998b). On the contrary, 8-kHz tone exposure increased these coherences, indicating that MET became more linear. However, exposures to 4-AP and furosemide resulted in more nonlinearity. Increasing the number of polynomial orders failed to improve the total coherence. Thus, the drop in the total coherence may suggest either the response is small and contains noise contamination, or a third-order polynomial function is not the best model of MET. The former is less likely, since the decrease in total coherence also presented at the higher stimulus level, suggesting that the latter is the case and other possible models, e.g., Boltzmann function, should be explored. Future experiments and model simulations are needed to determine the exact mechanisms of the alterations of the physiologic indices of MET.

V. SUMMARY AND CONCLUSION

The results of this study showed that even though the CAP thresholds were shifted equally and the amplitude of CM was reduced similarly by the RW applications of 4-AP and furosemide, the polynomial parameters and the physiologic indices of the MET model still demonstrated remarkable differences. The OP of the cochlear MET shifted towards SV in the 4-AP group, and towards ST in the furosemide group. There were frequency- and level-dependent differences between the two treatments on the symmetry, slope, and OR of the polynomial model of MET. The polynomial coefficients were also altered in different ways such that the coherences of the model showed stronger quadratic dominance in the furosemide group, and more cubic contribution in the 4-AP group. The outcomes of the present study show that the polynomial model and its parameters and associated physiologic indices are both sensitive and specific to discern changes in the cochlear micromechanics. Further research is needed to develop clinical tools in the differential diagnosis of cochlear hearing losses, leading to new rehabilitation techniques for hearing-impaired individuals.

ACKNOWLEDGMENTS

This work was based on part of the doctoral dissertation of Lin Bian. The authors wish to thank Jeff Radel, Xing Yi, John Belmont, Greg Ator, John Ferraro, Lawrence Sullivan, and Darren Wallace for their valuable comments on the manuscript and help in the experiments. This study was supported in part by the National Institute on Deafness and Other Communication Disorders of NIH, Grant No. 1 R29 DC02117-04A2, and USPHS Grant No. HD02528 to the Smith Mental Retardation and Human Development Research Center at the University of Kansas Medical Center.

AHFS (**1999**). ''Furosemide,'' in *American Hospital Formulatory Service: Drug Information® 1999* (American Society of Health-System Pharmacists, Inc., Bethesda, MD), pp. 2323–2327.

Bendat, J. S. (**1998**). *Nonlinear System Techniques and Applications* (Wiley-Interscience, New York).

Bendat, J. S. (**1990**). *Nonlinear System Analysis and Identification from Random Data* (Wiley-Interscience, New York).

Bendat, J. S., and Palo, P. (**1990**). ''Practical techniques for nonlinear system analysis/identication,'' Sound Vib. **24**, 28–33.

Bendat, J. S., and Piersol, A. G. (**1993**). *Engineering Applications of Correlation and Spectral Analysis*, 2nd ed. (Wiley, New York).

Bendat, J. S., and Piersol, A. G. (**1986**). *Random Data: Analysis and Measurement Procedures*, 2nd ed. (Wiley, New York).

Bian, L., and Chertoff, M. E. (**1998a**). ''Similar hearing loss, different physiology: Characterizing cochlear transduction in pure tone and salicylate damaged ears,'' Abstr. Assoc. Res. Otolaryngol. **21**, 84.

Bian, L., and Chertoff, M. E. (**1998b**). ''Differentiation of cochlear pathophysiology in ears damaged by salicylate or a pure tone using a nonlinear systems identification technique,'' J. Acoust. Soc. Am. **104**, 2261–2271.

Brown, R. D., and Feldman, A. M. (**1978**). ''Pharmacology of hearing and ototoxicity,'' Annu. Rev. Pharmacol. Toxicol. **18**, 233–252.

Chertoff, M. E., Steele, T. C., and Bian, L. (**1997**). ''Characterizing cochlear mechano-electric transduction in ears damaged with pure tones,'' J. Acoust. Soc. Am. **102**, 441–450.

Chertoff, M. E., Steele, T., Ator, G. A., and Bian, L. (**1996**). ''Characterizing cochlear mechano-electric transduction using a nonlinear system identification procedure,'' J. Acoust. Soc. Am. **100**, 3741–3753.

Chertoff, M. E., and Chen, J. (**1996**). ''An *in situ* calibration procedure for click stimuli,'' J. Am. Acad. Audiol. **7**, 130–136.

Chertoff, M. E., and Brownell, W. E. (**1994**). ''Characterization of cochlear outer hair cell turgor,'' Am. J. Physiol. **266**, C467–C479.

Cohen, J. (**1988**). ''The effect size index: f,'' in *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Erlbaum, Hillsdale, NJ), pp. 274–288.

Comis, S. D., Osborne, M. P., and Jeffries, D. J. (**1990**). ''Effect of furosemide upon morphology of hair bundles in guinea pig cochlear hair cells,'' Acta Otolaryngol. (Stockh). **109**, 49–56.

Corey, D. P., and Hudspeth, A. J. (**1979**). ''Ionic basis of the receptor potential in a vertebrate hair cell,'' Nature (London) **281**, 675–677.

Dallos, P. (**1986**). ''Neurobiology of cochlear inner and outer hair cells: Intracellular recordings,'' Hear. Res. **22**, 185–198.

Dallos, P. (**1973**). ''Nonlinear distortion,'' in *The Auditory Periphery: Biophysics and Physiology* (Academic, New York), pp. 391–464.

Dulon, D., Aran, J. M., and Schacht, J. (**1988**). ''Potassium-depolarization induces motility in isolated outer hair cells by an osmotic mechanism,'' Hear. Res. **32**, 123–130.

Durrant, J. D., Wang, J., Ding, D. L., and Salvi, R. J. (**1998**). ''Are inner or outer hair cells the source of summating potentials recorded from the round windows?'' J. Acoust. Soc. Am. **104**, 370–377.

Forge, A., and Brown, A. M. (**1982**). ''Ultrastructural and electrophysiological studies of acute ototoxic effects of furosemide,'' Br. J. Audiol. **16**, 109–116.

Gitter, A. H. (**1993**). ''Cell potential of isolated outer hair cells as a function of extracellular K+ concentration,'' ORL J. Otorhinolaryngol. Relat. Spec. **55**, 1–2.

Hibino, H., Horio, Y., Inanobe, A., Doi, K., Ito, M., Yamada, M., Gotow, T., Uchiyama, Y., Kawamura, M., Kubo, T., and Kurachi, Y. (**1997**). ''An ATP-dependent inwardly rectifying potassium channel, K_{AB}-2 (Kir4.1), in cochlear stria vascularis of inner ear: its specific subcellular localization and correlation with the formation of endocochlear potential,'' J. Neurosci. **17**, 4711–4721.

Hille, B. (**1992**). ''Potassium channels and chloride channels,'' in *Ionic Channels of Excitable Membranes*, 2nd ed. (Sinauer, Sunderland, MA), pp. 115–139.

Housley, G. D., and Ashmore, J. F. (**1992**). ''Ionic currents of the outer hair cells isolated from the guinea-pig cochlea,'' J. Physiol. (London) **488**, 73–98.

Johnstone, B. M., Patuzzi, R., Syka, J., and Sykova, E. (**1989**). ''Stimulus-related potassium changes in the organ of Corti of guinea-pig,'' J. Physiol. (London) **408**, 77–92.
Keppel, G. (**1991**). ''The single-factor within-subjects design,'' in *Design and Analysis: A Researcher's Handbook*, 3rd ed. (Prentice Hall, Upper Saddle River, NJ), pp. 344–366.
Kirk, D. L., and Yates, G. K. (**1998**). ''4-aminopyridine in scala media reversibly alters the cochlear potentials and suppresses electrically evoked oto-acoustic emissions,'' Audiol. Neurootol. **3**, 21–39.
Kros, C. J., and Crawford, A. C. (**1990**). ''Potassium currents in inner hair cells isolated from the guinea-pig cochlea,'' J. Physiol. (London) **421**, 263–291.
Lin, X., Hume, R. I., and Nutall, A. L. (**1995**). ''Dihydropyridines and verapamil inhibit voltage-dependent K^+ current in isolated outer hair cells of the guinea pig,'' Hear. Res. **88**, 36–46.
Mammano, F., and Ashmore, J. F. (**1996**). ''Differential expression of outer hair cell potassium currents in the isolated cochlea of the guinea-pig,'' J. Physiol. (London) **496**, 639–646.
Mills, D. M., and Rubel, E. W. (**1994**). ''Variation of distortion product otoacoustic emissions with furosemide injection,'' Hear. Res. **77**, 183–199.
Mills, D. M., Norton, S. J., and Rubel, E. W. (**1993**). ''Vulnerability and adaptation of distortion product otoacoustic emissions to endocochlear potential variation,'' J. Acoust. Soc. Am. **94**, 2108–2122.
Nenov, A. P., Norris, C., and Bobbin, R. P. (**1997**). ''Outwardly rectifying currents in guinea pig outer hair cells,'' Hear. Res. **105**, 146–158.
Ohnishi, S., Hara, M., Inoue, M., Yamashita, T., Kumazawa, T., Minato, A., and Inagaki, C. (**1992**). ''Delayed shortening and shrinkage of cochlear outer hair cells,'' Am. J. Physiol. **263**, C1088–C1095.
Patuzzi, R., and Moleirinho, A. (**1998**). ''Automatic monitoring of mechano-electrical transduction in the guinea pig cochlea,'' Hear. Res. **125**, 1–16.
Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (**1989**). ''Outer hair cell current and sensorineural hearing loss,'' Hear. Res. **42**, 47–72.
Pike, D. A., and Bosher, S. K. (**1980**). ''The time course of the strial changes produced by intravenous furosemide,'' Hear. Res. **3**, 79–89.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.
Russell, I. J., and Kössl, M. (**1991**). ''The voltage responses of hair cells in the basal turn of the guinea-pig cochlea,'' J. Physiol. (London) **435**, 493–511.
Russell, I. J., Cody, A. R., and Richardson, G. P. (**1986**). ''The responses of inner and outer hair cells in the basal turn of the guinea-pig cochlea and in the mouse cochlea grown *in vitro*,'' Hear. Res. **22**, 199–216.
Rybak, L. P. (**1986**). ''Ototoxic mechanisms,'' in *Neurobiology of Hearing: the Cochlea*, edited by R. A. Altschuler, D. W. Hoffman, and R. P. Bobbin (Raven, New York), pp. 441–454.
Rybak, L. P., and Morizono, T. (**1982**). ''Effect of furosemide upon endolymph potassium concentration,'' Hear. Res. **7**, 223–231.
Salt, A. N., Melichar, I., and Thalmann, R. (**1987**). ''Mechanisms of endocochlear potential generation by stria vascularis,'' Laryngoscope **97**, 984–991.
Salt, A. N., and Konishi, T. (**1986**). ''The cochlear fluids: perilymph and endolymph,'' in *Neurobiology of Hearing: the Cochlea*, edited by R. A. Altschuler, D. W. Hoffman, and R. P. Bobbin (Raven, New York), pp. 109–122.
Santos-Sacchi, J., and Dilger, J. P. (**1988**). ''Whole-cell currents and mechanical responses of isolated outer hair cells,'' Hear. Res. **35**, 143–150.
Saunders, J. C., Cohen, Y. E., and Szymko, Y. M. (**1991**). ''The structural and functional consequences of acoustic injury in the cochlea and peripheral auditory system: A five year update,'' J. Acoust. Soc. Am. **90**, 136–146.
Sewell, W. F. (**1984**). ''The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cats,'' Hear. Res. **14**, 305–314.
Shehata, W. E., Brownell, W. E., and Dieler, R. (**1991**). ''Effects of salicylate on shape, electromotility and membrane characteristics of isolated outer hair cells from guinea pigs,'' Acta Otolaryngol. (Stockh.) **111**, 707–718.
Soria, B. (**1998**). ''The biophysical basis of K^+ channel pharmacology,'' in *Ion Channel Pharmacology*, edited by B. Soria and V. Ceña (Oxford University Press, New York), pp. 167–185.
Syka, J., and Melichar, I. (**1985**). ''The effect of loop diuretics upon summating potentials in the guinea pig,'' Hear. Res. **20**, 267–273.
Takeuchi, S., and Ando, M. (**1999**). ''Voltage-dependent outward K^+ current in intermediate cell of stria vascularis of gerbil cochlea,'' Am. J. Physiol. **277**, C91–C99.
van Emst, M. G., Klis, S. F., and Smoorenburg, G. F. (**1996**). ''4-Aminopyridine effects on summating potentials in the guinea pig,'' Hear. Res. **102**, 70–80.
van Emst, M. G., Klis, S. F., and Smoorenburg, G. F. (**1997**). ''Identification of the nonlinearity governing even-order distortion products in the cochlear potentials,'' Hear. Res. **114**, 93–101.
Wang, J., Li, Q.-H., Dong, W.-J., and Chen, J.-S. (**1993**). ''Effects of K^+-channel blockers on cochlear potentials in the guinea pig,'' Hear. Res. **68**, 152–158.
Young, M. A. (**1993**). ''Supplementing tests of statistical significance: Variation accounted for,'' J. Speech Hear. Res. **36**, 644–656.
Zimmermann, U., and Fermin, C. (**1996**). ''Shape deformation of the organ of Corti associated with length changes of outer hair cell,'' Acta Otolaryngol. (Stockh.) **116**, 395–400.
Zucca, G., Valli, P., and Casella, C. (**1982**). ''Ionic mechanisms sustaining activity in ampulla receptors of the frog,'' Acta Otolaryngol. (Stockh.) **93**, 355–362.

Temporal pitch perception and the binaural system

Robert P. Carlyon[a)]
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom

Laurent Demany
Laboratoire de Neurophysiologie, UMR CNRS 5543, Université Victor Segalen, 146 rue Léo Saignat, F-33076 Bordeaux Cedex, France

John Deeks
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom

(Received 13 June 2000; revised 10 October 2000; accepted 25 November 2000)

Two experiments examined the relationship between temporal pitch (and, more generally, rate) perception and auditory lateralization. Both used dichotic pulse trains that were filtered into the same high (3900–5400-Hz) frequency region in order to eliminate place-of-excitation cues. In experiment 1, a 1-s periodic pulse train of rate Fr was presented to one ear, and a pulse train of rate $2Fr$ was presented to the other. In the ''synchronous'' condition, every other pulse in the $2Fr$ train was simultaneous with a pulse in the opposite ear. In each trial, subjects concentrated on one of the two binaural images produced by this mixture: they matched its perceived location by adjusting the interaural level difference (ILD) of a bandpass noise, and its rate/pitch was then matched by adjusting the rate of a regular pulse train. The results showed that at low Fr (e.g., 2 Hz), subjects heard two pulse trains of rate Fr, one in the ''higher rate'' ear, and one in the middle of the head. At higher Fr ($>$25 Hz) subjects heard two pulse trains on opposite sides of the midline, with the image on the higher rate side having a higher pitch than that on the ''lower rate'' side. The results were compared to those in a control condition, in which the pulses in the two ears were asynchronous. This comparison revealed a duplex region at $Fr > 25$ Hz, where across-ear synchrony still affected the perceived locations of the pulse trains, but did not affect their pitches. Experiment 2 used a 1.4-s 200-Hz dichotic pulse train, whose first 0.7 s contained a constant interaural time difference (ITD), after which the sign of the ITD alternated between subsequent pulses. Subjects matched the location and then the pitch of the ''new'' sound that started halfway through the pulse train. The matched location became more lateralized with increasing ITD, but subjects always matched a pitch near 200 Hz, even though the rate of pulses sharing the new ITD was only 100 Hz. It is concluded from both experiments that temporal pitch perception is not driven by the output of binaural mechanisms. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1342074]

PACS numbers: 43.66.Hg, 43.66.Mk, 43.66.Pn, 43.66.Qp [DWG]

I. INTRODUCTION

A. Aims and approach

Many decades of research have been devoted to the studies of pitch perception and of the lateralization of sounds (for recent reviews see Grantham, 1995; Houtsma, 1995). With a few notable exceptions, which will be described later in this Introduction, these two important auditory processes have been largely studied separately. In contrast, the present article investigates the relationship between the temporal mechanisms responsible for the perception of pitch (and, more generally, repetition rate), and the binaural mechanisms that underlie the perception of spatial location. This approach allows us to address a number of issues, including the extent to which temporal pitch perception is driven by the output of binaural lateralization mechanisms. In adopting it, we draw two crucial distinctions that constrain the potential mechanisms responsible for our findings. The first of these is that all of our stimuli consist of pulse trains that have been bandpass filtered in such a way that no spectral (''place of excitation'') cues to their repetition rate are available (Shackleton and Carlyon, 1994). That is, we are addressing the purely temporal processing of pulse rate. We do, however, study several different rates, ranging from those perceived as a series of separate clicks to those resulting in a single sound having a clear pitch. The second distinction is between a rate/pitch percept that is genuinely driven by the output of a binaural lateralization mechanism, and that which could arise simply from the listener attending to each ear separately. For example, if one transforms a diotic, filtered 200-Hz pulse train by playing alternate pulses to each ear, one hears two spatially separate pulse trains, each with a pitch of about 100 Hz. One would clearly be wrong to conclude from this demonstration that temporal pitch perception is driven by the output of binaural lateralization mechanisms, because the perceived pitches could be derived from the actual repetition rate of the pulses applied to each ear.

The experiments described here were motivated by some preliminary observations obtained with the stimulus shown in Fig. 1(a). It consists of two periodic trains of impulses, of repetition rates Fr and $2Fr$, respectively, presented to oppo-

a)Electronic mail: bob.carlyon@mrc-cbu.cam.ac.uk

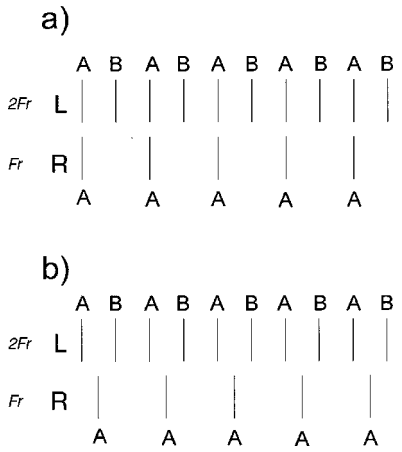


FIG. 1. (a) Schematic representation of the reference stimulus used in experiment 1, for a trial in the synchronous condition where the lower-rate (Fr) pulse train is presented to the right ear, and the higher-rate ($2Fr$) train is presented to the left ear. (b) As part (a), but for the asynchronous condition.

site ears over headphones. (The figure shows the Fr stimulus being presented to the right ear and the $2Fr$ stimulus to the left, but the observations and formal experiments were performed both with this and with the opposite configuration). The repetition rate in different conditions ranges from a few Hz, where the percept of each pulse train presented in isolation is of a series of separate clicks, up to 200 Hz, where listeners hear a single sound having a clear pitch. This change from a series of separate clicks to a single ''buzz'' can be interpreted as an increase in temporal ''binding''[1] between successive pulses as the interval between them decreases. The informal observations revealed that an interesting pattern of percepts occurs when both pulse trains are played together. At low rates, another form of binding takes place, in that simultaneous clicks in the two ears (marked ''A'' in the figure) fuse to form a centered percept. At these low rates ($Fr = $ a few Hz), two series of pulses, both with a rate of about Fr, can be heard: one near the middle of the head and the other near one ear. (At these low rates subjects can also choose to perceive a single pulse train whose location alternates between the two locations, but here we are primarily concerned with the rate heard at each location.) Note that the existence of two perceived pulse trains, each with a rate of Fr, does indicate that at least one of these perceived rates is driven by binaural processes. Unlike our example in which alternate pulses are presented to each ear, it is not the case that each ear is presented with a pulse train that corresponds to the perceived rate. As Fr is increased, however, the strength of this binaural fusion decreases, so that at $Fr = 100$ Hz one hears two pulse trains, one on either side of the midline. At the same time, another difference between the two percepts emerges, in that the pitch of the sound on the side of the head receiving the ''$2Fr$'' pulse train is approximately an octave above that receiving the ''Fr'' pulse train.

One way of interpreting this pattern of informal results is that there exists a trade-off between perceptual binding across space and binding over time, and that, as Fr increases, the balance between the two processes swings towards the latter. This general idea, that different organizational principles compete with each other, has recently been implemented in a computational approach to auditory scene analysis (Godsmark and Brown, 1999). Ways in which this competition might take place, together with alternative interpretations, will be considered in Sec. IV. For the time being, it suffices to note that increasing Fr has two effects: it produces a change in spatial perception, and it introduces a difference in perceived rate between the two concurrent pulse trains.

The above informal observations allow one to rule out some schemes whereby temporal and spatial aspects of perception might be related. For example, one might intuitively assume that diotic pulses [labeled ''A'' in Fig. 1(a)] would always fuse into a centered binaural image, whose perceived rate/pitch would then correspond to that of these centered pulses. A separate rate/pitch analysis would then be applied to the remaining pulses (labeled ''B''), which would be perceived to one side of the head. Clearly, such a scheme is precluded by the informal reports of two lateralized images, with different perceived pitches, that occur at high Fr. Other schemes, consistent with the informal observations, can be evaluated by the more detailed measures described in this paper. These schemes will be described in Sec. II A, which outlines the rationale for our first experiment.

B. Related research

1. Pitch, lateralization, and integration across frequency

Some experiments have investigated the interaction between pitch and lateralization mechanisms for stimuli containing resolved spectral components. An important conclusion resulting from these studies is that differences in perceived location do not have a large effect on the integration of different components into a perceived pitch. For example, if two consecutive harmonics from two separate fundamentals are presented simultaneously, listeners are not markedly better at identifying the two fundamentals when the harmonics are segregated appropriately by ear than when each ear receives one harmonic from each fundamental (Beerends and Houtsma, 1989). Also, Darwin and Ciocca (1992) measured the shift in the pitch of a complex tone produced by mistuning one of its harmonics, and reported that this shift was only slightly reduced when the mistuned harmonic was presented contralaterally to the rest of the complex.

Despite the validity of the above conclusion, there is evidence that the auditory system can derive a pitch from a ''central spectrum,'' which itself is derived from the output of binaural mechanisms. For example, Culling *et al.* (1998b, 1998a) have argued that a modified version of Durlach's (1972) equalization-cancellation model can account for three types of dichotic pitch: Huggins pitch, binaural edge pitch, and Fourcin pitch. In each case, the model produces peaks in the central spectrum at frequencies where the outputs of the corresponding auditory filters are interaurally decorrelated. In the case of Fourcin pitch, the model produces a series of peaks which, when subjected to de Boer's (1956, 1976) ''pattern matching'' rule, can account for the reported pitch.

Culling *et al.* also showed that this procedure successfully accounted for the two ambiguous pitches that are generated by some versions of the Fourcin pitch stimulus.

The results of all of the above experiments are consistent with the following scheme. For stimuli containing resolved harmonics, a central pitch mechanism operates on all available components, irrespective of their ITD, ILD, or ear of entry. It also operates on ''components,'' derived from binaural lateralization mechanisms, that occur in the central spectrum. The auditory system then groups together harmonically related components, and combines information on their ITDs and ILDs in order to obtain a perceived location for that group (Darwin and Carlyon, 1995).[2] Note, however, that a similar conclusion is unlikely to apply to the combination of pulses across time by the purely temporal pitch mechanisms being investigated here: as described in our initial example, assigning alternate pulses in a pulse train to opposite ears does have a dramatic effect on its pitch, reducing it by about an octave.

2. Perceived rate of alternating pulses

A second type of related research concerns the perceived rate of pulses that alternate between the two ears. Axelrod *et al.* (1967) used the method of limits to measure the rate of a dichotically alternating pulse train that was perceived as having the same rate as a monotic pulse train. Their approach differed from ours in that they asked subjects to judge the overall (total) rate of the dichotic train, rather than the perceived rates of two separately localizable images. They reported that the ratio between the perceived rates of the dichotic and monotic stimuli was close to 1 at a rate of 1 Hz, but dropped smoothly to a value of 0.6 at rates between 20 and 40 Hz (the highest tested). Subsequently, Huggins (1974) asked subjects to adjust the rate of a diotic pulse train so that its perceived total rate (''of pulses into the head'') matched that of a dichotically alternating train. The matched rate was equal to that of the dichotic stimulus at rates below 10 Hz, above which there was a sharp transition to subjects adjusting the diotic train to half the total dichotic rate. He argued that Axelrod *et al.* may have missed this sharp transition by averaging their results over a large number of subjects.

Finally, Akerboom *et al.* (1983) measured subjects' reaction times to the termination of a train of brief 800-Hz tone pips that was presented either monotically or in a dichotically alternating pattern. They reasoned that the reaction time would consist of a fixed ''response activation time,'' plus the duration of the perceived interpulse interval. This reasoning was based on the assumption that subjects would have to wait until the perceived interpulse interval (''subjective onset asynchrony,'' SOA) had passed before realizing that the pulse train had ended. They reported that reaction times were 25 ms longer for the alternating compared to the monotic sequence at all repetition rates studied, which ranged from 0.47 to 25 Hz, and concluded that the dichotic alternation increased the SOA by this amount. However, it seems likely that reaction times were determined not only by the mean SOA but also by the variance of its internal representation: subjects may delay responding until they can decide with a given degree of confidence that the pulse train has ended. As Akerboom *et al.* (1983) cite evidence (Nakao and Axelrod, 1976) consistent with this internal variance being larger for alternating compared to monotic pulse trains, it seems that their reasoning may well be flawed. In contrast, the present experiments, which used a pitch-matching procedure, allowed the mean of the internal representation of the stimulus period to be measured separately from the variance of that representation—these two values being reflected in the mean and variance of the pitch matches, respectively.

II. EXPERIMENT 1

A. Method

1. Rationale

As discussed in the Introduction, our informal observations allowed us to rule out some accounts of the way in which temporal and binaural processing may be related. The aim of experiment 1 was, in addition to quantifying the preliminary observations, to evaluate a number of alternative schemes. According to one of these, increasing *Fr* would decrease the tendency for diotic pulses to be fused, perhaps due to increased perceptual binding between temporally adjacent pulses in the same ear, but the perceived rate/pitch would still be determined from the output of this binaural analysis. According to this hypothesis, the changes in perceived location and in relative rate should always be consistent: there should be no value of *Fr* at which one simultaneously hears one pulse train in the middle of the head (''A'' pulses binaurally fused) and the remaining pulse train as having a rate higher than *Fr*. Another hypothesis is that the perception of rate is driven by the output of the binaural system only when the rate is too low to evoke a sense of pitch. That is, the perception of low rates operates on pulses sharing the same perceived location, whereas temporal pitch mechanisms operate on the input to each ear. If this is so, the value of *Fr* at which one starts to hear temporal differences between the two pulse trains should correspond to the minimum pulse rate which can support a perception of pitch. If one uses a loose definition of pitch—''that subjective ordering of sound which admits a rank ordering from low to high'' (Ritsma, 1963)—this value is approximately 19 Hz (Guttman and Pruzansky, 1962).[3]

2. Experimental conditions

Experiment 1 consisted of two conditions. In the ''synchronous'' condition, the reference stimulus was a dichotic, bandpass-filtered pulse train similar to that illustrated in Fig. 1(a). We obtained estimates of the perceived rate/pitch and lateralization of the percepts evoked by this stimulus, as a function of the baseline rate (*Fr*). The intention was to provide information on the ways in which perceived rate/pitch and location could be affected by competition between sequential binding of successive pulses presented to one ear, and binaural binding between simultaneous pulses in opposite ears. However, the judgment of the pitch of a sound presented to one ear can be influenced by the presence of another sound in the opposite ear, even in the absence of any binaural fusion between them (Thurlow, 1943; Terhardt,

1977). In addition, the presence of a sound at one location can affect lateralization judgments at another (McFadden and Pasanen, 1976; Warren and Bashford, 1976; Dye *et al.*, 1996; Heller and Trahiotis, 1996), presumably due to central mechanisms unrelated to the effects of temporal binding on perceived location. For these reasons, we included an ''asynchronous'' condition, in which the reference stimulus was as shown in Fig. 1(b). It was similar to that used in the synchronous condition, except that the pulses in the lower rate ear were delayed by $0.25/Fr$ s, so that no two pulses were ever simultaneous in the two ears. (The closest they ever got was a 2.5-ms separation when $Fr = 100$ Hz.) This condition controlled for any ''nonspecific'' effects of one pulse train on the perception of a concurrent train in the other ear that might occur even in the absence of any binaural fusion.

3. General method

In both the synchronous and asynchronous conditions, each stimulus had a total duration of 1 s, and the rate (Fr) of the slower of the two pulse trains was, in different trials, 2, 3.125, 6.25, 12.5, 25, 50, or 100 Hz. The overall level of this ''lower'' pulse train was always 72.5 dB SPL, and all pulse trains were bandpass filtered between 3900 and 5400 Hz (Kemo VBF25.03; attenuation 48 dB/octave) to remove any frequency components that could be resolved by the peripheral auditory system (Shackleton and Carlyon, 1994). All stimuli were presented against a background of 10-kHz-wide pink noise, having a spectrum level of 8.5 dB SPL at 4 kHz. Sennheiser HD414 headphones were used.

On each trial, subjects matched the perceived location of either the ''rightmost'' or ''leftmost'' sound that they could hear, and then matched its perceived rate/pitch. It was stressed that this sound could be perceived either in the middle of the head, completely to one side, or anywhere in between. A light on either the right- or left-hand side of the response box reminded subjects which sound to match. Prior to making any matches, they could listen to the stimulus as many times as they liked, and then press a button to start the location matching. This began with a presentation of the dichotic pulse train preceded, 500 ms earlier, by a white noise that had been passed through the same bandpass filter as the pulse train. The noise was diotic except for an interaural level difference (ILD) drawn at random from a rectangular distribution between ± 25 dB. Its overall level in one ear was 72.5 dB SPL (equal to that of the lower-rate pulse train), and was reduced in the opposite ear by an amount equal to the ILD. The subject could then press one of four buttons to adjust the ILD of the noise for the next presentation, by either ± 1 or ± 4 dB. The dichotic pulse trains and the noise were then presented again, the subject made a new adjustment, and this process was repeated until s/he was satisfied that the location of the noise matched that of the sound (leftmost or rightmost) appropriate for that trial. Subjects were encouraged to ''bracket'' the perceived location of the pulse train before accepting a match. The final ILD of the noise was used as a measure of perceived lateralization.

In the pitch-matching stage of the trial, the adjustable stimulus was a pulse train that was diotic, except for an ILD equal to that obtained in the lateralization-matching stage. This ''quasidiotic'' pulse train was presented 500 ms before the dichotic pulse train. On its first presentation in the pitch-matching stage, its rate was drawn at random from a distribution which had limits of $0.5Fr$ and $4Fr$, and which was rectangular on log(Fr) vs probability coordinates. The subject could then increase or decrease the rate of the quasidiotic pulse train for the next presentation, by a factor of 1.1 or 1.4, so that its perceived rate more nearly matched that of the currently attended component of the dichotic mixture. (The ILD imposed on the quasidiotic pulse train meant that it had approximately the same perceived location as that component, thereby facilitating the match.) The dichotic and quasidiotic pulse trains were then presented again, and the matching process continued until the subject was satisfied that the perceived rate of the quasidiotic stimulus approximated that of the appropriate pulse train in the mixture. Again, subjects were told to bracket the match before terminating the trial. The final rate of the quasidiotic pulse train was used as an estimate of the perceived rate or pitch of the appropriate pulse train in the dichotic mixture. Note that at very low values of Fr, subjects could perform the match either by simply counting the number of pulses heard in the appropriate location, or by estimating the interval between these pulses. Although the written instructions for all values of Fr were to ''match the rate or pitch'' of the stimulus, no special attempt was made to discourage the listener from using a ''counting'' strategy.

As discussed in Sec. II A 1, the judgment of the pitch of a sound presented to one ear may be influenced by the presence of another sound in the opposite ear, even in the absence of any binaural fusion between them. The asynchronous condition controlled for such nonspecific effects, and, when describing the results, we plot not only the raw data but also the ratio of the matches obtained in the synchronous and asynchronous conditions. Similarly, the asynchronous condition also controlled for nonspecific effects of one sound on the perceived location of another, simultaneous sound. Therefore, when describing the lateralization matches, we describe not only the raw data but also the difference (in dB) between the ILD matches obtained in the synchronous and asynchronous conditions.

B. Procedure and preliminary observations

Each value reported here is derived from the arithmetic (ILD) or geometric (matched rate) mean of ten matches for each subject and condition. In each 2-h session the listener was instructed to attend to the same relative location (leftmost or rightmost sound). Blocks of trials alternated between the two attended relative locations. Half of the blocks started with, in the following order: (i) a match with $Fr = 2$ Hz in the synchronous condition, lower pulse train presented to the attended side; (ii) as (i), but for the asynchronous condition; (iii) as (i), but with the higher pulse train presented to the attended side; (iv) as (iii), but in the asynchronous condition. This sequence was repeated at increasingly higher values of Fr, until $Fr = 100$ Hz was reached. This order was reversed for the other half of the blocks of trials. For subjects GN and TP, the matches with $Fr = 2$ Hz and the matches to the

lower-rate pulse train at $Fr = 3.125$ Hz took place after the matches to the other stimuli had been completed.

When piloting this experiment it became apparent that two subjects, who were highly experienced in other (purely monaural) psychoacoustic tasks, found it extremely difficult to perform the pitch matches at the highest values of *Fr* tested. Those pitch matches were extremely variable, both in the synchronous and asynchronous conditions, even though the subjects could make accurate pitch matches to monotic stimuli presented to either ear. This finding, which we interpret as a failure of selective attention, will not be pursued here. We did not continue measurements with those two subjects. At high values of *Fr* a third subject, GN, could make accurate matches to stimuli perceived on the left, but not those on his right, despite absolute thresholds at 4590 Hz (the geometric center of the filter passband used here) that differed by less than 1 dB between the two ears. Again, he could make highly accurate matches to monotic stimuli in either ear. This subject was one of the three who participated in the main experiment, but his matches were obtained only to the leftmost sound heard in each dichotic mixture. All subjects had absolute thresholds, in both ears and at all audiometric frequencies, within 20 dB of laboratory norms for 16 healthy young subjects.

C. Results

1. ILD matches

The raw (untransformed) lateralization matches are plotted for the ''leftmost'' and ''rightmost'' percepts in Figs. 2(a) and (b), respectively. Throughout this article, the sign of the ILD is defined as positive whenever the matching noise is more intense on the ''side'' to which the subject is matching (leftmost or rightmost sound). Accordingly, the ordinates are labeled as ''absolute'' ILD matches. Note also that it is not the case that each panel combines matches for the two stimuli that were present on any one trial (e.g., a $2Fr$ train on the left and an *Fr* train on the right): rather, the matches to a $2Fr$ train on the left and to an *Fr* train on the left are combined in each panel of Fig. 2(a) with corresponding rightmost matches being plotted in Fig. 2(b).

For the synchronous condition (open symbols), the informal observations described in the Introduction are largely confirmed. The higher-rate pulse train (squares) always gives rise to a lateralized percept (large ILD). The lower-rate pulse train (triangles) is perceived near midline at low *Fr*, and becomes more lateralized at high *Fr*. Note, however, that even at the highest rate its perceived location corresponds to an ILD of at most 5–10 dB, and it is never as lateralized as the higher-rate stimulus. In addition, some aspects of the results obtained in the synchronous condition are also obtained in the asynchronous condition, in which the large (2.5–125-ms) temporal separation between adjacent pulses in the two ears makes explanations in terms of binaural fusion—and its modification by within-ear temporal binding—unlikely. One such effect—the decrease in the lateralization of the leftmost higher-rate pulse train [open and filled squares, Fig. 2(a)] as *Fr* increases—is reminiscent of previous reports that the location of a sound can be ''pulled'' towards that of a contralateral stimulus (Warren and Bashford, 1976; Heller and Trahiotis, 1996). Others include the nonmonotonic trend observed in listener TP's right-ear matches to the lower-rate pulse train [open and filled triangles, Fig. 2(b)], and the dip at $Fr = 3.125$ Hz in JD's lateralization function to the higher-rate pulse train in his left ear [open and filled squares, Fig. 2(a)]. These features are removed in the transformed data shown by the symbols in Figs. 2(c) (leftmost sounds) and (d) (rightmost sounds), where the ILD matches for the asynchronous condition have been subtracted from those in the synchronous condition [the meaning of the solid lines with no symbols in Fig. 2(c) will be revealed later]. Because all asynchronous pulse trains were heard to one side of the head, a value of zero on these coordinates corresponds to a lateralized percept, and a large negative value corresponds to a sound close to the middle of the head. The transformed plots are in general more orderly than the raw data, and show a smaller variation across listeners and ears. They indicate that the synchronicity of the ''A'' pulses [Fig. 1(a)] in the higher-rate pulse train affects the perceived location of the lower-rate train (triangles) at low *Fr*, and that this effect decreases but does not disappear as *Fr* is raised to 100 Hz.

One limitation of the method by which we have transformed the lateralization data, by subtracting the matches in the asynchronous condition from those in the synchronous condition, arises from the existence of some extreme ILD matches in the raw data [Figs. 2(a), (b)]. Yost (1981) has shown that perceived lateralization increases linearly with ILD only up to 15 dB, with much smaller changes observed as the ILD is increased further. In our experiment 1, matches to the lower-rate stimulus in the asynchronous condition [filled triangles, Figs. 2(a), (b)] were often substantially larger than 15 dB, whereas those in the synchronous condition (open triangles) were closer to zero. This means that subtracting the asynchronous matches from the synchronous ones may not give an accurate representation of the difference in perceived lateralization in the two conditions. To control for this, we repeated the differencing operation, but with a ''ceiling'' of 15 dB applied to the raw ILD matches. The results of this new transform are shown in Figs. 2(e) and (f) for the leftmost and rightmost percepts, respectively. The general pattern is very similar to that shown in Figs. 2(c) and (d), the main difference being the elimination of a nonmonotonicity in listener JD's rightmost matches to the lower-rate stimulus [triangles, compare Figs. 2(d) and (f)].

2. F0 matches

Figures 3(a) and (b) show, respectively, the pulse rate matched to the leftmost and rightmost sounds in the dichotic mixture. The matches have been divided by *Fr* but are otherwise untransformed. Again, the results obtained in the synchronous condition (open symbols) generally confirm the informal observations described in the Introduction. The lower-rate stimulus (open triangles) is perceived at approximately its correct rate at all values of *Fr*. The higher-rate stimulus, which has a rate equal to $2Fr$, is perceived at *Fr* at low rates and an increasingly higher relative value as *Fr* increases to 100 Hz. Once again, however, the data in the

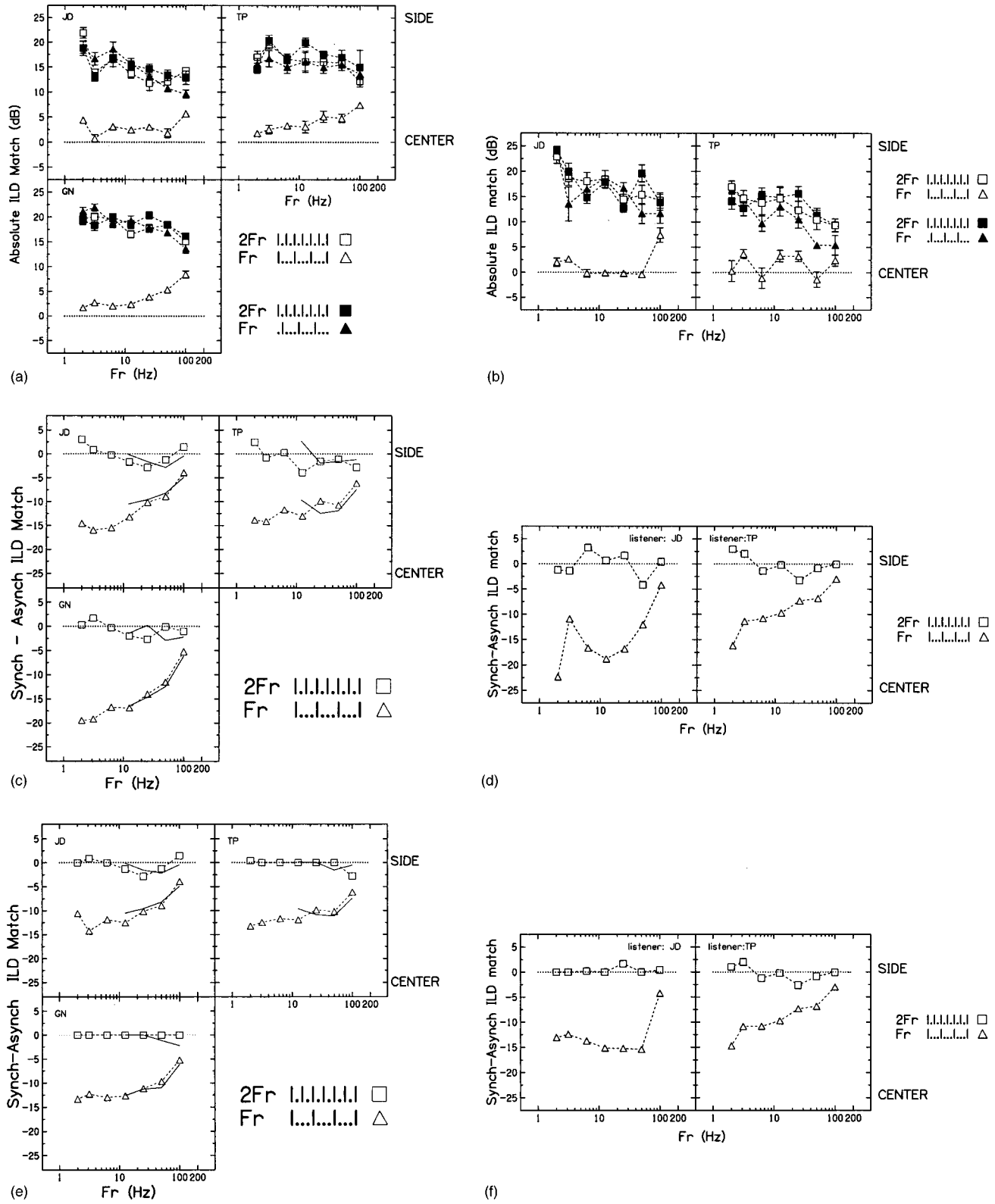


FIG. 2. (a) Untransformed ILD matches to the leftmost percept in experiment 1. (b) As (a), but for the rightmost percept. (c) Transformed ILD matches to the leftmost percept in experiment 1. (d) As (c), but for the rightmost percept. Parts (e) and (f) show the transformed data calculated with the raw scores limited to a maximum of 15 dB. The error bars in parts (a) and (b) show ± one standard error. The solid lines with no symbols in parts (c) and (f) are for the supplementary experiment with the reduced pulse-train level and the additional low-pass noise. In the key, instances where a pulse occurs in a given sequence (*Fr* or 2*Fr*) are indicated by a vertical line, and instances where a pulse occurs in the opposite ear are shown by a period.

asynchronous condition reveal the operation of central processes, which are unlikely to reflect an influence of binaural fusion on the perception of perceived rate/pitch. First, for the leftmost percepts [Fig. 3(a)], there is a tendency for the perceived rate of the asynchronous higher-rate stimulus (filled squares) to decrease relative to *Fr* as *Fr* is increased from 2 to 12.5–25 Hz. This tendency is even more marked for the rightmost percepts [Fig. 3(b)]. In some cases (JD and TP,

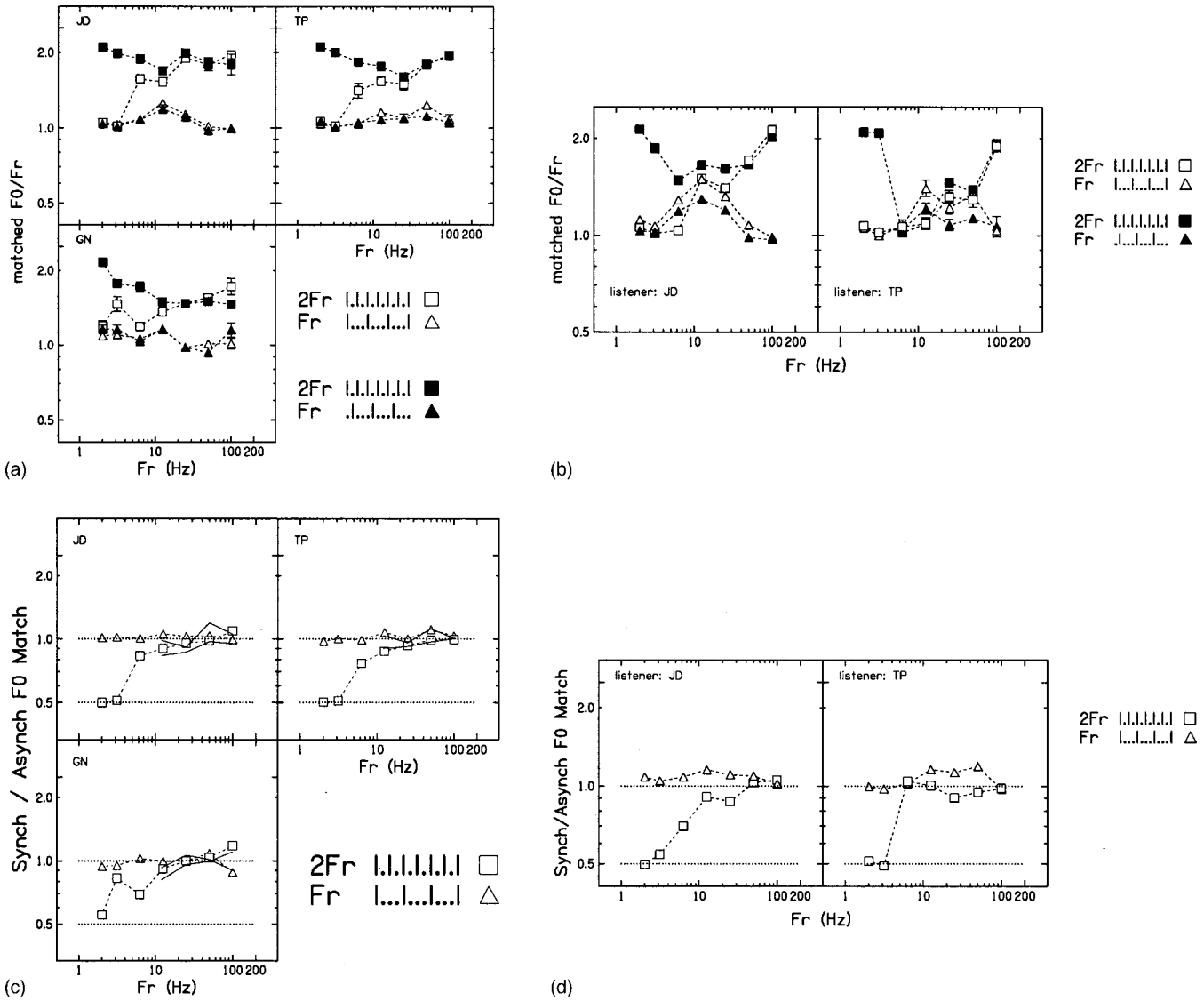


FIG. 3. (a) Rate ($F0$) matches to the leftmost percept in experiment 1, divided by *Fr* but otherwise untransformed. (b) As (a), but for the rightmost percept. (c) Transformed $F0$ matches (synch/asynch) to the leftmost percept in experiment 1. (d) As (c), but for the rightmost percept. The error bars in parts (a) and (b) show ± one standard error. The solid lines with no symbols in part (c) are for the supplementary experiment with the reduced pulse-train level and the additional low-pass noise.

leftmost percept), the matched rate increases as a proportion of *Fr* as *Fr* is raised further, from 25–100 Hz. In addition, there is some evidence of nonmonotonicities in the function for the lower-rate stimulus in the leftmost-percept data for listeners GN and JD [triangles, Fig. 3(a)], which are very similar in the synchronous and asynchronous conditions.

A more consistent and orderly pattern of results is shown in the plots of the transformed rate/pitch matches in Figs. 3(c) and (d). [Again, the transformed data are shown by symbols, and the solid lines with no symbols in Fig. 3(c) will be described later.] Here, the matches in the synchronous condition have been divided by those in the asynchronous condition, so that a value of unity indicates no effect of binaural fusion on perceived rate or pitch. For the lower-rate stimulus (triangles), the ratio between the matches in the synchronous and asynchronous conditions remains close to unity at all values of *Fr*. For the higher-rate stimulus (squares), the synchronicity of the contralateral ''A'' pulses [Fig. 1(a)] halves the perceived rate when $Fr = 2$ Hz, but ceases to have any effect once *Fr* is equal to or greater than 12.5–25 Hz.

3. General discussion

Two conclusions are suggested by the results of experiment 1. First, binaural fusion between pulses in opposite ears has no effect on the perceived rate of pulse trains once that rate is high enough for listeners to perceive a pitch. This occurs once subjects are matching to a higher-rate pulse train with a rate (equal to $2Fr$) of 25–50 Hz or higher. Second, there exists a duplex region, at $Fr > 12.5–25$ Hz, where the pitch of the higher-rate stimulus [squares in Figs. 3(c) and (d)] is unaffected by binaural fusion, but where the perceived location of the lower-rate stimulus is at least partially af-

fected by that fusion [triangles in Figs. 2(c) and (d)].[4] Before accepting these conclusions, however, some further considerations must be taken into account.

One factor that complicates the interpretation of the duplex region concerns the pitch matches to the higher-rate pulse train at $Fr > 25$ Hz. Although these matches were unaffected by whether or not the pulses in the two ears were synchronous, the matched values were not consistently equal to $2Fr$. Rather, the raw data shown in Figs. 3(a) and (b) indicate that these values were often below $2Fr$ in both the synchronous and asynchronous conditions. Although we interpreted this in terms of a common central factor reducing pitch matches in one ear when a lower-pitched sound is in the other, alternative interpretations are possible. Specifically, it could be that pitch matches at $Fr > 25$ Hz in the synchronous condition are reduced by the binaural fusion between the ''A'' pulses shown in Fig. 1(a), and that they are coincidentally also reduced in the asynchronous condition by a similar amount, but by some other mechanism, which, in turn, does not operate when the pulses are synchronous. One way of evaluating this is to look for cases where the match to the higher-rate pulse train is very close to $2Fr$, and then see whether the corresponding lower-rate train is localized in the center of the head or in the opposite ear. Two instances, where the 95% confidence limits of the higher-rate match encompass $2Fr$, can be seen in Fig. 3. One of these occurs for listener TP's match in his left ear, at $Fr = 100$ Hz [Fig. 3(a), open square, occluded by filled square]. The corresponding right-ear stimulus [Fig. 2(b), open triangle] was matched to an ILD of only 3 dB. The other occurs for JD in his right ear, again at Fr = 100 Hz [Fig. 3(b)], open square. The corresponding lower-rate stimulus, in his left ear, was matched to an ILD of 5 dB [Fig. 2(a) triangle]. Both of these comparisons indicate that it is possible for the higher-rate stimulus to have a pitch of $2Fr$, and for the lateralization of the corresponding lower-rate pulse train to be affected by simultaneous pulses in the opposite ear.

Finally, we considered the possibility that that the pitch of the filtered pulse trains may have been influenced by the presence of combination tones (CTs). Recently, Wiegrebe and Patterson (1999) have shown that the pitch of amplitude-modulated noise, bandpass filtered into a high frequency region, can be greatly weakened by a low-pass noise whose passband covers the modulation rate. They argued that the pitch of such modulated noises is dependent on a CT having a frequency equal to the modulation rate, and that the level of this CT is higher than would be predicted by traditional models of cochlear nonlinearities. Although they only measured the CTs produced by modulated noises, the fact that our bandpass-filtered pulse trains would have also generated modulated responses on the basilar membrane suggests that they too may have given rise to CTs at the modulation rate (equal to the repetition rate). We were particularly concerned that a CT at the modulation rate may have affected judgments of pitch but not of lateralization, and so performed the supplementary experiment described below.

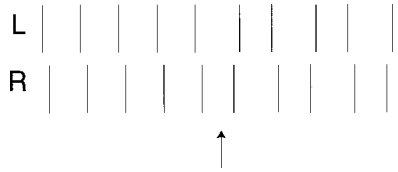


FIG. 4. Schematic representation of the reference stimulus used in experiment 2, for a trial where the pulse train lags in the right ear for the first 700 ms. The point where the ITDs start to alternate is indicated by an arrow. In both parts of the figure, only a few pulse pairs are shown, and the ITDs are exaggerated, for reasons of clarity.

D. Supplementary experiment

The supplementary experiment repeated the measures obtained at a subset of values of *Fr* (12.5, 25, 50, and 100 Hz) in both the synchronous and asynchronous conditions of experiment 1. The levels of the pulse trains were reduced by 15 dB, and an extra continuous low-pass noise (cutoff 1000 Hz, Kemo VBF25.01, attenuation 100 dB/octave) was added to the pink noise background. The spectrum level of this noise was 33 dB SPL. For these stimuli, the level of each spectral component of a 200-Hz pulse train within the filter passband was 51 dB SPL. Given that the critical ratio at 200 Hz is approximately 18 dB (Zwicker *et al.*, 1957), the low-pass noise would have masked any CT having a level lower than that of the primaries. The results of that experiment are shown by the solid lines without symbols in Figs. 2(c), (e), and 3(c). They are very similar to those obtained in the main experiment (open symbols), indicating that CTs are unlikely to have had a strong influence on the pitch or lateralization matches.

III. EXPERIMENT 2

A. Rationale and overview

The aim of experiment 2 was to provide a further test of whether temporal pitch perception is driven by the output of binaural lateralization mechanisms. As in experiment 1, subjects adjusted the ILD of a noise to match the perceived location of one part of a dichotic pulse train, and then adjusted the rate of a pulse train that was diotic (except for an ILD) to match its pitch. A new reference stimulus, illustrated in Fig. 4, was used. The first 700 ms of this stimulus consists of a dichotic pulse train with a constant ITD leading one ear. After 700 ms the ITD starts to alternate, switching sign between subsequent pairs of pulses. Subjects were instructed to make location and pitch matches to the ''new'' sound that they heard coming on halfway through the 1400-ms stimulus. When the ITD during the first 700 ms was leading in the left ear, they were told that the new sound would be to the right of the original sound, and vice versa. The nominal pulse rate of the stimulus, defined as the pulse rate that would have occurred had a zero ITD been used, was always 200 Hz. The rationale was that if the temporal pitch mechanism is driven by the output of the lateralization process, subjects should match to a pitch of about 100 Hz, because only every other dichotic pulse pair has an ITD corresponding to the new location. If, on the other hand, the processes underlying the estimation of pitch and location are separate,

then subjects may assign a location corresponding to the new ITD, but a pitch corresponding to the input to one ear. Note that paradoxical percepts of this general sort have previously been reported for stimuli where spectral cues to pitch are available, and where streaming by pitch and location have been pitted against each other (Deutsch, 1974; Efron and Yund, 1974; Yund and Efron, 1975; Deutsch, 1976).

The experiment was performed with a range of ITDs, from 0.6 to 1.4 ms. Preliminary experiments revealed that these quite large values were needed for two binaural images to be produced that were sufficiently far apart for subjects to be able to focus on them separately. This may be related to other reports showing that a sound may be perceived more centrally when accompanied by a second stimulus presented either to the opposite ear (Warren and Bashford, 1976) or diotically (Heller and Trahiotis, 1996). Of particular relevance is Dye *et al.*'s (1996) finding that, when subjects are asked to judge the laterality of a target tone, while ignoring that of a distracting tone, their judgments are often biased by the ITD at which the distractor is presented.[5] If a similar phenomenon applied to our stimuli, it would result in larger ITDs being needed to lateralize the different percepts produced by the pulse trains than would suffice for stimuli containing a single, unambiguous, ITD.

A potential problem arising from the large ITDs used in experiment 2 is that they might prevent the pulses in the two ears from fusing binaurally, causing subjects just to treat the two inputs as separate events. This in turn would lead to matches near 200 Hz, but one could not conclude that the two pulse trains were being localized by binaural mechanisms (because subjects may have matched both the location and the pitch to the input to each ear separately). By using a range of ITDs we were able to check for this possibility: If lateralization is proceeding via binaural mechanisms, then the ''new'' stimulus should become progressively more lateralized as ITD increases. This check was further facilitated by the presence of the first 700 ms of the stimulus, which contained a constant ITD equal in absolute value to the alternating ITD at the end of the stimulus. If the ITDs were sufficiently large to cause ''splitting,'' then this should have occurred throughout the stimulus and subjects would not be able to follow the instruction ''match to the sound in the new location.'' A further advantage of the leading 700 ms was that it promoted some perceptual segregation of the sounds in the two locations, thereby helping subjects to focus on the new sound.

Finally, it is worth noting that the stimulus described in Fig. 4 is not perfectly isochronous in each ear, but instead alternates between interpulse intervals that are $\Delta t/2$ ms longer and shorter than the nominal period. In a preliminary experiment we asked subjects to match an isochronous pulse train to a diotic stimulus that contained the same long–short–long–short interval pattern in both ears. As the deviation from isochrony increased, we found that the pitch match decreased steadily, as if subjects were matching to the longer of the two intervals (cf. Kaernbach and Demany, 1998, p. 2304). This finding, although interesting in itself, was tangential to our main purpose. In the pitch-matching part of the experiment, therefore, the adjustable stimulus also alternated between long and short intervals, corresponding to $V(P \pm \Delta t/2)$ ms, where P was the nominal period and V was the parameter that was adjusted. The ''matched $F0$'' was then defined as being equal to $1/VP$. This stimulus was diotic except for an ILD that was determined during the first, location-matching part of each trial.

B. Method and stimuli

The method of stimulus generation, including the filter specifications, was the same as in experiment 1. Some differences in the stimuli and procedure were introduced. The level of the pulse trains was the reduced level used in the supplementary part of that experiment, and the same low-pass noise was present throughout. However, no pink noise background was used, as we suspected that the task would prove quite difficult and we did not want to exacerbate this difficulty by presenting stimuli at a low sensation level. Another difference was that both the reference and matching stimuli were gated on and off with 100-ms raised-cosine ramps. Also, because subjects were matching to the end of the reference stimulus, the matching sounds were presented 500 ms after, rather than 500 ms before, the reference. In order to maximize the reliability of the location matches, an additional procedure (cf. Buell *et al.*, 1991) was introduced at the beginning of each session. During this, subjects listened to a continuous diotic version of the bandpass-filtered noise that was used in the location-matching part of the study, and adjusted the headphones to produce a centered image.

During each trial, subjects listened to the reference stimulus at will, adjusted the ILD of a 700-ms bandpass-filtered noise to match the ''sound with the new location,'' and then adjusted the (nominal) rate of a pulse train that was diotic save for an ILD determined from the location-matching stage. In a given session, the initial 700 ms contained an ITD to one side only (always leading to the right or always leading to the left). In each 2-h session the subject would perform one trial at each ITD in either an ascending or descending pattern, repeat this in reverse order, and continue in a similar fashion until the session was over. A total of ten matches was made at each ITD and for each side (new sound heard on the left or right). Four listeners took part, including the three who had participated in experiment 1.

C. Results

The absolute values of the ILD matches for each subject are shown in Fig. 5. Three out of the four listeners perceived the new sound in a progressively more lateralized position as ITD increased (squares and triangles in Fig. 5), although there were some nonsystematic differences between listeners and ears. Listener TP showed a generally erratic pattern of ILD matches. Figure 6 shows that all listeners produced pitch matches close to 200 Hz, at all ITDs. As discussed in Sec. III A, a prerequisite for interpreting the pitch matches in experiment 2 is that the lateralization of the new (matched) sound should increase systematically with increasing ITD. The results of the three subjects who did show this pattern, averaged across ears, are shown in Fig. 7. It can be seen that

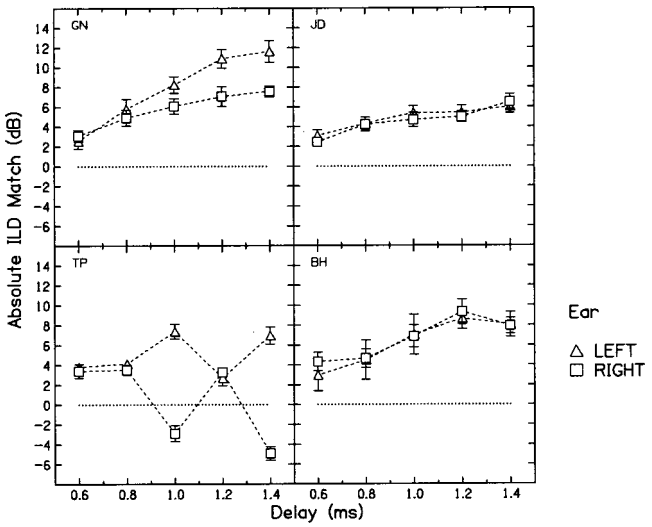


FIG. 5. Absolute ILD matches obtained in experiment 2, shown separately for each subject.

the perceived location of the new sound changes with ITD, indicating that the perceived location of the new sound is determined by binaural mechanisms, although there is some evidence of a leveling off at 1.4 ms. In contrast, listeners hear a pitch of about 200 Hz—double that which would be expected from the rate of the pulses having the new ITD. It therefore seems that the temporal pitch mechanism does not receive its input from the binaural mechanism responsible for processing ITDs.

The stimuli used in experiment 2 bear some resemblance to one condition of an experiment reported by Freyman *et al.* (1997). They presented subjects with pulse trains in which the ITD alternated between ±0.5 ms, and asked them to adjust an acoustic pointer so that its location matched that of the ''strongest'' image heard. For the range of pulse rates tested, 200–1000 Hz, subjects adjusted the pointer to the location that would be expected based on the first pair of pulses in the test stimulus. However, because of the nature of the instructions, this does not preclude the possibility that their subjects were in fact perceiving multiple images, with

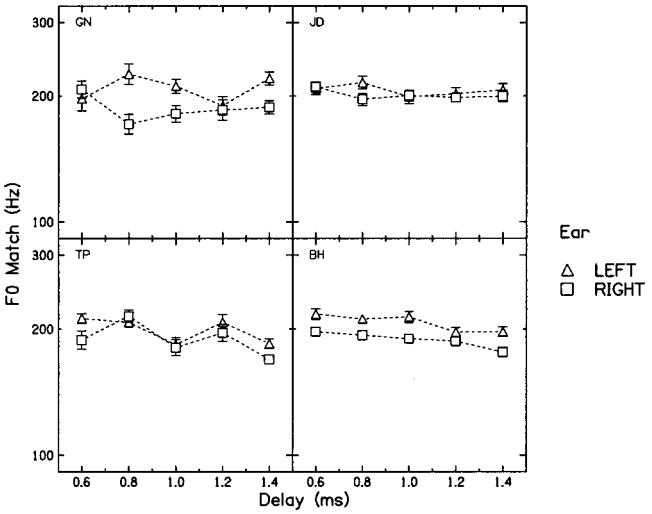


FIG. 6. *F*0 matches obtained in experiment 2, shown separately for each subject.

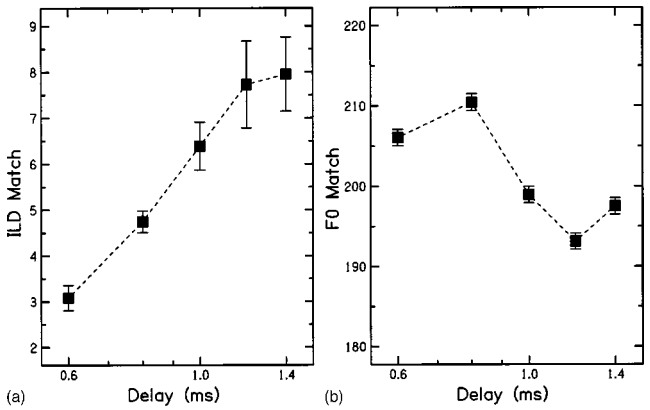


FIG. 7. Absolute ILD [part (a)] and *F*0 [part (b)] matches obtained in experiment 2, averaged across ears and across subjects GN, JD, and BH.

the one corresponding to the initial pulse(s) being strongest. As mentioned in Sec. III A, the ability of our subjects to perceive multiple images may have been facilitated by the fact that the initial (unambiguous) ITD was presented for 700 ms, before the ITDs started to alternate.

Two further points are worth making. First, in this experiment, listener GN could make accurate pitch matches to a stimulus heard on the right, even though he could not do so in experiment 1. We attribute this difference to the fact that, in experiment 2, the percept to which he was matching started 700 ms after the beginning of the entire stimulus. This may have helped him to focus attention on the appropriate binaural image. Second, the matches made by the subjects at each ITD were distributed unimodally. This is illustrated by Fig. 8, which shows the distribution of matches to a 1.2-ms ITD for subjects GN, JD, and BH, relative to the mean match made by each subject. The unimodality of the matches confirms that the ILD matches reported here were not due, for example, to subjects failing to form a binaural image on some trials, therefore making extreme ILD matches, and these extreme values being averaged with another, more centralized, subset of matches. It is important to rule out such an explanation, because, if the probability of

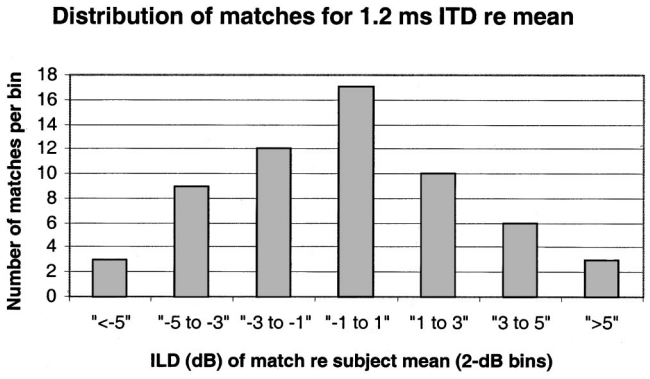


FIG. 8. Number of ILD matches to the 1.2-ms ITD of experiment 2, falling in each 2-dB bin. Prior to calculation, each ILD was subtracted from the mean for that subject and ear. This was done to avoid a situation where individual data showed bimodal matches, but where the peaks of the distributions of different subjects fell in different places, thereby obscuring the bimodality. The bin size of 2 dB was chosen to be smaller than the standard deviation (3.1 dB) of the normalized data set.

failing to form an image increased with increasing ITD, it could account for the pattern of increasingly lateralized average matches.

Finally, it is worth remarking on the range of ITDs used here. These values (0.6–1.4 ms) are considerably larger than the minimum ITDs that can be detected, even for stimuli filtered into high frequency regions ($>=4$ kHz), where the jnd ranges from about 60 μs (for 1/3rd-octave noise centered on 4 kHz: Koehnke *et al.*, 1995) to about 0.3 ms [for a single click filtered above 5 kHz (Yost *et al.*, 1971)]. They are also close to, or larger than, the maximum of about 0.65 ms that results in real-life situations from the time taken for sounds in this frequency range to travel the distance between the two ears (Kuhn, 1987). This might lead one to doubt whether the changes in location truly reflect the output of a binaural lateralization mechanism. However, in addition to the procedural control mentioned in Sec. III A, it is worth noting that, for high-pass-filtered noise stimuli, subjects can reliably make lateralization judgments based on ITDs up to 3 ms (Mossop and Culling, 1998). Another relevant finding is that, although subjects can detect small ITDs in the envelopes of high-frequency stimuli (Henning, 1974), ITDs in the high-frequency part of broadband stimuli have a much smaller effect on localization judgments than do ILDs in that part of the spectrum (Wightman and Kistler, 1989).[6] This in turn suggests that, for stimuli filtered into the high-frequency region and containing a zero ILD, large ITDs will be required to produce a substantial change in perceived location, even though much smaller ITDs may be detectable.

IV. DISCUSSION

A. Binding

In the Introduction, we discussed the informal observations that led to experiment 1 in terms of ''binding'' between pulses. If one assumes that the binding between temporally adjacent pulses becomes stronger as the interpulse interval is reduced, and that this sequential binding competes with the binaural fusion between the ''A'' pulses [Fig. 1(a)] in the two ears, then one can account *qualitatively* for the informally reported effects of *Fr* on perception. Specifically, as *Fr* increases, the binaural fusion is reduced by competition, thereby (i) increasing the lateralization of the lower-rate stimulus, and (ii) allowing the higher-rate stimulus to be perceived at something more closely approaching its true rate ($2Fr$). However, as discussed below, the quantitative data reported here would require such an account to be modified.

One way in which the competition between sequential and simultaneous binding could be implemented is if a proportion of each ''A'' pulse [Fig. 1(a)] in the lower-rate ear were fused with the contralateral A pulse, and the remainder formed part of a temporal sequence with the ''B'' pulses. This is illustrated in Fig. 9(a) for a stimulus in which the higher-rate train is presented to the left ear. The figure shows that the dichotic pulse train would decompose into two parts: a pulse train of rate *Fr* localized between midline and the right ear, plus a modulated pulse train having a carrier rate of $2Fr$ and lateralized to the left. When the pulse rate increases [Fig. 9(b)], less of each left-ear ''A'' pulse fuses with the

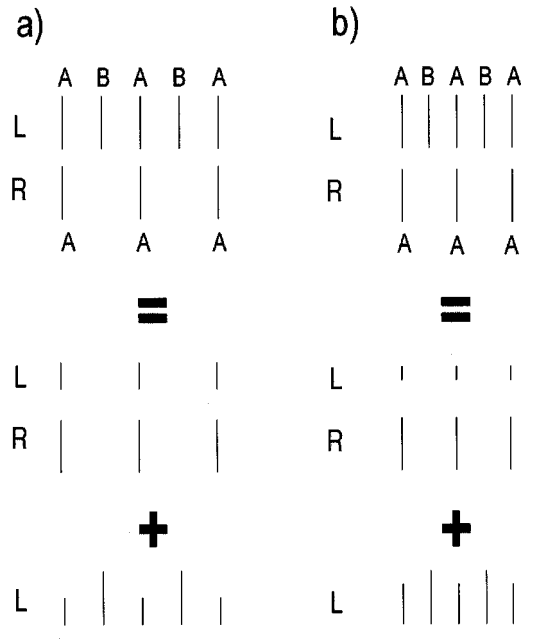


FIG. 9. Schematic representation of one way in which perceptual binding could account for the pattern of pitch/rate and location matches observed in experiment 1 for low *Fr* [part (a)] and high *Fr* [part (b)].

right-ear A pulse, which leads to two effects: the binaurally fused percept is more lateralized to the right ear, and the ''residual'' pulse train in the left ear is less modulated. This reduction in modulation depth will, in turn, cause the matched pitch of the ''higher-rate'' stimulus to increase (McKay and Carlyon, 1999). Note that this analysis, although accounting qualitatively for the pattern of results observed in experiment 1, fails to account for the ''duplex'' region observed in that experiment. For example, the ILD matches made to the lower-rate stimulus were often close to 0 dB, even when *Fr* was as high as 50 Hz—cf. for example the rightmost percepts experienced by listeners JD and TP [open triangles, Fig. 2(b)]. This would predict that the residual pulse train, heard on the left (Fig. 9, bottom row), would have every other pulse completely attenuated, and would have a perceived rate of *Fr*. Instead, the matches to the corresponding higher-rate pulse train [open squares, Fig. 3(a)] were closer to $2Fr$.

In order to reconcile the general scheme described above with the duplex region observed in experiment 1, one could assume that the links between sequential-same-ear and simultaneous-opposite-ear pulses do not have to be interpreted in a mutually consistent way by all mechanisms. In this sense, one can view the lateralization and temporal pitch mechanisms as ''grabbing what they can:'' at high pulse rates, the ''A'' pulses in the two ears are at least partially fused into an image that is not located at the ''ear of entry,'' but the pitch mechanism nevertheless completely incorporates the pulses in the higher-rate ear. This general idea, that different mechanisms interpret links between sounds in different ways, is not a new one, with various dissociations between pitch and location being present in the literature. For example, Deutsch (1974) presented subjects with an alternating pattern of 400- and 800-Hz tones in each ear, such that when one ear was receiving the 400-Hz tone the other

received the 800-Hz tone and vice versa. Subjects typically heard a sequence of high tones in one ear and a sequence of low tones in the other, with the ear in which the high tones were heard depending to some extent on the handedness of the listener. Deutsch (1976) suggested that, for right-handers, the perceived pitch at any one time was dominated by the sound presented to the right ear, whereas the perceived location was determined by the tone having the higher frequency. A consequence of this was that when 800 Hz was presented to the left ear and 400 Hz to the right, subjects heard the location corresponding to the left ear (as it received the higher tone) but the pitch corresponding to the right. A related phenomenon[7] has been reported for dichotic pairs of inharmonically related tones (Efron and Yund, 1974; Yund and Efron, 1975).

Overall, one can conclude that the concept of competition between binding processes, combined with the idea that the resulting links are assessed independently by different mechanisms, can account for the general pattern of results observed here. However, it must be conceded that such an account is purely qualitative, and that the way that it may be implemented by the auditory system is not specified in any detail. Below, we discuss a number of alternative explanations, which, while not necessarily inconsistent with the above account, have perhaps a more quantitative grounding in the psychoacoustic literature on binaural processing.

B. Interaural decorrelation

When two identical noises are presented to the two ears, listeners report a single, fused, centered percept. As the correlation between them is decreased (for example, by averaging one channel with an independent noise source), the percept becomes more diffuse, and listeners may eventually report hearing two separate sounds, one in each ear. In other words, below some interaural correlation, the binaural system ceases to form a fused image, and the inputs to the two ears may be treated as separate sources. This principle could account for the increasingly lateralized perception of the lower-rate pulse train in the synchronous condition of experiment 1 [Fig. 1(a)] as *Fr* is increased, if one assumes that the correlation is calculated over a finite duration. At low *Fr*, the interpulse duration of the higher-rate stimulus may be longer than the ''correlation window,'' so that, over the duration of a window centered on a pair of ''A'' pulses [Fig. 1(a)], the interaural correlation is 1.0. As *Fr* is increased beyond a certain value, adjacent ''B'' pulses start to fall into the window, and the resulting interaural correlation drops. The correlation should reach a minimum at a value of *Fr* where three pulses in one ear pass through the window unattenuated, and where the middle of these is accompanied by a simultaneous pulse in the opposite ear; this will occur when the duration of the window is equal to 1/*Fr*. The correlation should not drop systematically as *Fr* is increased further. This prediction can be tested against both the present data and the existing literature on binaural processing.

One way of measuring the effective duration of the binaural window was described by Summerfield and his colleagues (Akeroyd and Summerfield, 1998; Culling and Summerfield, 1998; Akeroyd and Summerfield, 1999). They measured the detection threshold for a brief, interaurally phase-shifted ($S\pi$) signal in a burst of diotic noise (N0), which was preceded and followed by two 200-ms bursts of uncorrelated noise (Nu). By measuring the increase in threshold as the duration of the N0 burst was reduced, Akeroyd and Summerfield derived a binaural window having an equivalent rectangular duration (ERD) of about 120 ms. A roughly similar value of 170 ms has been obtained in experiments where the listener is required to discriminate between a single burst of N0 noise and one in which a short portion has been replaced by Nu noise (Akeroyd and Summerfield, 1999). The correlation at the output of a binaural window, produced by the stimuli of experiment 1, should therefore reach a minimum at $Fr = 5.9 - 8.3$ Hz (1/0.170 to 1/0.120 s), and not drop further as *Fr* is increased above this value. This prediction is not borne out by the lateralization data [Figs. 2(c) and (d)], where the matches to the lower-rate stimulus change most over the range of *Fr* from 50–100 Hz. It is, however, more consistent with the effects on perceived rate/pitch, where the effect of synchrony between the pulses in the two ears does not vary markedly with *Fr* above about 12.5 Hz [Figs. 3(c), (d)].

The above observation, that a given window duration can account for the pitch but not the location data, is a consequence of the duplex region observed in experiment 1. It provides a further argument against decorrelation as a complete explanation for our results. Indeed, it leads to the slightly surprising observation that a binaural model is more effective at modeling the effects of interaural decorrelation on pitch than on lateralization. Furthermore, even if a modified version of the decorrelation hypothesis could account for our lateralization results, perhaps by assuming a different window duration, it is unlikely that it could then account for our pitch data.

C. Binaural sluggishness

Another potential explanation for the effects of *Fr* on lateralization comes from the concept of ''binaural sluggishness.'' For example, it is known that slow sinusoidal interaural phase modulation produces an image that oscillates between the ears, but that this percept of sound movement disappears once the modulation rate exceeds 10–20 Hz (Grantham and Wightman, 1978). The binaural processing of modulations in ILD is also sluggish, albeit to a lesser degree: as modulation rate exceeds 2–5 Hz, the perception of movement deteriorates (Blauert, 1972) and thresholds of delectability (*re* an unmodulated stimulus) increase (Grantham, 1984), but large time-varying ILDs can still be detected at modulation rates as high as 50–100 Hz (Grantham, 1984). As shown in Fig. 1(a), the ILD *re* each pulse in the higher-rate stimulus alternates between zero (for the ''A'' pulses) and infinity (''B'' pulses), and at high *Fr* it is possible that this rate is too fast for the binaural system to follow. It might then ''default'' to perceiving one sound on each side of the head. However, as shown in Fig. 2, what the listener actually hears is one sound on one side of the head and the other partway between the center of the head and the other ear. It seems hard to understand why the system would default to such a mode.

Some aspects of our results shed light on the nature of binaural sluggishness itself. At the highest *Fr* studied (100 Hz), subjects were still able to make reasonably reliable ILD matches to individual binaural images within each mixture, even though the ILD (experiment 1) or ITD (experiment 2) was alternating at a rate of 100 Hz. This suggests that although binaural sluggishness may apply to the perception of movement of continuous stimuli, it does not have a devastating effect on the lateralization of static images resulting from the segregation of subsets of elements within a pulse train. However, the need for listeners to perceive movement is clearly not essential for some temporal limitations on binaural processing to be observed, as evidenced by the long binaural windows reported in the works by Akeroyd, Culling, and Summerfield and described in Sec. IV A 1. Hence, it appears that the conditions under which binaural sluggishness does and does not occur still remain to be fully determined. Our results serve to further define the conditions under which this aspect of auditory processing affects performance.

D. Binaural adaptation

The experiments described in this article have led us to the conclusion that the temporal pitch mechanism is not driven by the output of binaural lateralization mechanisms. Another insight into the relationship between spatial and temporal processing comes from the phenomenon of binaural adaptation, first described by Hafter and Dye (1983). They required subjects to detect an ITD imposed on a click train, and found that the improvement in sensitivity that occurred as more clicks were added to the sequence was smaller than the improvement predicted from an optimal combination of independent observations. They concluded that the later clicks in a pulse train make a smaller contribution to sensitivity than do the earlier ones, provided the pulse rate exceeds a certain value. The minimum pulse rate at which this occurred varied somewhat across listeners, but, for most listeners, little or no binaural adaptation occurred at pulse rates of 200 Hz (the highest used here) or below. More generally, there is ample evidence that onsets can dominate suprathreshold lateralization judgments (Franssen, 1960; Saberi, 1996; Freyman *et al.*, 1997). This is not the case for pitch judgments; at least for discrimination measures, $F0$ discrimination of unresolved harmonics deteriorates markedly as signal duration drops below about 100 ms (Plack and Carlyon, 1995), and later pulses in a pulse train contribute to rate discrimination thresholds as much as the earlier pulses do (Hafter and Richards, 1988). Hence, it is possible that, in our experiments, the location of each pulse train was determined largely by the beginning of the sequence, whereas the pitch was derived from an analysis that was distributed more evenly in time.

V. SUMMARY AND CONCLUSIONS

In the Introduction we described preliminary observations showing that, as *Fr* is increased, two changes occur in the perception of the dichotic pulse train described in Fig. 1(a). The lower-rate train is perceived in a progressively more lateralized position, and the pitch of the higher-rate train increases relative to *Fr*. These findings were discussed in terms of competition between temporal and spatial binding in audition, and two specific hypotheses were described. The results of the more formal experiments described here allow a test of these two accounts.

First, it was suggested that increasing *Fr* could decrease the fusion between simultaneous pulses in opposite ears, but that the perceived rate/pitch might nevertheless be driven by the output of this binaural analysis. Two findings argue strongly against this scheme. Experiment 1 revealed that the perceived pitches and locations of the two percepts produced by the dichotic trains were not always consistent. Specifically, at $Fr > 12.5–25$ Hz, the binaural fusion between the two pulse trains did not affect the perceived pitch of the higher-rate train [Figs. 3(c), (d)], but had a strong effect on the lateral position of the lower-rate train [Figs. 2(c), (d), (e), (f)]. Experiment 2 showed that when the odd-numbered pulses in a dichotic stimulus (Fig. 4) had an opposite ITD to the even-numbered pulses, the lateral position of the odd pulses was dominated by their ITD, but that their pitch was roughly equal to the total pulse rate presented to each ear.

The results of both experiments are more consistent with a scheme whereby temporal pitch mechanisms operate on the input to each ear, rather than being driven by the output of binaural mechanisms This idea formed part of a hypothesis proposed in Sec. II A, which also stated that, when the value of *Fr* is too low to elicit a sense of pitch, the resulting perception of rate *is* driven by the output of binaural mechanisms. The highest value of *Fr* at which binaural fusion affected the pitch of the higher-rate pulse train was about 25 Hz, and given that the higher-rate train was presented at $2Fr$, this corresponds reasonably well to the 19 Hz at which listeners start to report hearing a pitch (Guttman and Pruzansky, 1962). However, this correspondence is not particularly strong evidence for a strict dissociation between the perception of roughness and of pitch, given evidence that the exact lower limit of pitch will depend on factors such as the task requirements and the spectral content of the stimuli (Guttman and Pruzansky, 1962; Pressnitzer *et al.*, 1999; Krumbholz *et al.*, 2000). The strongest distinction we wish to highlight is that, for $Fr > 25$ Hz, binaural fusion affects perceived location but not pitch.

Finally, it is worth pointing out that, although temporal pitch perception does not receive its input from lateralization mechanisms, this does not mean that it is a purely peripheral process. Indeed, the fact that the stimulus in one ear can affect the pitch matches in the other ear, even when there is no synchrony between the two ears [Fig. 1(b)], suggests that representations of temporal pitch interact at some central stage. In this regard, temporal pitch perception may be like other phenomena studied in the first author's laboratory, which appear to be dominated by ''ear of entry'' rather than perceived location, but which are also affected by stimuli presented to both ears; these phenomena include the detection of mistuning (Gockel and Carlyon, 1998) and modulation detection interference (Lyzenga and Carlyon, 2000). It differs from another group of phenomena, which is driven by the output of binaural processes; these include the build-up of auditory streaming (Rogers and Bregman, 1993), the

''overintegration'' of pitch (Gockel *et al.*, 1999), picking out one tune from a pair of interleaved melodies (Hartmann and Johnson, 1991), and tracking one spoken message in the presence of another (Darwin and Hukin, 1999). This categorization of phenomena in terms of whether they are or are not driven by binaural processes may well prove useful in uncovering the relationship between the neural processes responsible for different aspects of auditory perception.

ACKNOWLEDGMENT

We thank Michael Akeroyd for useful comments on a previous version of this manuscript.

[1]We use the term ''binding'' to refer to the formation of links between different sound elements (in our case, clicks). This usage is slightly different from that common in the visual literature, where it refers to the conjunction of different features (position, color, orientation) of the same object (e.g., Treisman, 1998).

[2]This general scheme is also consistent with Kubovy's (1988) ''thought experiment,'' in which each instrument in a sextet is played over a separate, hidden, loudspeaker. He argues that the listener would not spontaneously form a spatial image of the locations of the loudspeakers, but, rather, would use other cues to segregate the six sources and then (serially) attempt to assign a location to each source.

[3]Recently, Pressnitzer *et al.* (1999) and Krumbholz *et al.* (2000) have argued that the lowest rate that will support a sensation of melodic pitch varies with the frequency region into which the stimuli are filtered, and can be as high as 250 Hz.

[4]Another way of viewing the dissociation between the effects of binaural fusion on rate/pitch and on lateralization is to observe the range of *Fr* over which these two percepts change most: rate/pitch changes markedly with increasing *Fr* only up to about 25 Hz [squares, Figs. 3(c), (d)], whereas lateralization varies over the entire range, with some of the largest changes occurring at high *Fr* [triangles, Figs. 2(c) and (d)].

[5]Dye *et al.* (1996) measured the ''weights'' that subjects applied to the target and distractor when judging the laterality of the former. They did this for various combinations of target and distractor frequency. Interestingly, more than half of their subjects applied a greater weight to the ITD of the tone with the higher-frequency, even when that tone was the distractor and they were told to ignore it. This may be related to the finding that, in Deutsch's octave illusion, the perceived location is determined by that of the higher-frequency tone.

[6]Conversely, Wightman and Kistler also found that, in the low-frequency part of the spectrum, ITDs had a much larger effect on localization than did ILDs.

[7]Deutsch and Roll argued that different mechanisms were responsible for their phenomenon and that reported by Efron and Yund. One piece of evidence that they cited in support of this assertion was that handedness did not affect the pattern of results reported by Efron and Yund's subjects. However, it is worth noting that a study by Zwicker (1984), which successfully replicated Deutsch's main findings, failed to find an effect of handedness on her phenomenon either.

Akerboom, S., ten Hoopen, G., Olierook, P., and van der Schaaf, T. (**1983**). ''Auditory spatial alternation transforms auditory time,'' J. Exp. Psychol. Hum. Percept. Perform. **9**, 882–897.

Akeroyd, M. A., and Summerfield, A. Q. (**1998**). ''Predictions of signal thresholds in a frozen-noise masker using monaural and binaural temporal windows,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 433–439.

Akeroyd, M. A., and Summerfield, A. Q. (**1999**). ''A binaural analog of gap detection,'' J. Acoust. Soc. Am. **105**, 2807–2820.

Axelrod, S., Guzy, L. T., and Diamond, I. T. (**1967**). ''Perceived rate of monotic and dichotically alternating clicks,'' J. Acoust. Soc. Am. **43**, 51–55.

Beerends, J. G., and Houtsma, A. J. M. (**1989**). ''Pitch identification of simultaneous diotic and dichotic two-tone complexes,'' J. Acoust. Soc. Am. **85**, 813–819.

Blauert, J. (**1972**). ''On the lag of lateralization caused by interaural time and intensity differences,'' Audiology **11**, 265–270.

Buell, T. N., Trahiotis, C., and Bernstein, L. R. (**1991**). ''Lateralization of low-frequency tones: Relative potency of gating and ongoing interaural delays,'' J. Acoust. Soc. Am. **90**, 3077–3085.

Culling, J. F., Summerfield, A. Q., and Marshall, D. H. (**1998a**). ''Dichotic pitches as illusions of binaural unmasking. Huggins' pitch and the binaural edge pitch,'' J. Acoust. Soc. Am. **103**, 3509–3526.

Culling, J. F., Summerfield, A. Q., and Marshall, D. H. (**1998b**). ''Dichotic pitches as illusions of binaural unmasking. The Fourcin pitch and the dichotic repetition pitch,'' J. Acoust. Soc. Am. **103**, 3527–3539.

Culling, J. F., and Summerfield, Q. (**1998**). ''Measurements of the binaural temporal window using a detection task,'' J. Acoust. Soc. Am. **103**, 3540–3553.

Darwin, C. J., and Carlyon, R. P. (**1995**). ''Auditory Grouping,'' in *Hearing*, edited by B. C. J. Moore (Academic, Orlando), pp. 387–424.

Darwin, C. J., and Ciocca, V. (**1992**). ''Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component,'' J. Acoust. Soc. Am. **91**, 3381–3390.

Darwin, C. J., and Hukin, R. W. (**1999**). ''Auditory objects of attention: The role of interaural time differences,'' J. Exp. Psychol.: Hum. Percept. Perform. **25**, 617–629.

de Boer, E. (**1956**). ''On the 'Residue' in Hearing,'' Doctoral dissertation, Amsterdam.

de Boer, E. (**1976**). ''On the 'Residue' and Auditory Pitch Perception,'' in *Handbook of Sensory Physiology*, edited by W. D. Keindel and W. D. Neff (Springer, Berlin), Vol. 3, pp. 479–584.

Deutsch, D. (**1974**). ''An auditory illusion,'' Nature (London) **251**, 307–309.

Deutsch, D. (**1976**). ''Separate 'what' and 'where' decision mechanisms in processing a dichotic tonal sequence,'' J. Exp. Psychol.: Hum. Percept. Perform. **2**, 23–29.

Durlach, N. I. (**1972**). ''Binaural signal detection: Equalization and cancellation theory,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. II.

Dye, R. H., Stellmack, M. A., Grange, A. N., and Yost, W. A. (**1996**). ''The effect of distractor frequency on judgments of target laterality based on interaural delays,'' J. Acoust. Soc. Am. **99**, 1096–1107.

Efron, R., and Yund, E. W. (**1974**). ''Dichotic competition of simultaneous tones bursts of different frequency I. Dissociation of pitch from lateralization and loudness,'' Neuropsychologia **12**, 249–256.

Franssen, N. V. (**1960**). ''Some considerations of the mechanisms of directional hearing,'' Ph.D. thesis, Delft.

Freyman, R. L., Zurek, P. M., Balakrishnan, U., and Chiang, Y.-C. (**1997**). ''Onset dominance in lateralization,'' J. Acoust. Soc. Am. **101**, 1649–1659.

Gockel, H., and Carlyon, R. P. (**1998**). ''Effects of ear of entry and perceived location of synchronous and asynchronous components on mistuning detection,'' J. Acoust. Soc. Am. **104**, 3534–3545.

Gockel, H., Carlyon, R. P., and Micheyl, C. (**1999**). ''Context dependence of fundamental frequency discrimination: Lateralized temporal fringes,'' J. Acoust. Soc. Am. **106**, 3553–3563.

Godsmark, D., and Brown, G. J. (**1999**). ''A blackboard architecture for computational auditory scene analysis,'' Speech Commun. **27**, 351–366.

Grantham, D. W. (**1984**). ''Detectability of varying interaural intensity differences,'' J. Acoust. Soc. Am. **76**, 71–76.

Grantham, D. W. (**1995**). ''Spatial Hearing and Related Phenomena,'' in *Hearing*, edited by B. C. J. Moore (Academic, Orlando), pp. 297–346.

Grantham, D. W., and Wightman, F. L. (**1978**). ''Detectability of varying interaural temporal differences,'' J. Acoust. Soc. Am. **63**, 511–523.

Guttman, N., and Pruzansky, S. (**1962**). ''Lower limits of pitch and musical pitch,'' J. Speech Hear. Res. **5**, 207–214.

Hafter, E. F., and Dye, R. H. (**1983**). ''Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number,'' J. Acoust. Soc. Am. **73**, 644–651.

Hafter, E. F., and Richards, V. M. (**1988**). ''Discrimination of the rate of filtered impulses,'' Percept. Psychophys. **43**, 405–414.

Hartmann, W. M., and Johnson, D. (**1991**). ''Stream segregation and peripheral channeling,'' Music Percept. **9**, 155–184.

Heller, L. M., and Trahiotis, C. (**1996**). ''Extents of laterality and binaural interference effects,'' J. Acoust. Soc. Am. **99**, 3632–3637.

Henning, G. B. (**1974**). ''Detectability of interaural delay with high-frequency complex waveforms,'' J. Acoust. Soc. Am. **55**, 84–90.

Houtsma, A. J. (**1995**). ''Pitch Perception,'' in *Hearing*, edited by B. C. J. Moore (Academic, Orlando), pp. 267–298.
Huggins, A. W. F. (**1974**). ''On perceptual integration of dichotically alternated pulse trains,'' J. Acoust. Soc. Am. **56**, 939–943.
Kaernbach, C., and Demany, L. (**1998**). ''Psychophysical evidence against the autocorrelation theory of auditory temporal processing,'' J. Acoust. Soc. Am. **104**, 2298–2306.
Koehnke, J., Culotta, C. P., Hawley, M. L., and Colburn, H. S. (**1995**). ''Effects of reference interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing,'' Ear Hear. **16**, 331–353.
Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). ''The lower limit of pitch as determined by rate discrimination,'' J. Acoust. Soc. Am. **108**, 1170–1180.
Kubovy, M. (**1988**). ''Should we resist the seductiveness of the space:time::vision:audition analogy?'' J. Exp. Psychol.: Hum. Percept. Perform. **14**, 318–320.
Kuhn, G. F. (**1987**). ''Physical acoustics and measurements pertaining to directional hearing,'' in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer, New York), pp. 3–25.
Lyzenga, J., and Carlyon, R. P. (**2000**). ''Binaural effects in frequency modulation detection interference for vowel formants,'' J. Acoust. Soc. Am. **108**, 753–759.
McFadden, D., and Pasanen, E. G. (**1976**). ''Lateralization at high frequencies based on interaural time differences,'' J. Acoust. Soc. Am. **59**, 634–639.
McKay, C. M., and Carlyon, R. P. (**1999**). ''Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains,'' J. Acoust. Soc. Am. **105**, 347–357.
Mossop, J. E., and Culling, J. F. (**1998**). ''Lateralization of large interaural delays,'' J. Acoust. Soc. Am. **104**, 1574–1579.
Nakao, M. A., and Axelrod, S. (**1976**). ''Effects of bilateral alternation on perceived temporal uniformity of auditory and somasthetic pulse trains,'' Percept. Psychophys. **20**, 274–280.
Plack, C. J., and Carlyon, R. P. (**1995**). ''Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics,'' J. Acoust. Soc. Am. **98**, 1355–1364.
Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**1999**). ''The lower limit of melodic pitch with filtered harmonic complexes,'' J. Acoust. Soc. Am. **105**, 1152.
Ritsma, R. J. (**1963**). ''On pitch discrimination of residue tones,'' Int. Audiol. **2**, 34–37.
Rogers, W. L., and Bregman, A. S. (**1993**). ''An experimental evaluation of three theories of stream segregation,'' Percept. Psychophys. **53**, 179–189.
Saberi, K. (**1996**). ''Observer weighting of interaural delay in filtered impulses,'' Percept. Psychophys. **58**, 1037–1046.
Shackleton, T. M., and Carlyon, R. P. (**1994**). ''The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,'' J. Acoust. Soc. Am. **95**, 3529–3540.
Terhardt, E. (**1977**). ''Pitch shift of monaural pure tones caused by contralateral sounds,'' Acustica **37**, 56–57.
Thurlow, W. R. (**1943**). ''Studies in auditory theory. I. Binaural interaction and the perception of pitch,'' J. Exp. Psychol. **32**, 17–36.
Treisman, A. (**1998**). ''Feature binding, attention, and object perception,'' Philos. Trans. R. Soc. London, Ser. B **353**, 1265–1306.
Warren, R. M., and Bashford, J. A. (**1976**). ''Auditory contralateral induction: An early stage in binaural processing,'' Percept. Psychophys. **20**, 380–386.
Wiegrebe, L., and Patterson, R. D. (**1999**). ''Quantifying the distortion products generated by SAM noise,'' J. Acoust. Soc. Am. **106**, 2709–2718.
Wightman, F. L., and Kistler, D. J. (**1989**). ''The dominant role of low-frequency interaural time differences in sound localization,'' J. Acoust. Soc. Am. **91**, 1648–1661.
Yost, W. A. (**1981**). ''Lateralization position of sinusoids presented with interaural intensive and temporal differences,'' J. Acoust. Soc. Am. **70**, 397–409.
Yost, W. A., Wightman, F. L., and Green, D. M. (**1971**). ''Lateralization of filtered clicks,'' J. Acoust. Soc. Am. **50**, 1526–1531.
Yund, E. W., and Efron, R. (**1975**). ''Dichotic competition of simultaneous tones bursts of different frequency—II. Suppression and ear dominance functions,'' Neuropsychologia **13**, 137–150.
Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). ''Critical bandwidths in loudness summation,'' J. Acoust. Soc. Am. **29**, 548–557.
Zwicker, T. (**1984**). ''Experimente zur dichotischen Oktav-Tauschung,'' Acustica **55**, 128–136.

Influence of rate of change of frequency on the overall pitch of frequency-modulated tones[a)]

Hedwig Gockel[b)]
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, England

Brian C. J. Moore[c)]
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

Robert P. Carlyon[d)]
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, England

(Received 26 July 2000; revised 24 October 2000; accepted 22 November 2000)

The mechanism(s) determining pitch may assign less weight to portions of a sound where the frequency is changing rapidly. The present experiments explored the possible effect of this on the overall pitch of frequency-modulated sounds. Pitch matches were obtained between an adjustable unmodulated sinusoid and a sinusoidal carrier that was frequency modulated using a highly asymmetric function with the form of a repeating U ($\cup\cup$) or inverted U ($\cap\cap$). The amplitude was constant during the 400-ms presentation time of each stimulus, except for 10-ms raised-cosine onset and offset ramps. In experiment 1, the carrier level was 50 dB SPL and the geometric mean of the instantaneous frequency of the modulated carrier, f_c, was either 0.5, 1, 2, or 8 kHz. The modulation rate (f_m) was 5, 10, or 20 Hz. The overall depth (maximum to minimum) of the FM was 8% of f_c. For all carrier frequencies, the matched frequency was shifted away from the mean carrier frequency, downwards for the $\cup\cup$ stimuli and upwards for the $\cap\cap$ stimuli. The shift was typically slightly greater than 1% of f_c, and did not vary markedly with f_c. The effect of f_m was small, but there was a trend for the shifts to decrease with increasing f_m for $f_c = 0.5$ kHz and to increase with increasing f_m for $f_c = 2$ kHz. In experiment 2, the carrier level was reduced to 20 dB SL and matches were obtained only for $f_c = 2$ kHz. Shifts in matched frequency of about 1% were still observed, but the trend for the shifts to increase with increasing f_m no longer occurred. In experiment 3, matches were obtained for a 4-kHz carrier at 50 dB SPL. Shifts of about 1% again occurred, which did not vary markedly with f_m. The shifts in matched frequency observed in all three experiments are not predicted by models based on the amplitude- or intensity-weighted average of instantaneous frequency (EWAIF or IWAIF). The shifts (and the pitch shifts observed earlier for two-tone complexes and for stimuli with simultaneous AM and FM) are consistent with a model based on the assumption that the overall pitch of a frequency-modulated sound is determined from a weighted average of period estimates, with the weight attached to a given estimate being inversely related to the short-term rate of change of period and directly related to a compressive function of the amplitude.
[DOI: 10.1121/1.1342073]

PACS numbers: 43.66.Hg, 43.66.Ba, 43.66.Mk [LRB]

I. INTRODUCTION

Many sounds produced by musical instruments or the human voice can be characterized as complex tones in which the fundamental frequency ($f0$) undergoes quasiperiodic fluctuations (vibrato). If the fluctuations are of moderate depth, the fluctuation rate is not too high, and the tones are reasonably long, then the tones are perceived as having a single overall pitch (Miller and Heise, 1950; Shonle and Horan, 1976; d'Alessandro and Castellengo, 1994), which is sometimes referred to as the ''principal pitch'' (Iwamiya *et al.*, 1983, 1984); otherwise, multiple pitches may be heard (McClelland and Brandt, 1969; d'Alessandro and Castellengo, 1994). When a single overall pitch is perceived, the pitch value has usually been reported to correspond to the mean $f0$ (Tiffin, 1931; Seashore, 1938; Sundberg, 1978a, 1978b; Shonle and Horan, 1980; Iwamiya *et al.*, 1983; d'Alessandro and Castellengo, 1994), although Shonle and Horan (1980) reported that the geometric mean gave a slightly better fit to their results than the arithmetic mean. These findings have led to models of the pitch of vibrato tones in which the overall pitch is computed as a simple average of the pitches derived from brief samples of the sound (Iwamiya *et al.*, 1983), although d'Alessandro and Castellengo (1994) proposed that, for brief sounds, later parts of the sound are weighted more highly than earlier parts of the sound.

[a)]Some of the data for experiment 1 were presented in Gockel *et al.* [Br. J. Audiol. **34**, 99 (2000)].
[b)]Present address: CNBH, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK. Electronic mail: hedwig.gockel@mrc-cbu.cam.ac.uk
[c)]Electronic mail: bcjm@cus.cam.ac.uk
[d)]Electronic mail: bob.carlyon@mrc-cbu.cam.ac.uk

Much of the research on the pitch of vibrato tones made use of tones which were frequency modulated symmetrically on a linear frequency scale. However, Shonle and Horan (1980) included an experiment using asymmetrical modulation waveforms (described as ''flat top'' and ''flat bottom''), and concluded that overall pitch depended on ''an averaging of all frequencies present and not just the extreme frequencies'' (page 246). Iwamiya *et al.* (1983) came to a similar conclusion. Feth *et al.* (1982) have also studied the pitch of sounds in which the frequency modulation was highly asymmetric. However, their stimuli contained both amplitude modulation (AM) and frequency modulation (FM). The stimuli were created by adding two sinusoids with closely spaced frequencies (f_1 and f_2) and slightly different amplitudes (in what follows, the term ''amplitude'' will be used to denote the magnitude of the envelope). These stimuli can be described as a single sinusoid whose instantaneous frequency and amplitude fluctuate periodically at a rate equal to f_2-f_1. Rapid changes in instantaneous frequency occur close to envelope minima. As described earlier by Helmholtz (1863), they found that the pitch did not correspond to the mean of f_1 and f_2 but was shifted towards the frequency of the component with higher amplitude. They accounted for their results using a model based on calculation of the envelope-weighted (arithmetic) average of instantaneous frequency (EWAIF), described earlier by Feth (1974). In this model, portions of the sound with low amplitude receive less weight in the calculation of pitch than portions with high amplitude. In later work, models based on the intensity-weighted average of instantaneous frequency (IWAIF) have been proposed (Anantharaman *et al.*, 1993; Dai, 1993; see also Iwamiya *et al.*, 1984), but the general principle of the models remains the same.

Recent work on the perception of frequency modulated sounds suggests that the computation of overall pitch may not be so simple. It is often assumed that the detection of FM of sinusoidal carriers can be mediated by two mechanisms: a place mechanism based on FM-induced AM in the excitation pattern, and a temporal mechanism based on phase locking in the auditory nerve. It has been proposed that the mechanism for extraction of pitch is ''sluggish'' and has difficulty tracking rapid changes in frequency, especially when the pitch is extracted using temporal information (phase locking in the auditory nerve) (Moore and Sek, 1995, 1996; Sek and Moore, 1999, 2000). For example, Moore and Sek (1995) measured psychometric functions for the detection of FM and AM using quasitrapezoidal modulation with a rate of 5 periods per s and carriers of 250, 1000, and 6000 Hz. With quasitrapezoidal modulation, the stimuli spend more time at extremes of frequency or amplitude than is the case with sinusoidal modulation. They found that performance was better for 5-Hz trapezoidal modulation than for 5-Hz sinusoidal modulation. More importantly, for the two lower carrier frequencies only, the improvements were markedly greater for FM than for AM detection. This is consistent with the idea that the use of phase-locking information depends on the time that the stimuli spend at frequency extremes.

Evidence for sluggishness of the temporal pitch mechanism has also been obtained using carriers consisting of groups of high, unresolved harmonics (Carlyon and Shackleton, 1994; Shackleton and Carlyon, 1994; Plack and Carlyon, 1995). It is usually assumed that the pitch of such sounds is extracted by a purely temporal mechanism. Plack and Carlyon (1995) showed that detection of FM of $f0$ was much poorer for a complex tone consisting of unresolved harmonics than for one with resolved harmonics (5-Hz FM rate); for the latter, place information as well as temporal information would be available. It is noteworthy that FM detection thresholds for the complex tone with unresolved harmonics were similar to those found in an $f0$ discrimination task (between two steady tones) when the tone duration was short. It appears that the short time spent at frequency extremes is the critical factor (cf. Carlyon *et al.*, 2000).

The term sluggishness has often been used to characterize the poor ability of the pitch mechanism to track rapid changes in frequency. This terminology would be consistent with the idea that the poor ability is produced by an integration mechanism with a long integration time. If this were the case, the overall pitch of a frequency-modulated sound should correspond to the mean frequency of that sound. However, it is also possible that the initial analysis of pitch information occurs over relatively short durations. The pitch estimate for a given segment of the stimulus may be reduced in accuracy when the frequency changes rapidly during that segment. Subsequently, there may be a longer-term integration or averaging process, in which pitch estimates from segments of the sound where the frequency (or $f0$) is changing rapidly receive less weight than estimates from portions where the frequency is changing more slowly. For brevity, we will refer to this as ''stability-sensitive weighting.'' If such a thing does occur, results on the pitch of two-tone complexes (Feth, 1974; Feth *et al.*, 1982; Anantharaman *et al.*, 1993; Dai, 1993) may need reinterpretation. The shift in pitch found for these stimuli might have occurred partly because of stability-sensitive weighting. With these stimuli, it is not possible to tease apart the effects of stability-sensitive weighting and of fluctuations in amplitude on the overall pitch of a frequency-modulated sound; the stimuli are modulated in both frequency and amplitude, and amplitude covaries with the rate of change of instantaneous frequency.

In the present study, we used frequency-modulated tones with highly asymmetrical patterns of FM, comparable to the patterns of FM for the two-tone stimuli of Feth and coworkers. However, for our stimuli the amplitude was constant during the stimulus presentation. Our working hypothesis was that stability-sensitive weighting would result in the pitch mechanism placing less weight on portions of the waveform where the instantaneous frequency was changing rapidly. This would result in a shift in the perceived pitch away from the value corresponding to the mean frequency. The EWAIF and IWAIF models would predict no such shift.

II. EXPERIMENT 1

A. Stimuli

Pitch matches were obtained between an unmodulated sinusoid and a sinusoid that was frequency modulated with a repeated U-shaped function (∪∪) or a repeated inverted

U-shaped function ($\cap\cap$); see Fig. 1, top. The amplitude was constant during the 400-ms presentation time of each stimulus (see Fig. 1, bottom), except for 10-ms raised-cosine onset and offset ramps. The level was 50 dB SPL. The geometric mean of the instantaneous frequency of each FM tone was always equal to its carrier frequency (f_c) and was either 0.5, 1, 2, or 8 kHz. The modulation rate (f_m) was 5, 10, or 20 Hz. For f_c equal to 2 kHz, a rate of 50 Hz was also used, as preliminary data indicated that the overall pitch was affected by modulation rate, and we wished to determine whether this effect persisted at a higher rate. The overall depth (maximum to minimum) of the FM (f_{depth}) was 8% of f_c. For f_c equal to 0.5 kHz, initial measurements indicated that the overall pitch was less affected by the modulator waveform ($\cup\cup$ versus $\cap\cap$) than was the case for higher carrier frequencies; to obtain larger effects, we also used $f_{\text{depth}}=12\%$ for f_c equal to 0.5 kHz.

Equation (1) specifies the time waveform of the FM tones

$$x(t)=\sin 2\pi\left(f_c t+\int_0^t \Delta F(\lambda)\,d\lambda\right), \tag{1}$$

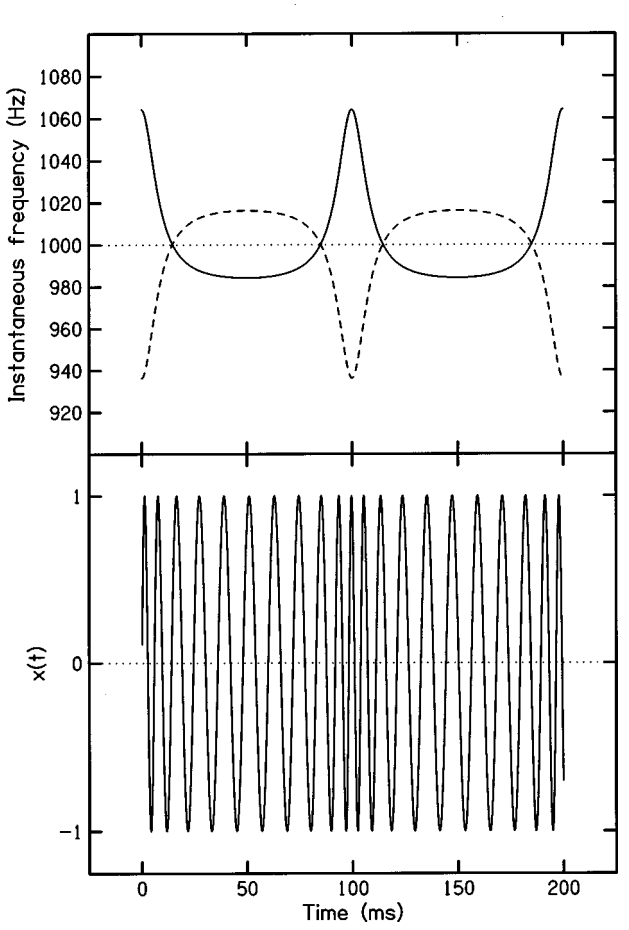


FIG. 1. The top panel shows the instantaneous frequency trajectories of the $\cup\cup$ (solid line) and the $\cap\cap$ (dashed line) stimuli. The two stimuli have the same geometric mean instantaneous frequency (1 kHz). In this example f_c equals 1 kHz, f_{depth} equals 8%, f_m is 10 Hz, and Φ equals zero. The bottom panel shows the waveform of a sinusoid modulated by two cycles of the $\cup\cup$ function. For clarity, a f_c of 100 Hz, and a greatly exaggerated f_{depth} of 50% are shown.

where λ is the integration variable, and $\Delta F(\lambda)$ is either the repeated U-shaped $[\Delta F_{\cup}(t)]$ or the repeated inverted U-shaped $[\Delta F_{\cap}(t)]$ function. The generic function used to define the instantaneous frequency for U-shaped FM with rate f_m is given by

$$f_{\cup}(t)=(1-a^2)/(1-2a\cos(2\pi f_m t+\Phi)+a^2), \tag{2}$$

where t is time in seconds and a is equal to 0.6. Time zero is defined as the start of the stimulus. The maximum of this function (Max) equals $(1+a)/(1-a)$. The minimum (Min) equals $(1-a)/(1+a)$. If the starting phase Φ equals zero, then the function starts at its maximum; if it equals π, then it starts at its minimum. The inverted U-shaped function is given by

$$f_{\cap}(t)=\text{Max}-f_{\cup}(t). \tag{3}$$

The U-shaped modulator is given by

$$\Delta F_{\cup}(t)=k(f_{\cup}(t)-Z_{\cup}), \tag{4}$$

and the inverted U-shaped modulator is given by

$$\Delta F_{\cap}(t)=k(f_{\cap}(t)-Z_{\cap}). \tag{5}$$

The parameter k determines the range over which the frequency varies, with

$$k=f_{\text{depth}}f_c/(100(\text{Max}-\text{Min})). \tag{6}$$

The parameters $Z_{\cup}$ and $Z_{\cap}$ determine the ''zero line'' of the modulator functions. They were chosen so that the geometric mean instantaneous frequency of each FM tone was equal to its carrier frequency. Note that, for the relatively small modulation depths used in these experiments, the geometric and arithmetic mean frequencies are essentially identical. For the 8% overall modulation depth, the arithmetic mean frequencies of the $\cup\cup$ and $\cap\cap$ stimuli are 0.0250% and 0.0262% above f_c, respectively. The difference between the arithmetic means of the $\cup\cup$ and $\cap\cap$ stimuli is 0.0012% of f_c.

The starting phase of the modulator function was either 0 or π. For Φ equal to zero, the modulation starts at the extreme of the fast frequency excursion (phase 1). For Φ equal to π, it starts half a cycle later, i.e., halfway through the plateau in instantaneous frequency (phase 2). These two phase conditions were included to check whether the end (or the start) of the stimulus contributes more to the overall pitch than the central portion.

All stimuli were generated digitally, played out by a 16-bit digital-to-analog converter (CED 1401 plus) at a sampling rate of 40 kHz, and passed through an antialiasing filter (Kemo 21C30) with a cutoff frequency of 17.2 kHz (slope of 100 dB/oct). Stimuli were presented monaurally (to the left ear of subjects), using Sennheiser HD250 headphones. Subjects were seated individually in an IAC double-walled sound-attenuating booth.

B. Procedure

A two-interval, two-alternative adaptive procedure was used to obtain pitch matches between the modulated tone and an unmodulated sinusoid; the frequency of the latter was adjusted to obtain the match. The FM tone was always pre-

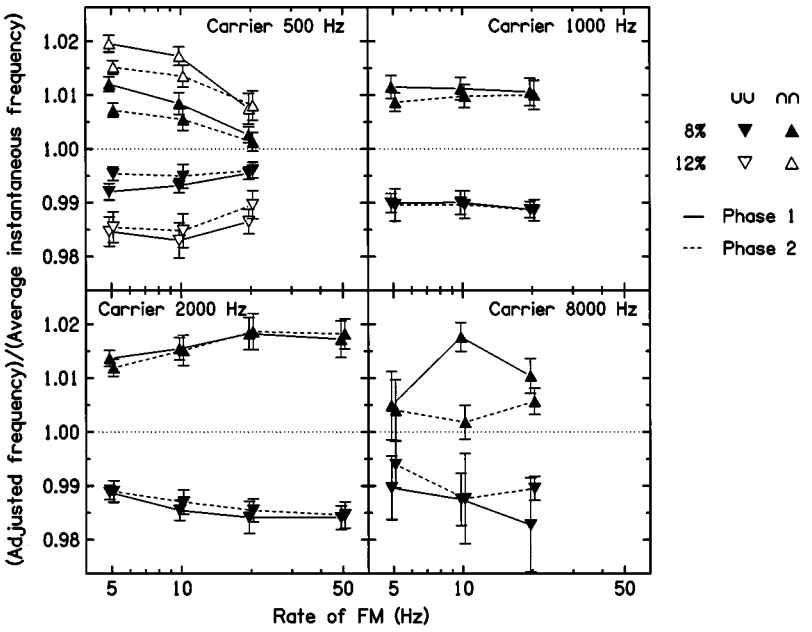


FIG. 2. Geometric mean data and standard errors across all four subjects (except for carrier of 8 kHz, where only three subjects are included). Each panel shows data for one carrier frequency. The value of f_{depth} was 8% (filled symbols) or 12% (open symbols; only measured for $f_c = 0.5$ kHz). Symbols for phase 1 are connected by a solid line. Symbols for phase 2 are connected by a dashed line. Downward-pointing triangles indicate the ∪∪ stimuli. Upward-pointing triangles indicate the ∩∩ stimuli.

sented in the first interval. After an interstimulus interval of 500 ms, the unmodulated sinusoid was presented. Each interval was marked by a light. Subjects were required to indicate the interval containing the tone with the higher pitch. A two-down, two-up rule was used to adjust the frequency of the unmodulated tone. After two consecutive identical answers, the frequency of the tone was increased if the first interval was judged higher and decreased if the second interval was judged higher. This procedure tracks the point corresponding to 50% ''higher'' (or ''lower'') judgments of the adjustable tone. At each turnpoint (a change in the direction of the frequency change), the step size was halved. The starting step size was 10% of f_c. The minimum step size was 0.2% of f_c at 0.5, 1, and 2 kHz, and 0.3% at 8 kHz.

The starting frequency of the unmodulated tone was varied quasirandomly in the range $f_c \pm 0.4$ octaves. For each FM tone there were always two runs in a sequence; in one, the unmodulated tone started at a frequency higher than f_c, and in the other it started at a frequency lower than f_c. The order of the two was balanced. Two buttons on a response box, marked 1 and 2, respectively, were used to indicate the interval with the higher pitch. By pressing a third button (underneath a red LED), subjects initiated the presentation of the same stimulus pair again. By pressing a fourth button (underneath a green LED) subjects indicated when they were satisfied with the pitch match. Subjects were encouraged to ''bracket'' the overall pitch of the FM tone with the pitch of the unmodulated tone, that is, to go from higher to lower to higher, etc. They were also encouraged (and did so) to listen a few times to the same stimulus pair before indicating the interval with the higher pitch. Only after they reached the minimum step size was their response ''Pitch of the two tones is the same'' accepted. If a subject pressed the fourth button before the minimum step size was reached, the response was ignored and the same stimulus pair was presented again. No feedback was provided. Subjects were informed about the procedure.

A pitch match was defined as the frequency of the unmodulated tone presented immediately before the subject indicated that the pitch in the two intervals was equal. At least ten pitch matches were obtained for each condition and subject. The data reported are the geometrical means of these ten (or more) pitch matches.

The total duration of a single session was about 2 h, including rest times. The different carrier frequencies were run in separate blocks. FM rate and starting phase of the modulator function were run in a quasirandomized order. The order of the conditions was counterbalanced over subjects. To familiarize subjects with the procedure and equipment, they were given three practice runs (i.e., made three matches) before data collection was started.

C. Subjects

Four subjects, ranging in age from 18 to 38 years, participated in all conditions. Their quiet thresholds at octave frequencies between 500 and 8000 Hz were within 15 dB of the ANSI (1969) standard. All subjects were musically trained. They differed in their degree of training. Subject JJ was an eighth-grade singer, and had the highest degree of training. The other ones, ranked from more to less training, were PK, CO, and US.

D. Results

Except for $f_c = 8$ kHz, the results were similar across subjects and mean results for the three lower carrier frequencies (with associated standard errors) are plotted in Fig. 2. The mean adjusted frequency of the sinusoidal matching tone is expressed relative to the average instantaneous frequency of the modulated tone (equal to f_c) and plotted as a function of modulation rate. For the 8-kHz carrier, one subject (CO) produced much more erratic results than the other subjects, and she also showed strong bias effects. The mean data for the 8-kHz carrier shown in the bottom right panel of Fig. 2 were calculated excluding her results. For the ∪∪ stimuli (downward-pointing triangles), the adjusted fre-

quency was below the mean frequency of the modulated stimuli, while for the ∩∩ stimuli (upward-pointing triangles), the adjusted frequency was above the mean frequency of the modulated stimuli. This was true for all modulation rates and all carrier frequencies. These pitch shifts are in the direction expected from the hypothesis described in the Introduction, that portions of the sound where the frequency was changing rapidly would receive less weight in the computation of pitch than portions where the frequency was changing more slowly. The pitch shifts are not predicted by the EWAIF or IWAIF models. The existence of pitch shifts for the 8-kHz carrier will be discussed later. Note that although the results for CO at 8 kHz showed a bias effect, the adjusted frequency always falling below the mean frequency of the modulated carrier, the ∪∪ stimuli were still generally matched with a lower frequency than the ∩∩ stimuli.

Within-subjects analyses of variance (ANOVAs) were conducted on the logarithms of the ratios (adjusted frequency/f_c) with factors shape of the modulator (∪∪ versus ∩∩), starting phase, and modulation rate. Logarithms were used, as the variability of the matches tended to increase with increasing pitch shift. The Huynh–Feldt (HF) correction was used when the condition of sphericity was not satisfied. Note that, since the shifts in pitch match for the ∪∪ stimuli were roughly a ''mirror image'' of those for the ∩∩ stimuli (i.e., the shifts were roughly symmetric about a value of 1), we would not generally expect to find a significant main effect of modulation rate or starting phase. Effects of modulation rate or starting phase would be revealed as an interaction of these factors with shape. Separate analyses were conducted for each carrier frequency, as more modulation rates were used for the 2-kHz carrier than for the other carriers, and as two modulation depths were used for the 0.5-kHz carrier only.

The main effect of shape was highly significant for all carrier frequencies: for $f_c = 0.5$ kHz, $F(1,3) = 92.4$, $p = 0.002$ for the 8% depth and $F(1,3) = 176.9$, $p < 0.001$ for the 12% depth; for $f_c = 1$ kHz, $F(1,3) = 980.7$, $p < 0.001$; for $f_c = 2$ kHz, $F(1,3) = 106.3$, $p = 0.002$; for $f_c = 8$ kHz (analysis based on the results for all four subjects), $F(1,3) = 35.6$, $p = 0.009$. There was no significant main effect of starting phase for any carrier frequency. However, the interaction of starting phase and shape was significant for the 0.5-kHz carrier [$F(1,3) = 77.7$, $p = 0.003$ for the 8% depth and $F(1,3) = 11.7$, $p = 0.042$ for the 12% depth] and the 8-kHz carrier [$F(1,3) = 22.4$, $p = 0.018$]. The interactions occurred because the difference in pitch match between the ∪∪ and ∩∩ stimuli was larger for phase 1 (solid lines) than for phase 2 (dashed lines). For phase 1, the stimuli started and finished with an instantaneous frequency close to a point where the frequency was changing rapidly (see Fig. 1). The overall pitch shifts measured (for both phases) indicate that these points in the waveform were weighted less than the portions of the stimuli where the frequency was changing more slowly. The larger shifts for phase 1 indicate that the rapidly changing portions received even less weight when they were at the start or end of the stimuli than when they were contained within the stimuli. Overall, though, the effects of starting phase were small, which supports the idea that the overall pitches of the stimuli were not dominated by any specific segment of the stimuli relative to the onset, for example, the start or the end.

The interaction of shape and modulation rate was significant for $f_c = 0.5$ kHz [$F(2,6) = 21.0$, HF = 0.58, $p = 0.014$ for the 8% depth and $F(2,6) = 42.0$, HF = 0.63, $p = 0.003$ for the 12% depth] and for $f_c = 2$ kHz [$F(3,9) = 40.5$, $p < 0.001$], but not for the other carrier frequencies. However, these interactions reflect opposite effects: for $f_c = 0.5$ kHz, the difference in pitch match between the ∪∪ and ∩∩ stimuli decreased with increasing modulation rate, whereas for $f_c = 2$ kHz the difference increased with increasing modulation rate.

For the 0.5-kHz carrier, the main effect of modulation rate was just significant; $F(2,6) = 11.1$, HF = 0.55, $p = 0.038$ for the 8% depth and $F(2,6) = 8.07$, HF = 0.65, $p = 0.045$ for the 12% depth, perhaps reflecting a bias effect. The main effect of modulation rate was not significant for any other carrier frequency.

An additional ANOVA was conducted on the combined data for the 0.5-, 1-, and 2-kHz carriers, for modulation rates up to 20 Hz and for the single FM depth of 8%. The data for the 8-kHz carrier were excluded because of the sizable individual differences at that frequency. The analysis showed a significant main effect of shape; $F(1,3) = 618.7$, $p < 0.001$. There was also a significant interaction of carrier frequency and shape; $F(2,6) = 16.6$, HF = 0.55, $p = 0.021$, confirming that the difference in pitch between the ∪∪ and ∩∩ stimuli increased with increasing carrier frequency up to 2 kHz. Finally, there was a significant three-way interaction of carrier frequency, shape, and modulation rate; $F(4,12) = 37.4$, $p < 0.001$, confirming that the variation of the pitch shifts with modulation rate differed across carrier frequency.

In summary, the ANOVAs confirm the clear effect of modulator shape on the pitch matches, for all four carrier frequencies. The interaction of modulation rate with shape was significant for the 0.5- and 2-kHz carriers, but not for the other carriers. The overall magnitude of the pitch shifts increased somewhat with increasing carrier frequency up to 2 kHz.

III. EXPERIMENT 2: EFFECTS OF REDUCED LEVEL

The spectra of frequency-modulated sounds generally contain many frequency components (Hartmann, 1997). The levels of these components generally decrease with increasing frequency separation from the carrier frequency, although the decrease is not always monotonic. When the modulation rate and depth are low (typically below about 20 Hz and 10%, respectively, but depending somewhat on carrier frequency), these components lie close to the carrier frequency and are not resolved in the auditory system. However, for larger rates or depths, one or more components may be resolved and heard as separate tone(s) (McClelland and Brandt, 1969). Our subjects reported hearing only a single overall pitch in each condition (although the modulation was heard as a fluctuation in pitch). However, it is possible that performance was affected to some extent by the presence of resolvable components. Specifically, the effects of modula-

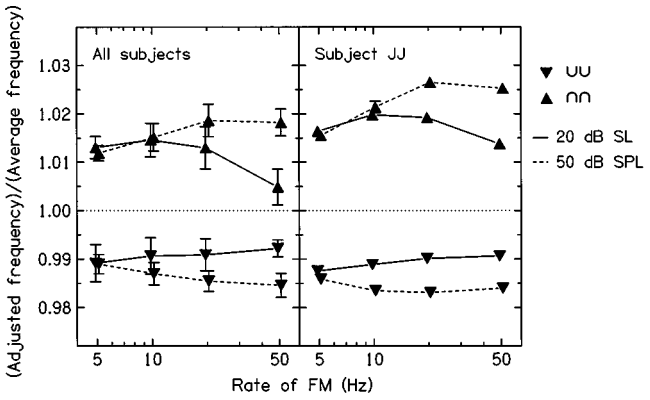


FIG. 3. Geometric mean data and standard errors for $f_c = 2$ kHz, and $f_{\text{depth}} = 8\%$, using a level of 20 dB SL (solid lines) or 50 dB SPL (dashed lines). In the left panel data for each level are averaged across four subjects, only one of whom (JJ) was tested at both levels. The right panel shows results for JJ. All stimuli had starting phase 2.

tion rate found for the 2-kHz carrier might reflect an increasing influence of resolved components with increasing modulation rate. To test this idea, we reran some of the conditions of experiment 1, but using a much lower sensation level of 20 dB SL. At such a low SL, resolved spectral components would fall below the absolute threshold. Note that resolved components would not contribute to the modulation produced by the stimuli in the auditory system, so placing the resolved components below threshold would not alter the modulation pattern.

A. Subjects

One subject from the first experiment (JJ) and three new subjects (EE, TE, TP) were tested. All of the new subjects had some musical training. Their absolute thresholds at octave frequencies between 500 and 4000 Hz were within 15 dB of the ANSI (1969) standard.

B. Stimuli

The stimuli were similar to those of the first experiment, except that all sounds were presented at 20 dB SL. All subjects were tested using a carrier frequency of 2 kHz, with an FM depth of 8%, and modulation rates of 5, 10, 20, and 50 Hz. Only phase 2 was used, with the modulation starting halfway through the plateau in instantaneous frequency.

C. Results

The results were similar across subjects. The left panel of Fig. 3 shows geometric means (and standard errors) across subjects for the data obtained at 20 dB SL as solid lines. The dashed lines reproduce the data obtained in experiment 1 at 50 dB SPL. Recall that only one subject (JJ) was common to the two experiments. The individual results for JJ at the two levels are shown in the right panel of Fig. 3. His results are similar to the mean results, suggesting that the effects of level are not due to the use of different subjects at the two levels. The progressive increase in pitch shift with increasing modulation rate which was found at 50 dB SPL is not present at 20 dB SL; if anything, the pitch shifts decline slightly for modulation rates above 10 Hz. A within-subjects ANOVA of the data obtained at 20 dB SL gave a significant effect of shape; $F(1,3) = 99.6$, $p = 0.002$. There was also a significant interaction of modulation rate and shape; $F(3,9) = 31.8$, $p = 0.007$.

The pattern of results is consistent with the idea that the progressive increase in pitch shift with increasing modulation rate found at 50 dB SPL was due to the influence of resolvable components in the spectrum. However, if that was the case, it is puzzling that the pitch shift decreased with increasing modulation rate at 0.5 kHz, as the bandwidth of the auditory filter is markedly lower at 0.5 than at 2 kHz (Glasberg and Moore, 1990). Probably, some other factor is responsible for the changes in pitch shift with modulation rate found at 50 dB SPL. Indeed, analysis of data in the literature (see later for details) suggests that pitch shifts of asymmetrically frequency-modulated tones become smaller when the components are partially resolved. In any case, the fact that pitch shifts occurred at 20 dB SL, and these shifts were in the same direction as and comparable in magnitude to those found at 50 dB SPL, suggests that the pitch shifts occur due to some factor other than the presence of resolvable components in the spectrum.

IV. EXPERIMENT 3: SUPPLEMENTARY DATA AT 4 kHZ

When we started these experiments, we thought that the effects of stability-sensitive weighting on pitch perception might only be observed over the frequency range where phase locking is available, as previous work suggested that poor ability to track frequency changes is observed mainly when pitch is extracted via temporal mechanisms (Carlyon and Shackleton, 1994; Shackleton and Carlyon, 1994; Moore and Sek, 1995, 1996; Plack and Carlyon, 1995; Sek and Moore, 1999, 2000). To our surprise, pitch shifts were also found for the 8-kHz carrier. However, subjects reported that the pitch-matching task was much more difficult at 8 kHz than at the lower frequencies tested, and one subject gave rather erratic results at 8 kHz. Therefore, we decided to gather some supplementary data using a 4-kHz carrier, for which pilot data indicated that reproducible results could be obtained. At this frequency, phase-locking cues are usually assumed to be weak, but perhaps still present (Moore, 1997).

A. Subjects and stimuli

Two of the subjects from experiment 2 were used; JJ and TE. The level was 50 dB SPL, the carrier frequency was 4 kHz, the modulation depth was 8% and modulation rates of 5, 10, and 20 Hz were used. One subject (JJ; filled symbols) was tested using both starting phases, and the other (TE; open symbols) was tested only using phase 1.

B. Results

The results are shown in Fig. 4. Clear pitch shifts were observed that (for JJ) did not vary with modulator phase, and varied only slightly with modulation rate. The overall magnitude of the pitch shifts was similar to that found for the

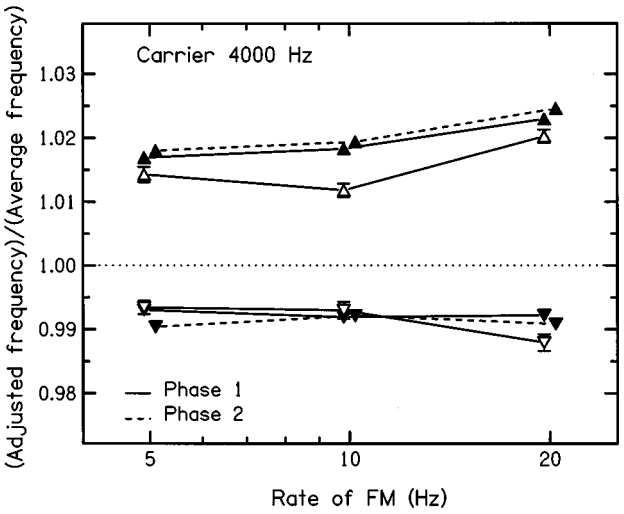


FIG. 4. Data for two subjects run at 50 dB SPL, with $f_c = 4$ kHz, $f_{\text{depth}} = 8\%$, and $f_m = 5$, 10, and 20 Hz. One subject (JJ, filled symbols) was tested with both starting phases, and the other (TE, open symbols) was tested only with phase 1. Otherwise as Fig. 2.

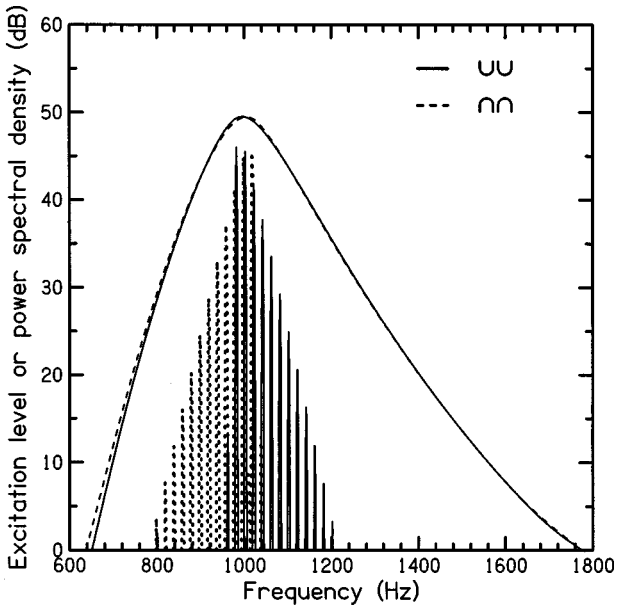


FIG. 5. The solid and dashed lines between 800 and 1200 Hz show long-term spectra for the ∪∪ and ∩∩ stimuli, respectively. To allow the individual spectra to be seen more clearly, the spectrum for the ∪∪ stimulus is shifted 2 Hz to the right, and the spectrum for the ∩∩ stimulus is shifted 2 Hz to the left. The smooth curves show corresponding excitation patterns (but without any shift).

lower carrier frequencies. We conclude that clear pitch shifts can occur for a carrier frequency where phase-locking cues are likely to be weak.

V. DISCUSSION

The results show that the pitch of a sinusoidal carrier frequency modulated with a highly asymmetric modulator does not correspond to the geometric or arithmetic mean frequency. The results are not consistent with previous models proposed to account for the overall pitch of frequency modulated sounds (Feth, 1974; Feth *et al.*, 1982; Iwamiya *et al.*, 1984; Anantharaman *et al.*, 1993; Dai, 1993; d'Alessandro and Castellengo, 1994). We consider next possible ways of accounting for the pitches observed for our stimuli.

A. Possible excitation-pattern cues

It is important to consider whether some cue in the long-term spectra or excitation patterns of the stimuli might give rise to the pitch shifts found in these experiments. Figure 5 shows long-term spectra for two examples of our stimuli. The overall level was 50 dB SPL, the carrier frequency was 1000 Hz, the modulation rate was 20 Hz, and the depth of modulation was 8%. Solid and dashed lines show spectra for the ∪∪ and ∩∩ stimuli, respectively. Figure 5 also shows excitation patterns for these stimuli, calculated using the procedure described by Glasberg and Moore (1990). The excitation patterns do not show any ripples corresponding to the individual components in the spectrum. This is consistent with our interpretation of the results of experiment 2, that the pitch shifts were not due to the presence of resolvable components in the spectrum.

The excitation patterns for the ∪∪ and ∩∩ stimuli are very similar, but are not identical. Two differences can be discerned. First, the low-frequency edge of the excitation pattern is slightly higher in frequency for the ∪∪ stimulus than for the ∩∩ stimulus. Although some researchers have proposed that pitch is related to the position of the low-frequency edge of the excitation pattern (Zwislocki and Nguyen, 1999), the shift observed here is actually in the opposite direction to that needed to explain the observed pitch shifts; the ∪∪ stimulus had a lower pitch than the ∩∩ stimulus, but the low-frequency edge of the excitation pattern was shifted upwards for the ∪∪ stimulus relative to that for the ∩∩ stimulus. Another reason for believing that a shift on the low-frequency side of the excitation pattern is not responsible for the observed pitch shifts is that the shift in the low-frequency side is negligible for points between the peak and 20 dB below the peak, but pitch shifts were observed for stimuli presented at 20 dB SL, and these shifts were comparable to those obtained at 50 dB SPL.

The second observable difference in the excitation patterns is a shift in the position of the peak; the peak is slightly higher in frequency for the ∩∩ stimulus than for the ∪∪ stimulus. This shift is in the right direction to explain the observed pitch shifts, but its magnitude is not sufficient. For example, for a carrier frequency of 1000 Hz, the differences in frequency at the peaks of the excitation patterns were 9.2, 9.2, and 7.9 Hz, for modulation rates of 5, 10, and 20 Hz, respectively. The observed differences in pitch for the ∪∪ and ∩∩ stimuli correspond to frequency differences of about 20 Hz.

We can conclude that there are no obvious features of the excitation patterns that can be used to explain the observed pitch shifts.

B. Possible cues in the long-term autocorrelation function

Several researchers have proposed that pitch might be determined from the autocorrelation function (ACF) of the

stimulus, or of its representation in the auditory system (Licklider, 1951; Meddis and O'Mard, 1997; Yost, 1996). Yost (1996) proposed that the first main peak in the ACF is the determinant of pitch, but it is possible that a few of the higher-order peaks also play a role.

To assess whether the pitch shifts observed in our experiments could be predicted using the long-term ACF, we performed two analyses. The first was based on the ACFs of the stimulus waveforms. This analysis showed that the first peak in the functions was always at a delay corresponding to $1/f_c$. The higher-order peaks showed shifts from integer multiples of $1/f_c$, but these were always very small. For example, for the $\cup\cup$ stimulus with $f_c = 1000$ Hz modulated at an 8-Hz rate with 8% depth, peaks in the ACF up to the fifth were shifted upwards by less than 0.24%. For the corresponding $\cap\cap$ stimulus, peaks up to the fifth were shifted downwards by less than 0.34%.

The second analysis was based on the use of an auditory model whose initial stages were identical to those in the models described by Meddis and Hewitt (1991) and Meddis and O'Mard (1997) (software available from ftp://ftp.essex.ac.uk/pub/omard/dsam). These stages included a filter simulating the transfer function of the outer and middle ear, an array of gammatone filters, and a hair-cell model. They calculated a running ACF for each channel of the model, and then summed across channels to obtain a ''summary ACF.'' The running ACF made use of an exponential decay, with a time constant of 2.5 ms (in Meddis and Hewitt, 1991) or 10 ms (in Meddis and O'Mard, 1997). In practice, they restricted the proportion of the stimulus (the frame length) over which the ACF was calculated to three times the value of the time constant, i.e., to 7.5 or 30 ms. To evaluate the predictions of the model, only the last frame, i.e., the last 7.5 or 30 ms of the stimulus, was used. Thus, their model was designed to determine pitch using relatively brief samples of the stimulus, and it did not include any mechanism which would allow determination of the overall pitch of frequency-modulated stimuli, like those used in our experiments. To assess if the long-term ACF could be used to predict the overall pitch, we increased the time constant to 6203 s, and calculated the ACF over the whole stimulus duration. The first main peak in the summary ACF showed shifts away from $1/f_c$, but these shifts were small (typically less than 0.3%), and did not explain the differences in pitch between the $\cup\cup$ and the $\cap\cap$ stimuli. The higher-order peaks usually showed small shifts in the direction required to explain the data, but of insufficient magnitude (by a factor of 3 for peaks up to the fourth).

We conclude that neither the long-term ACFs of the stimuli nor the long-term summary ACFs derived from an auditory model can be used to explain the observed pitch shifts.

C. A new model for the pitch of sounds with simultaneous frequency and amplitude modulation

One way of interpreting the pitch shifts observed in our experiments is based on the idea that portions of the sound where the frequency is changing rapidly receive less weight than portions where the frequency is changing more slowly. However, it seems plausible that the auditory system would derive pitch from estimates of the period of the sound, which would be directly reflected in interspike intervals in the auditory nerve, rather than from estimates of the instantaneous frequency, as assumed in the EWAIF and IWAIF models. Hence, we assumed that the overall pitch is based on a weighted mean of successive estimates of the waveform period, where the weights, w_i, are a function of the rate of change of period. The periods were estimated from the digital representation of a given waveform. We tried two methods for estimating the periods. The first involved locating maxima in the waveform, and estimating periods from the times between successive maxima. In order to locate the maxima accurately in time, it was necessary to use very high sampling rates (typically above 20 MHz). The second method involved estimating the periods from the times between successive positive-going zero crossings, which did not require such high sampling rates. In what follows, we present the results obtained using this second method.

As a measure of the short-term rate of change of period for period i, P_i, we calculated the mean of five successive periods (P_{mi}), from $i-2$ to $i+2$, and the standard deviation (s.d.$_i$) across those periods, and took the ratio s.d.$_i/P_{mi}$, which will be denoted R_i. We chose to estimate the rate of change over a fixed number of periods rather than a fixed duration for several reasons. First, several aspects of pitch perception seem to depend more on number of periods than on duration. These include the point at which a very brief tone starts to sound tone-like rather than click-like (Doughty and Garner, 1948) and the duration below which frequency discrimination starts to worsen markedly with decreasing duration (Moore, 1973). Second, models of pitch perception based on the extraction of timing information in the auditory nerve often assume a matched filter or template that uses information over a range of interspike intervals corresponding to a fixed number of periods rather than a fixed duration (Srulovicz and Goldstein, 1983).

We assumed that the weight associated with a given period, w_i, was a function of R_i. The weighted mean period, P, is given by

$$P = \Sigma(w_i(R_i) \times P_i)/\Sigma w_i(R_i). \tag{7}$$

The perceived pitch is assumed to be related to $1/P$. We tried several different functions relating the weights $w_i(R_i)$ to R_i. Our goal was to find a function that would lead to predictions consistent with the main features of our results, namely that the shifts in pitch were almost invariant with carrier frequency and modulation rate, and were typically slightly over $\pm 1\%$. We found that the following function achieved this goal:

$$w_i(R_i) = 1/(1 + KR_i^{0.5}), \tag{8}$$

where K is a constant. We found that a value of $K = 600$ led to predicted pitch shifts of about $+1.1\%$ for the $\cap\cap$ stimuli and -1% for the $\cup\cup$ stimuli, and the predicted shifts were almost independent of carrier frequency and modulation rate; the only ''deviant'' values were shifts of $+1.0\%$ and -0.9% for the 4-kHz carrier modulated at a rate of 5 Hz. Note that,

for a sinusoidal stimulus, the model predicts a pitch value equal to the frequency of the stimulus.

It is clear that the general pattern of our results can be accounted for using the assumption that the overall pitch of a frequency-modulated sound is determined from a weighted sum of short-term period estimates, with the weight attached to a given estimate being inversely related to the short-term rate of change of period. However, this cannot be the whole story. In particular, the short-term amplitude must also play a role. This is indicated by the results of Tiffin (1931) and Iwamiya *et al.* (1984). They used sinusoidal carriers that were simultaneously amplitude and frequency modulated, and showed that the overall pitch was affected by the phase of the AM relative to the FM; the pitch was higher when the AM and FM were in phase (amplitude increases coinciding with frequency increases) than when they were 180° out of phase.

To take account of the effect of amplitude fluctuations, we propose a more general version of Eq. (7) in which the weight associated with a given period, P_i, is a joint function of amplitude, A_i, and the ratio R_i. The amplitude, A_i, corresponding to a given period, P_i, was taken as the value at the waveform maximum following the first zero crossing defining P_i. The weighted average period is given by

$$P = \Sigma(W_i(A_i, R_i) \times P_i) / \Sigma W_i(A_i, R_i). \tag{9}$$

The overall weight for a given period is defined as

$$W_i(A_i, R_i) = w_i(R_i) \times w_i(A_i). \tag{10}$$

The weight associated with a given value of A_i was assumed to have the form

$$w_i(A_i) = (A_i / \max A)^{\alpha}, \tag{11}$$

where $\max A$ is the maximum value of A_i over all A_i, and the exponent, α, is a free parameter of the model. We describe the model defined by Eqs. (9)–(11) as the weighted average period (WAP) model. Note that, when the amplitude is constant, Eq. (9) predicts the same pitch as Eq. (7), and hence accounts for our data.

To estimate an appropriate value for α, we used the data of Feth *et al.* (1982). They measured pitch shifts for stimuli composed of two sinusoidal components. The component frequencies were centered at 500, 1000, 2000, or 4000 Hz, and were separated by 10, 20, 50, or 100 Hz. For each pair of components, the level differences between the components were 0.5, 1, or 3 dB. It is a feature of the data of Feth *et al.* that for a given frequency separation and level difference, the pitch shifts were roughly invariant with center frequency when the frequency separation was small. However, when the frequency separation was larger than about half the equivalent rectangular bandwidth of the auditory filter (ERB, see Glasberg and Moore, 1990), the shift was smaller than for the same frequency separation at a higher center frequency, for which the components would have been separated by less than 0.5 ERB. It seems likely that this effect occurred because the tones in a pair were partially resolved by the peripheral auditory system when their separation exceeded 0.5 ERB. When the tones are partially resolved, the effective difference in level between the tones varies markedly across auditory filters with different center frequencies, and it is no longer reasonable to treat the tone pair as a single narrow-band stimulus. For this reason, we concentrated on fitting the data of Feth *et al.* for frequency separations of the two tones of 20 Hz or less at 500 Hz, 50 Hz or less at 1000 Hz, and 100 Hz or less at 2000 Hz. We did not try to fit the data for a center frequency of 4000 Hz, as the results varied markedly across subjects. We also did not fit the data for a level difference of 0.5 dB as these were incomplete. Finally, we excluded the data of one of their subjects (S3), whose data were very erratic.

We used the WAP model to predict the pitch shifts observed by Feth *et al.* for the conditions specified above. They used as a matching stimulus a two-component complex with equal-amplitude components and the same frequency separation as for the test stimulus. Hence, to predict their results, we determined what center frequency of the matching stimulus was required to predict the same pitch as a given test stimulus. The value of K was fixed at 600, a value giving a good fit to our data. We found that a value of α of 0.42 gave the best overall fit to the data.

One feature of the data of Feth *et al.* was that the positive pitch shifts, produced when the component with the higher frequency was also higher in level (called hereafter the SH stimulus), were often markedly larger than the negative shifts occurring when the higher-frequency component was lower in level (the SL stimulus). Such an asymmetry is not predicted by the EWAIF or IWAIF models and is also not predicted by our model (the version of the model in which periods were estimated from waveform maxima did predict larger pitch shifts for the SH than for the SL stimuli, but it also predicted positive shifts for the matching stimulus. Such shifts for the matching stimulus were not found to occur by Stover and Feth (1983), so we conclude that period estimates based on zero crossing are more consistent with the empirical data). Figure 6 compares the pitch shifts obtained by Feth *et al.* with the predictions of the WAP model. Following the convention adopted by Feth *et al.*, the figure shows the difference in pitch match between the SH and SL stimuli. The parameter is the difference in level of the tones, which was 1 dB (open symbols data, dashed lines predictions) or 3 dB (filled symbols data, solid lines predictions). Cases are only shown where the component frequencies were separated by less than 0.5 ERB. It is clear that the model fits the data very well, in terms of the magnitudes of the shifts and the way that they vary with frequency separation of the components and with difference in level of the components.

We turn now to other data in the literature. Keeping the same parameter values ($K = 600$, $\alpha = 0.42$), we used the WAP model to predict the mean data of Dai (1993). Dai also measured pitch shifts for stimuli composed of two sinusoidal components. The component frequencies were centered around 500 or 1000 Hz; the separation was 40 Hz at 500 Hz and 80 Hz at 1000 Hz. For each pair of components, the level differences between the components were 2.4, 3.4, or 4.4 dB. The starting phases were always 0 and π, for the lower and the higher frequency component, respectively. All of Dai's conditions involved frequency separations of the components greater than 0.5 ERB. Thus, the components

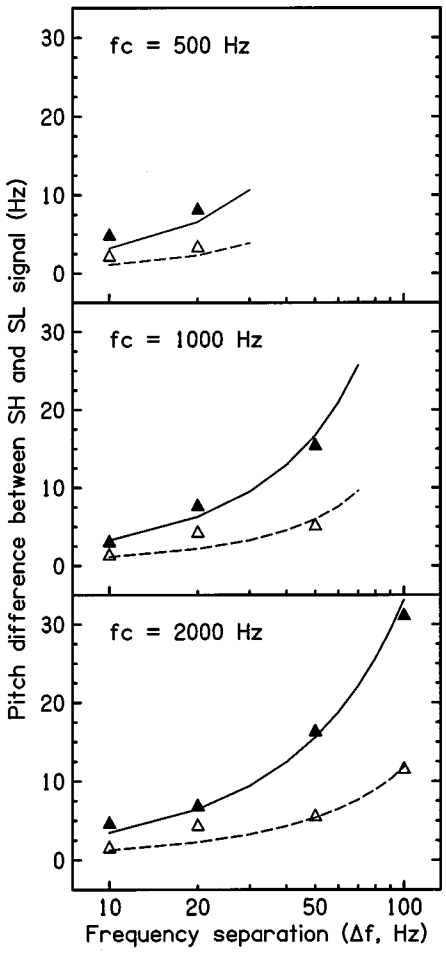


FIG. 6. Comparison of the data of Feth *et al.* (1982) on the pitch of two-tone complexes with the predictions of the WAP model. The differences between the pitch matches for the SH and SL two-tone complexes are plotted as a function of the frequency separation between the two tones, Δf. The parameter is the difference in level of the tones, which was 1 dB (open symbols data, dashed lines predictions) or 3 dB (filled symbols data, solid lines predictions).

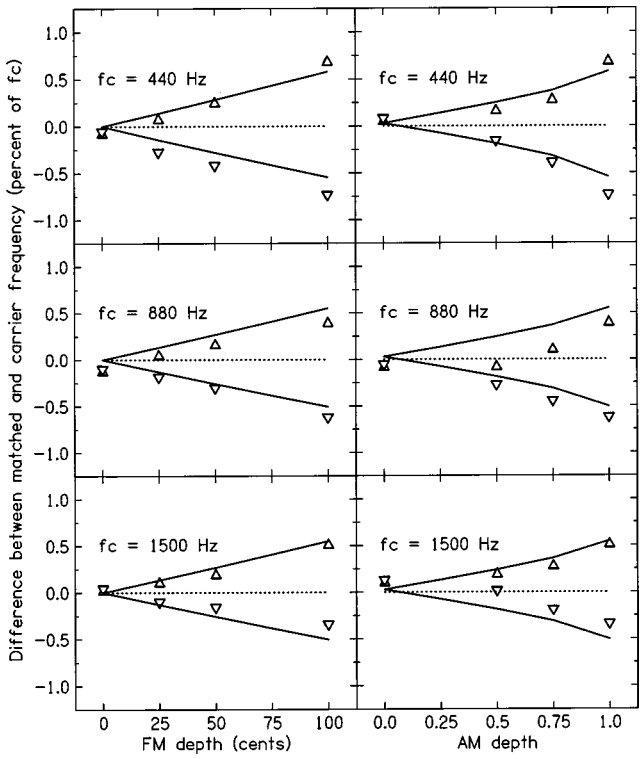


FIG. 7. Comparison of the results of Iwamiya *et al.* (1984) with the predictions of the WAP model. Subjects were required to adjust a sinusoid to match the pitch of a sinusoidal carrier (frequency $f_c = 440$, 880, or 1500 Hz) with simultaneous 6-Hz triangular AM and FM. For the results presented in the left column, the AM depth was fixed at 1 (100%) and the FM depth was varied. For the results presented in the right column, the FM depth was fixed at 100 cents, and the AM depth was varied. The difference between the matched frequency and f_c is plotted as a percentage of f_c. Open symbols show the mean data across subjects; upward-pointing triangles indicate that the AM and FM were in phase, while downward-pointing triangles indicate that the AM and FM were in antiphase. Solid lines show the predictions of the WAP model.

would have been partially resolved in the peripheral auditory system. For this reason, it is to be expected that the model predicts larger pitch shifts than observed, and this was the case. For each test stimulus (SH or SL), we predicted the pitch and then calculated the center frequency of the matching stimulus (a two-component complex with equal-amplitude components) needed to obtain the same pitch. Table I compares differences between obtained matches to the SH and SL stimuli with predicted differences.

It is noteworthy that the discrepancy between obtained and predicted values is greater for the center frequency of 1000 Hz than for the center frequency of 500 Hz. This is consistent with the idea that the discrepancy depends on partial resolution of the components in the peripheral auditory system. The frequency separation of 40 Hz at 500 Hz corresponds to about 0.51 ERB, whereas the separation of 80 Hz at 1000 Hz corresponds to 0.6 ERB.

TABLE I. Comparison of the differences of the matching frequency (in Hz) between SH and SL signals obtained by Dai (1993) with the differences predicted by the WAP model. The center frequency is denoted f_c (Hz) and the frequency separation of the components is denoted Δf (Hz).

			Level difference, dB		
f_c	Δf		2.4	3.4	4.4
500	40	Obtained	9.0	13.4	16.5
		Predicted	12.4	16.8	20.7
1000	80	Obtained	16.8	25.4	30.1
		Predicted	24.9	33.8	41.5

We also used the WAP model to predict the mean data of Iwamiya *et al.* (1984), again keeping the same parameter values. They used a sinusoidal carrier ($f_c = 440$, 880, or 1500 Hz) simultaneously modulated with triangular AM and FM with a rate of 6 Hz, where the AM and FM were either in phase or in antiphase. A sinusoid was adjusted in frequency to match the pitch of the modulated carrier. In one case, they kept the AM depth fixed at 1 (100%), and varied the amount of FM up to 100 cents (one semitone, corresponding to a frequency ratio between the maximum and minimum frequencies of 1.059). The left column of Fig. 7 shows the results for this case. The difference between the matched frequency and f_c is plotted as a percentage of f_c. Open symbols show the mean data across subjects (excluding the results for their subject B at 1500 Hz, for whom there were ''large and irregular biases''; upward-pointing triangles: AM and FM in phase; downward-pointing triangles: antiphase),

and solid lines show the predictions of the WAP model. The predictions fit the data very well. In a second case, the amount of FM was fixed at 100 cents, and the AM depth was varied up to 1. The right column of Fig. 7 shows the results for this case. Again, the fit is good, although there are some deviations for $f_c = 880$ Hz, which appear to reflect a bias in the matches (a downward shift). However, it should be noted that the standard deviation of the pitch matches across subjects was typically about 0.4%, and that the largest deviation of any data point from the predicted value is about 0.3%. Thus, there is no substantial deviation of the data from the predicted values.

The best-fitting value of $\alpha = 0.42$ deserves some comment. The fact that the value is less than 1 indicates that amplitude fluctuations play a smaller role in determining pitch than assumed in the EWAIF or IWAIF models. Effectively, the amplitude is compressed. The model could equally well be formulated by saying that the weighting associated with a given period is proportional to the short-term intensity raised to the power 0.21. This compression may reflect the compression that has been observed on the basilar membrane for midrange sound levels (Rhode and Robles, 1974; Sellick *et al.*, 1982; Robles *et al.*, 1986; Ruggero *et al.*, 1997). Basilar-membrane compression for midrange sound levels typically corresponds to a power function with an exponent between about 0.2 and 0.35 (Oxenham and Moore, 1994; Moore and Oxenham, 1998; Oxenham and Plack, 1997). Our exponent falls within this range if the model is formulated in terms of short-term intensity rather than short-term amplitude. If the exponent does reflect compression on the basilar membrane, then the pitch shifts observed for stimuli of the type used by Feth *et al.* (1982), Dai (1993), and Iwamiya *et al.* (1984) should be level dependent, as basilar-membrane compression is reduced or absent at very low levels and (perhaps) at very high levels. This prediction remains to be tested.

In summary, we have proposed a model, the WAP model, in which the overall pitch of a narrow-band stimulus is determined as a weighted average of estimates of the period; the weights are a joint function of amplitude and the short-term rate of change of period. The WAP model fits both the data presented in this paper, and earlier data of Feth *et al.* (1982) and Iwamiya *et al.* (1984), at least for stimuli where the components would not be well resolved in the peripheral auditory system.

D. Interpretation of pitch shifts at 8 kHz

We have interpreted the pitch shifts observed for our stimuli as the result of a stability-sensitive weighting mechanism based on the use of phase-locking information. This interpretation is somewhat weakened by our finding of significant pitch shifts for a carrier frequency of 8 kHz, a frequency where phase-locking cues are usually assumed not to play a role in pitch perception (Moore, 1997). However, it is possible that very weak phase-locking cues persist even at 8 kHz (Johnson, 1980; Palmer and Russell, 1986). Some researchers have modeled the frequency discrimination of pulsed tones in terms of these residual phase-locking cues (Goldstein and Srulovicz, 1977; Heinz *et al.*, 2001). Indeed, Heinz *et al.* (2001) concluded ''There is significant information in the auditory nerve for frequency discrimination up to at least 10 kHz, and thus temporal schemes cannot be rejected at high frequencies based on the decrease in phase locking in the auditory nerve.'' The weak phase-locking cues may be used in assigning a pitch value to a stimulus in a pitch-matching task, like that used in our experiments. For other types of tasks, for example, the detection of FM imposed on high carrier frequencies, subjects may perform better by making use of the AM of the excitation pattern that is induced by the FM (Zwicker, 1952; Moore and Sek, 1994). In such a case the subject simply has to detect the AM, and it is not necessary to assign a precise pitch value to the stimulus.

An alternative possibility is that the stability-sensitive weighting suggested by our results occurs for pitches extracted from place cues as well as for pitches extracted using temporal cues. It may be difficult to extract precise estimates of the place of stimulation when the frequency is changing rapidly, so overall estimates of place-based pitch may depend mostly on portions of the stimulus where the frequency is relatively stable.

VI. CONCLUSIONS

The following conclusions can be drawn from this study:

(1) For sinusoidal carriers with constant amplitude and asymmetrical patterns of FM, the pitch was shifted away from the pitch corresponding to the mean frequency. The shift was upwards for the $\cap\cap$ stimuli and downwards for the $\cup\cup$ stimuli. This is inconsistent with the predictions of the EWAIF and IWAIF models. The direction of the pitch shifts is consistent with the idea that less weight is given to portions of the stimulus where the frequency is changing more rapidly.
(2) Although some small effects of modulation rate were found, to a first approximation the shifts were independent of modulation rate.
(3) For an overall FM depth of 8% of the carrier frequency, the shifts were roughly a constant proportion (about 1%) of the carrier frequency, over the range 500 Hz to 8 kHz.
(4) For a carrier frequency of 2 kHz, the shifts were similar for levels of 50 dB SPL and 20 dB SL.
(5) Calculated excitation patterns based on the long-term spectra of the stimuli did not show any features that could be used to explain the observed pitch shifts.
(6) Neither the long-term autocorrelation functions (ACFs) of the stimuli, nor the long-term summary ACFs derived from an auditory model can be used to explain the observed pitch shifts.
(7) The main features of both the present results, and of earlier results on the pitch of two-tone complexes and tones with simultaneous AM and FM, can be explained by a model, the WAP model, in which the overall pitch is assumed to be based on a weighted mean of estimates of the period. The weight attached to a given estimate is

assumed to be inversely related to the short-term rate of change of period and directly related to a compressive function of the amplitude at that period.

ACKNOWLEDGMENTS

This work was supported by the MRC (UK). The work was partially conducted while the first author was employed at the CNBH, Dept. of Physiology, University of Cambridge, on MRC Grant No. G9900362. We thank Brian Glasberg for his unstinting assistance, Katrin Krumbholz for assistance with modeling the summary ACFs, and Geoff Moore and Stefan Uppenkamp for comments on an earlier version of this paper. We also thank Les Bernstein, Chris Plack, and an anonymous reviewer for very helpful comments and constructive suggestions.

Anantharaman, J. N., Krishnamurthy, A. K., and Feth, L. L. (**1993**). ''Intensity-weighted average of instantaneous frequency as a model for frequency discrimination,'' J. Acoust. Soc. Am. **94**, 723–729.

ANSI (**1969**). *S3.6. Specifications for Audiometers* (American National Standards Institute, New York).

Carlyon, R. P., Moore, B. C. J., and Micheyl, C. (**2000**). ''The effect of modulation rate on the detection of frequency modulation and mistuning of complex tones,'' J. Acoust. Soc. Am. **108**, 304–315.

Carlyon, R. P., and Shackleton, T. M. (**1994**). ''Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?,'' J. Acoust. Soc. Am. **95**, 3541–3554.

Dai, H. (**1993**). ''On the pitch of two-tone complexes,'' J. Acoust. Soc. Am. **94**, 730–734.

d'Alessandro, C., and Castellengo, M. (**1994**). ''The pitch of short-duration vibrato tones,'' J. Acoust. Soc. Am. **95**, 1617–1630.

Doughty, J. M., and Garner, W. M. (**1948**). ''Pitch characteristics of short tones. II. Pitch as a function of duration,'' J. Exp. Psychol. **38**, 478–494.

Feth, L. L. (**1974**). ''Frequency discrimination of complex periodic tones,'' Percept. Psychophys. **15**, 375–378.

Feth, L. L., O'Malley, H., and Ramsey, J. J. (**1982**). ''Pitch of unresolved, two-component complex tones,'' J. Acoust. Soc. Am. **72**, 1403–1412.

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

Gockel, H., Carlyon, R. P., and Moore, B. C. J. (**2000**). ''Pitch of asymmetrically frequency-modulated tones,'' Br. J. Audiol. **34**, 99.

Goldstein, J. L., and Srulovicz, P. (**1977**). ''Auditory-nerve spike intervals as an adequate basis for aural frequency measurement,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).

Hartmann, W. M. (**1997**). *Signals, Sound, and Sensation* (AIP, Woodbury, NY).

Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2000**). ''Evaluating auditory performance limit: I. One-parameter discrimination using a computational model for the auditory nerve,'' Neural Comput. (in press).

Helmholtz, H. L. F. (**1863**). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (F. Vieweg, Braunschweig).

Iwamiya, S., Kosugi, K., and Kitamura, O. (**1983**). ''Perceived principal pitch of vibrato tones,'' J. Acoust. Soc. Jpn. (E) **4**, 73–82.

Iwamiya, S., Nishikawa, S., and Kitamura, O. (**1984**). ''Perceived principal pitch of FM–AM tones when the phase difference between frequency modulation and amplitude modulation is in-phase and anti-phase,'' J. Acoust. Soc. Jpn. (E) **5**, 59–69.

Johnson, D. H. (**1980**). ''The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones,'' J. Acoust. Soc. Am. **68**, 1115–1122.

Licklider, J. C. R. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–133.

McClelland, K. D., and Brandt, J. F. (**1969**). ''Pitch of frequency-modulated sinusoids,'' J. Acoust. Soc. Am. **45**, 1489–1498.

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

Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.

Miller, G. A., and Heise, G. A. (**1950**). ''The trill threshold,'' J. Acoust. Soc. Am. **22**, 637–638.

Moore, B. C. J. (**1973**). ''Frequency difference limens for short-duration tones,'' J. Acoust. Soc. Am. **54**, 610–619.

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, San Diego).

Moore, B. C. J., and Oxenham, A. J. (**1998**). ''Psychoacoustic consequences of compression in the peripheral auditory system,'' Psychol. Rev. **105**, 108–124.

Moore, B. C. J., and Sek, A. (**1994**). ''Effects of carrier frequency and background noise on the detection of mixed modulation,'' J. Acoust. Soc. Am. **96**, 741–751.

Moore, B. C. J., and Sek, A. (**1995**). ''Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (AM vs FM),'' J. Acoust. Soc. Am. **97**, 2468–2478.

Moore, B. C. J., and Sek, A. (**1996**). ''Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking,'' J. Acoust. Soc. Am. **100**, 2320–2331.

Oxenham, A. J., and Moore, B. C. J. (**1994**). ''Modeling the additivity of nonsimultaneous masking,'' Hear. Res. **80**, 105–118.

Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.

Palmer, A. R., and Russell, I. J. (**1986**). ''Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells,'' Hear. Res. **24**, 1–15.

Plack, C. J., and Carlyon, R. P. (**1995**). ''Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics,'' J. Acoust. Soc. Am. **98**, 1355–1364.

Rhode, W. S., and Robles, L. (**1974**). ''Evidence from Mössbauer experiments for non-linear vibration in the cochlea,'' J. Acoust. Soc. Am. **55**, 588–596.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input–output functions, tuning curves, and response phases,'' J. Acoust. Soc. Am. **80**, 1364–1374.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.

Seashore, C. (**1938**). *Psychology of Music* (McGraw-Hill, New York).

Sek, A., and Moore, B. C. J. (**1999**). ''Discrimination of frequency steps linked by glides of various durations,'' J. Acoust. Soc. Am. **106**, 351–360.

Sek, A., and Moore, B. C. J. (**2000**). ''Detection of quasitrapezoidal frequency and amplitude modulation,'' J. Acoust. Soc. Am. **107**, 1598–1604.

Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). ''Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique,'' J. Acoust. Soc. Am. **72**, 131–141.

Shackleton, T. M., and Carlyon, R. P. (**1994**). ''The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,'' J. Acoust. Soc. Am. **95**, 3529–3540.

Shonle, J. I., and Horan, K. E. (**1976**). ''Trill threshold revisited,'' J. Acoust. Soc. Am. **59**, 469–471.

Shonle, J. I., and Horan, K. E. (**1980**). ''The pitch of vibrato tones,'' J. Acoust. Soc. Am. **67**, 246–252.

Srulovicz, P., and Goldstein, J. L. (**1983**). ''A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum,'' J. Acoust. Soc. Am. **73**, 1266–1276.

Stover, L. J., and Feth, L. L. (**1983**). ''Pitch of narrow-band signals,'' J. Acoust. Soc. Am. **73**, 1701–1707.

Sundberg, J. (**1978a**). ''Effects of the vibrato and the singing formant on pitch,'' Musicologica Slovaca **6**, 51–69.

Sundberg, J. (**1978b**). ''Synthesis of singing,'' Swedish J. Musicol. **60**, 107–112.

Tiffin, J. (**1931**). ''Some aspects of the psychophysics of the vibrato,'' Psychol. Rev. Monogr. **41**, 152–200.

Yost, W. A. (**1996**). ''Pitch of iterated rippled noise,'' J. Acoust. Soc. Am. **100**, 511–518.

Zwicker, E. (**1952**). ''Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones,'' Acustica **2**, 125–133.

Zwislocki, J. J., and Nguyen, N. (**1999**). ''Place code for pitch: A necessary revision,'' Acta Oto-Laryngol. **119**, 140–145.

Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants

Luc Geurts[a] and Jan Wouters[b]
Laboratory for Experimental ORL, KULeuven, Kapucijnenvoer 33, B 3000 Leuven, Belgium

(Received 6 December 1999; revised 31 August 2000; accepted 21 November 2000)

In this study the perception of the fundamental frequency ($F0$) of periodic stimuli by cochlear implant users is investigated. A widely used speech processor is the Continuous Interleaved Sampling (CIS) processor, for which the fundamental frequency appears as temporal fluctuations in the envelopes at the output. Three experiments with four users of the LAURA (Registered trade mark of Philips Hearing Implants, now Cochlear Technology Centre Europe) cochlear implant were carried out to examine the influence of the modulation depth of these envelope fluctuations on pitch discrimination. In the first experiment, the subjects were asked to discriminate between two SAM (sinusoidally amplitude modulated) pulse trains on a single electrode channel differing in modulation frequency ($\delta f = 20\%$). As expected, the results showed a decrease in the performance for smaller modulation depths. Optimal performance was reached for modulation depths between 20% and 99%, depending on subject, electrode channel, and modulation frequency. In the second experiment, the smallest noticeable difference in $F0$ of synthetic vowels was measured for three algorithms that differed in the obtained modulation depth at the output: the default CIS strategy, the CIS strategy in which the $F0$ fluctuations in the envelope were removed (FLAT CIS), and a third CIS strategy, which was especially designed to control and increase the depth of these fluctuations ($F0$ CIS). In general, performance was poorest for the FLAT CIS strategy, where changes in $F0$ are only apparent as changes of the average amplitude in the channel outputs. This emphasizes the importance of temporal coding of $F0$ in the speech envelope for pitch perception. No significantly better results were obtained for the $F0$ CIS strategy compared to the default CIS strategy, although the latter results in envelope modulation depths at which sub-optimal scores were obtained in some cases of the first experiment. This indicates that less modulation is needed if all channels are stimulated with synchronous $F0$ fluctuations. This hypothesis is confirmed in a third experiment where subjects performed significantly better in a pitch discrimination task with SAM pulse trains, if three channels were stimulated concurrently, as opposed to only one. [DOI: 10.1121/1.1340650]

PACS numbers: 43.66.Ts, 43.66.Fe, 43.66.Hg [SPB]

I. INTRODUCTION

Voiced speech phonemes are the result of a periodic vibration of the vocal chords, followed by a filtering in the vocal tract. Such signals can be decomposed into a series of discrete sinusoids, called the harmonics, and the frequencies of these harmonics are an integer multiple of a common frequency, the fundamental frequency ($F0$). $F0$ is related to the pitch of a sound, and provides several cues for the perception of speech and music. It provides cues to intonation, question–statement contrasts, the identity of the speaker, and emotional state. Also, it is essential for the recognition of a melody, sung or played on a musical instrument. Yet, not all the harmonics—even not the first one—have to be present in the sound, in order to evoke a pitch corresponding to $F0$. This low pitch associated with a group of harmonics is often called ''residue pitch'' or ''periodicity pitch'' (Schouten *et al.*, 1962). The phenomenon of the residue pitch demonstrates that $F0$ is not solely coded by the auditory nerves at the tonotopic place corresponding to $F0$, but also in the other regions of the cochlea. The response of a specific auditory nerve fiber to a vowel, or any harmonic complex, depends on the fiber's characteristic frequency and on the amplitudes of the harmonics in the vicinity of that frequency. In the case where the nearest harmonic is resolved by the cochlear filter or where this harmonic is very dominant compared to the surrounding ones (e.g., the harmonic closest to a formant frequency), the time representation of the neural response shows little or no modulation according to $F0$, but to the harmonic (Miller and Sachs, 1984; Palmer, 1990). On the other hand, nerve fibers with a characteristic frequency between two formants respond to several harmonics, and due to beating, their response shows a clear envelope modulation according to $F0$ (Miller and Sachs, 1984; Palmer, 1990). The important role of the temporal coding of $F0$ was illustrated in a study of Cariani and Delgutte (1996). They predicted the pitch of several complex tones on the basis of interspike intervals between successive and nonsuccessive spikes, so-called all-order intervals. The most frequent all-order interval present in the distribution pooled over many fibers corresponds to the pitch heard in psychophysical experiments.

Only a limited number of electrode channels are available in a cochlear implant system, and if the whole speech

[a]Electronic mail: Luc.Geurts@uz.kuleuven.ac.be
[b]Electronic mail: Jan.Wouters@uz.kuleuven.ac.be

frequency range is divided over these channels, a relatively large frequency region is assigned to each of them. This results in poor frequency selectivity, or in other words, harmonics are likely to be unresolved. In that case, the outputs of the band-pass filters will show a modulation according to $F0$. In most common speech processors, the envelope in each frequency band is typically estimated by low-pass filtering the rectified output of the band-pass filters. If the cutoff frequency of the low-pass filter is chosen above the range of typical $F0$'s of speech, then $F0$ appears as a temporal fluctuation in the speech envelope (for an illustration, see Fig. 3). In a Continuous Interleaved Sampling (CIS) processor (Wilson *et al.*, 1991), brief pulses are presented to each electrode channel at a high fixed rate, typically around 1000 pps (pulses per second). These pulse trains are modulated with the extracted speech envelopes. In feature extraction algorithms that are or were used in the Nucleus cochlear implant system, $F0$ itself is also estimated from the input signal and determines the repetition rate of the pulses applied on the electrodes (Blamey *et al.*, 1987; Skinner *et al.*, 1991). The sensitivity of a cochlear implantee to changes in pulse presentation rate or modulation frequency is described in several psychophysical studies. An unmodulated pulse train with a higher repetition rate will elicit a higher pitch (Eddington *et al.*, 1978; Shannon, 1983; Tong and Clark, 1985; Townshend *et al.*, 1987). This is often called ''rate pitch'' as opposed to the ''place pitch,'' the pitch according to the tonotopic organization of the cochlea. The highest frequency limit for which a subject still can detect a change in pitch when the rate changes is strongly dependent on the subject, even if the same type of device and stimuli are used. This limit generally lies above 300 Hz (Eddington *et al.*, 1978; Simmons *et al.*, 1981; Blamey *et al.*, 1984), which implies that temporal coding of $F0$ is feasible. Wilson (1997) presented SAM pulse trains in the cochlea and investigated the neural response by measuring action potentials. He concluded that the ratio between carrier rate and modulation rate should be at least four in order to obtain a good representation of the stimulus envelope. This ratio, obtained from physiological measurements, is in agreement with psychophysical data (McKay *et al.*, 1994; Wilson, 1997). McKay *et al.* (1994) investigated the pitch associated with SAM pulse trains and reported that these stimuli could be ranked by modulation frequency in the same way that unmodulated stimuli could be ranked by pulse rate, provided that the carrier rate is high enough (at least 800 Hz for a modulation frequency of 200 Hz, so a ratio of four). The LAURA[1] cochlear implant can stimulate at an overall rate of 10 000 pps, so implementing the CIS strategy on all eight channels results in a pulse rate of 1250 pps on each channel. Taking the ratio of four into account implies that modulation frequencies up to about 300 Hz can be encoded correctly on a single channel, which is well above the range of fundamental frequencies of normal speech.

In the first part of this study, it is investigated how the modulation depth of the envelope fluctuation influences the discrimination ability of two SAM pulse trains on a single channel with different modulation frequencies. McKay *et al.* (1995) showed that if the modulation depth of a modulated pulse train is too small, a pitch corresponding to the carrier frequency is perceived, instead of the modulating frequency.[2] So, it is expected that performance degrades for smaller values of the modulation depth, since the carrier frequency is kept constant. In the second part of this study, the just noticeable $F0$ difference is measured for three CIS strategies, which stimulate all available channels, using synthetic vowels. A change of $F0$ has two effects at the output of the CIS strategy. First, the average amplitude in each channel will change, mainly due to a different attenuation of the harmonics in the sidebands of the band-pass filters (for an illustration, see Fig. 5). This possible source of information will be referred to as the ''average amplitude cue.'' Second, the period of the fluctuations in the envelope, which corresponds to the period of $F0$, will change too. This possible source of temporal information will be referred to as the ''envelope modulation cue.'' In the first strategy (FLAT CIS), the temporal $F0$ fluctuations are removed, so there is only an ''average amplitude cue.'' The second strategy (CIS) is the default strategy, in which ''envelope modulation cues'' are present also. In the third strategy ($F0$ CIS), the envelope fluctuations are enhanced compared to the default strategy, by increasing the modulation depth of these fluctuations. A comparison of the first two algorithms permits the determination as to what extent the ''envelope modulation cue'' improves pitch perception corresponding to $F0$. Jones *et al.* (1995) conducted a related study based on the SMSP strategy, developed for the Nucleus device (McDermott *et al.*, 1992). They concluded that the explicit inclusion of a temporal cue according to $F0$ did not increase the performance in several tasks involving pitch perception. However, they did not accurately measure the pitch discrimination abilities of their subjects for the different strategies. By comparing the results of the second and the third CIS algorithm, it is possible to determine whether the enhancement of the ''envelope modulation cue'' results in better frequency discrimination.

II. EXPERIMENT I

A. Subjects

Four post-lingually deafened users of the LAURA cochlear implant participated in this study. Some details about the subjects can be found in Table I. It was essential for this study that they were sensitive to changes of 20% of the modulation frequency of a SAM pulse train, at least in the lower frequency range around 150 Hz and on one electrode pair. This sensitivity was investigated in a pilot experiment, which resulted in the rejection of four other subjects for further testing. No relation between this sensitivity and any objective parameter could be found. Moreover, one subject, who was an experienced musician before he became suddenly deaf, could not consistently identify the stimulus with the higher pitch. However, some of them were sensitive to larger differences of the modulation frequency.

Two versions of the electrode array of the LAURA implant exist, and both consist of eight bipolar electrode pairs, called channels. These are numbered from apex to base. Subjects SV and JH use type 5.6: the distance between the two

TABLE I. Some details of the four LAURA subjects who participated in this study. Columns 6 and 7 indicate the channels that were selected for experiment I. The last column indicates the number of active channels used in experiment II. The same channels are used daily in the subjects' clinical processors.

Subject	Age (yrs)	Duration of profound deafness (yrs)	Etiology	Implant experience (yrs)	Apical channel	Basal channel	Number of active electrode channels
SV	30	6	unknown	5	2	6	8
WM	49	5	progressive	2	2	7	8
AE	42	3	Menière	3	4	8	7
JH	17	3	meningitis	4	2	7	8

electrodes of one channel is 1.3 mm, and the distance between the midpoints of adjacent channels is 2.05 mm. For subjects AE and WM, who use type 5.8, these values are, respectively, 3 mm and 2 mm, which means that their channels physically overlap. For experiment I, two channels were selected: a channel close to the apical end of the electrode array and a channel close to the basal end.

B. Stimuli and procedure

All stimuli were SAM pulse trains with a carrier pulse rate of 1250 pps. Pulses were biphasic and had a pulse width of 40 μs per phase. In each trial, two stimuli with different modulation frequencies (df=20%) were presented and the subjects had to indicate the stimulus with the higher pitch. An answer was correct if the stimulus with the higher modulation frequency was selected. Correct response feedback was always given on the computer screen. The length of each stimulus was 400 ms, and the silent interval between the stimuli was 500 ms. The two frequencies in each pair were either 150 Hz and 180 Hz, or 250 Hz and 300 Hz. These frequencies are within the range of typical fundamental frequencies of the human voice. The order of presentation in each trial was randomized. The level of the largest pulse of each SAM pulse train always corresponded to the upper limit of the dynamic range. Several values of the modulation depth were used, but were kept constant within each pair of stimuli. The exact values were chosen from the set {5%, 10%, 20%, 40%, 99%} and were determined in a pilot experiment. The highest value was always 99% and for the lowest value, the subject should score around chance level ($p=50\%$). In this experiment, the number of presentations was 80 in each condition. The total number of conditions varied across subjects. For each of the two channels and each of the two frequency pairs, one to four modulation depths were selected. Each next trial consisted of a stimulus pair from any condition, randomly chosen.

The relative modulation depth m' is here defined as the difference between the minimum and maximum pulse amplitude in the SAM pulse train, divided by the dynamic range on the stimulated channel. The threshold, or the minimum level of the dynamic range, was the current level for which a 500-ms[3] unmodulated pulse train of 1250 pps was just audible. The maximum level of the dynamic range equaled the amplitude of the largest pulse of a 500-ms long SAM pulse train (1250-pps carrier rate, 100-Hz modulation frequency) at a comfortable loudness level. The level of the smallest pulse of this SAM pulse train was always kept at threshold. Table II represents the limits of the dynamic range for each subject and channel. Also, the absolute modulation depths (m) are given, following the definition in Eq. (1). The minimum and the maximum current level of the pulses are denoted by $I_{\min}$ and $I_{\max}$, respectively:

$$m=\frac{I_{\max}-I_{\min}}{I_{\max}}. \tag{1}$$

The absolute modulation depth m can be derived from the relative modulation depth m' using Eq. (2). The upper limit of the dynamic range is denoted by I_{MCL}, the threshold level is denoted by I_{THR}:

TABLE II. Lower (I_{THR}) and upper (I_{MCL}) limits of the dynamic range for each subject and channel (third and fourth column) as used in experiment I. The last five columns indicate the absolute modulation depth that corresponds to the relative modulation depth indicated at the top of each column. If the condition was not tested, no figure is given (···).

Subject	Channel	I_{THR} (μA)	I_{MCL} (μA)	Absolute modulation depths m at... $m'=5\%$	$m'=10\%$	$m'=20\%$	$m'=40\%$	$m'=99\%$
SV	2	80	500	···	8.4	16.8	33.6	83.2
	6	260	750	···	6.5	13.1	26.1	64.7
WM	2	150	570	···	7.4	14.7	29.5	72.9
	7	140	570	···	···	15.1	30.2	74.7
AE	4	350	1100	3.4	6.8	13.6	···	67.5
	8	450	1150	3.0	6.1	12.2	···	60.3
JH	2	100	470	···	7.9	15.7	31.5	77.9
	7	110	620	···	8.2	16.5	32.9	81.4

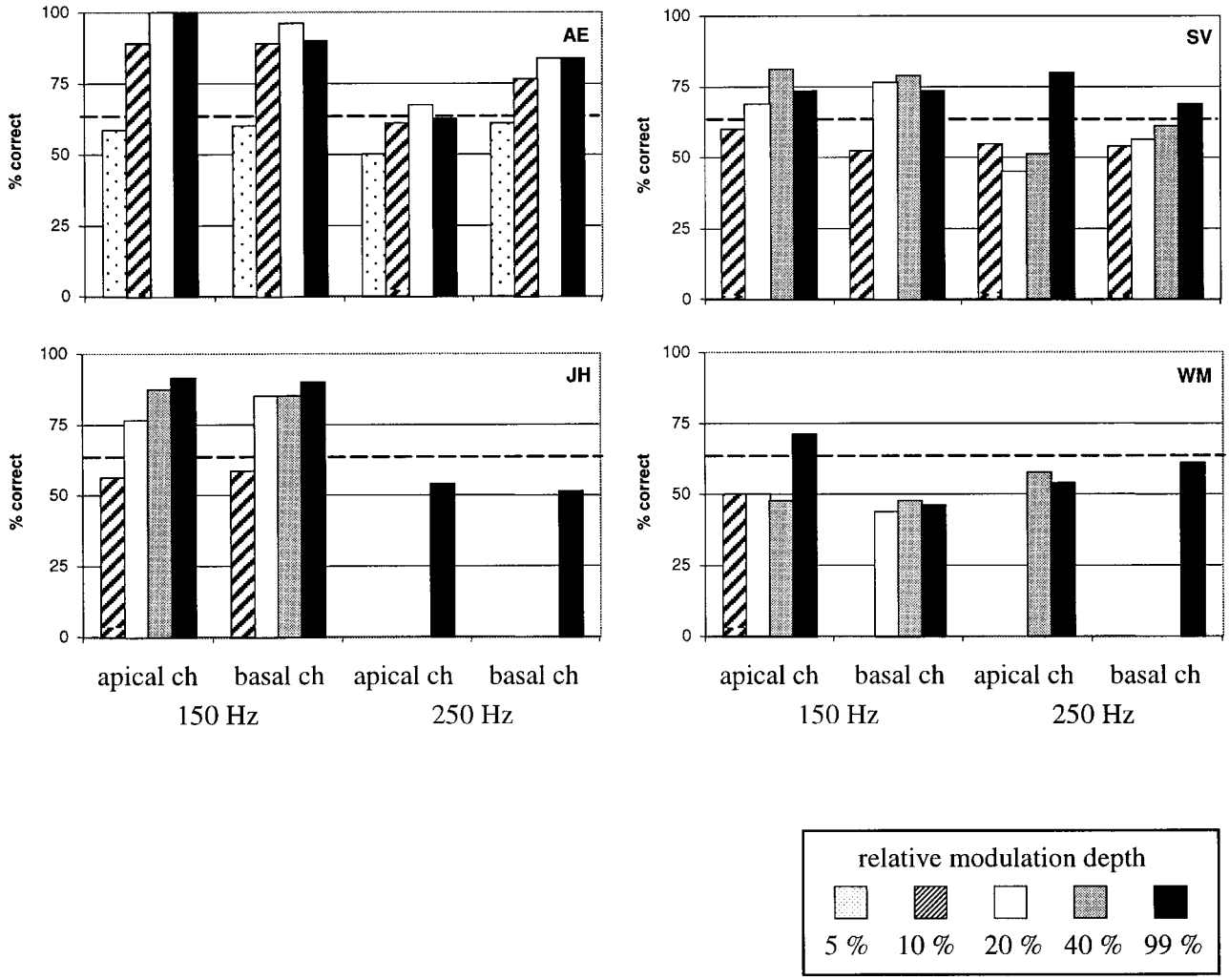


FIG. 1. Percentage correct scores for experiment I. Each plot shows the results of one subject. The channel and lowest modulation frequency of the stimulus pair is indicated at the bottom. Scores above the dashed line are significantly higher than the 50% chance level ($p<0.01$).

$$m = m' \cdot \frac{I_{\mathrm{MCL}} - I_{\mathrm{THR}}}{I_{\mathrm{MCL}}}. \tag{2}$$

Stimuli were not balanced in loudness for the following reasons. For both stimuli in a trial, duration, carrier rate, minimum and maximum current level—and thus modulation depth—were identical, and so mean current level, root mean square current level, and total injected electric charge was the same. Therefore, it is not likely that the two stimuli have a different loudness. This was confirmed by all the subjects. In the case in which loudness would be slightly different, it would be totally due to the difference in modulation frequency. The subjects are allowed to use any cue that emerges from a change of this parameter.

C. Results

Each plot in Fig. 1 shows the percentage correct answers for one subject. The conditions are indicated at the bottom of the figure. In general, it is clear that if the modulation depth is too small, the subjects performed at chance level. The minimum value of the modulation depth needed for a reliable discrimination (e.g., a score greater than 75%) depends on the condition and the subject. This value varies from 10% (e.g., subject AE, 150 Hz on both channels; 250 Hz on basal channel) to more than 40% (e.g., subject SV, 250 Hz on apical channel). Notice that for each subject, there are one or more conditions for which the criterion of 75% correct responses is not met for any modulation depth (e.g., subjects JH and WM at 250 Hz). Also, there seems to be a saturation effect in some conditions: increasing the modulation depth beyond a certain value does not noticeably increase the score (e.g., subject AE, 250 Hz; subject SV, 150 Hz). Subject AE was the best performer. Her scores were the highest on average, and she could hear pitch differences in both frequency intervals (150 Hz and 250 Hz). Subject SV's results were similar, but his scores were somewhat lower. Subject JH was clearly sensitive to modulation frequency changes in the low-frequency region (150 Hz), but not in the high-frequency region (250 Hz). His scores were still at chance level at the largest modulation depths. Subject WM was the poorest performer. He never reached the 75% criterion, and only scored significantly above chance ($p<0.01$) in one condition: 150 Hz, apical channel, 99% modulation depth.

D. Discussion

Two subjects were not sensitive to 20% deviations of the modulation frequency around 250 Hz, while the two others performed still above chance at 250 Hz, but worse compared to the 150-Hz condition. Either lower scores were obtained at 250 Hz for a given modulation depth (subject AE) or optimal performance was reached at higher modulation depths (subject SV). There are two possible explanations for this effect. First, in order to hear a pitch difference, the relative change in frequency or pulse rate must be larger at higher frequencies (Tong and Clark, 1985; Townshend *et al.*, 1987; McKay *et al.*, 1994). In this experiment, the relative change was always 20%, so this value might be below the detection limit at 250 Hz for subjects JH and WM. A second explanation is that a temporal modulation itself is more easily detected at lower frequencies (Shannon, 1992; Busby *et al.*, 1993). This might imply that also higher modulation depths are needed at higher frequencies in order to perform equally well as for lower frequencies in a pitch discrimination task.

At first sight, subjects with a larger dynamic range seem to obtain better overall scores. Indeed, the correlation between the absolute dynamic range ($I_{\mathrm{MCI}}-I_{\mathrm{THR}}$) for each channel of each subject and the average of the scores at 10%, 20%, and 40% modulation depth[4] is 0.723, which is significant ($p=0.043$). This indicates that subjects with a larger dynamic range need less relative modulation depth to judge the pitch of a temporally modulated train. Obviously, the variance in the data cannot be explained only in terms of the dynamic range, but also the subject's sensitivity to rate or frequency variations has to be taken into account.

McKay *et al.* (1995) concluded that the modulation depth of a modulated pulse train should be greater than 100 μA in order to elicit a pitch that equals the pitch of an unmodulated pulse train with the rate equal to the modulation frequency. This value corresponded to at least a quarter of the dynamic range up to almost the whole dynamic range, depending on the subject. Although one should always be cautious when comparing inter-subject results, their findings can be related to the results of experiment I. The optimal performance was obtained at modulation depths ranging from 20% of the dynamic range (subject AE) via 40% (subject SV, 150 Hz) to 99% (subject SV, 250 Hz; subjects JH and WM). So, it is possible that for these values of the modulation depth the stimulus elicited a pitch that corresponds to the modulation frequency.

From physiological measurements of the neural response to modulated electric pulse trains, Litvak *et al.* (1999) concluded that the modulation gain[5] is much higher than the one obtained in acoustical stimulation, and that the synchronization to the modulation already saturated at modulation depths of 10%. More variability, i.e., less synchronization, in the response to electrical stimuli would result in a better approximation of natural acoustical hearing. However, the best performances in experiment I were obtained for modulation depths much larger than the aforementioned 10%. In other words, desynchronization might improve the mimicking of normal hearing, but will not necessarily result in a better performance of the electric listener.

III. EXPERIMENT II

To evaluate the influence of the modulation depth on pitch judgments in a real-life situation, a second experiment was run. The smallest noticeable difference in $F0$ was measured for synthetic vowels, processed by one of the three CIS-based processing strategies. Unlike experiment I, the stimuli consisted of modulated pulse trains on all available active channels. At the output of one of the algorithms, FLAT CIS, no temporal fluctuations according to $F0$ were present. A comparison of the results between FLAT CIS and the other two will permit the determination of the importance of temporal coding of $F0$. As will be shown further, the standard CIS strategy results in modulation depths at which sub-optimal performance is obtained in some cases of experiment I. So, a new algorithm, $F0$ CIS, is designed to control and increase the modulation depth. It is expected that the subjects will perform better using this new algorithm.

A. Speech processing

The three speech processing algorithms in this study consisted of the following steps: band-pass filtering, envelope extraction, logarithmic compression, current mapping, and pulse train modulation. The processing schemes for one frequency band are shown in Fig. 2. The scheme at the top corresponds both to the default CIS and the FLAT CIS algorithm, the scheme at the bottom corresponds to the $F0$ CIS algorithm. The input of the speech processing algorithms were synthetic vowels, downsampled from 20 kHz to 10 kHz and with a resolution of 16 bits.

1. FLAT CIS and CIS

The signal is first filtered by several band-pass filters (Butterworth fourth order) covering the speech frequency range. The number of band-pass filters corresponds to the number of active electrode channels in the cochlea of the implanted subject (see Table I). The cutoff frequencies are linearly spread below 1 kHz and logarithmically above 1 kHz. For eight channels, these are 100, 325, 550, 775, 1125, 1634, 2372, 3444, and 5000 Hz, and for seven channels, these are 100, 357, 614, 934, 1421, 2161, 3287, and 5000 Hz. Next, the envelope is calculated by full-wave rectification and a smoothing filter. This filter is a Butterworth fourth-order low-pass filter with a cutoff frequency of 50 Hz for the FLAT CIS algorithm and 400 Hz for the CIS algorithm (Wilson *et al.*, 1991). The former is well below the range of $F0$ values of adult speech, the latter is clearly above that range. As a result, the $F0$ fluctuations are clearly present in the envelope of the CIS algorithm, and are absent in the envelope of the FLAT CIS algorithm. This is illustrated in Fig. 3 where both envelopes for channel 5 are shown on the same plot. The input was the synthetic vowel /a/ with $F0$ equal to 250 Hz. Note that the low cutoff frequency of 50 Hz used in the FLAT CIS algorithm is unlikely to affect speech intelligibility. In several studies, it was concluded that speech reception scores are nearly unaffected after low-pass filtering the speech envelope at 50 Hz, for normal-hearing listeners (Drullman *et al.*, 1994; Shannon *et al.*, 1995), as well as for cochlear implantees (Fu and Shannon, 2000).

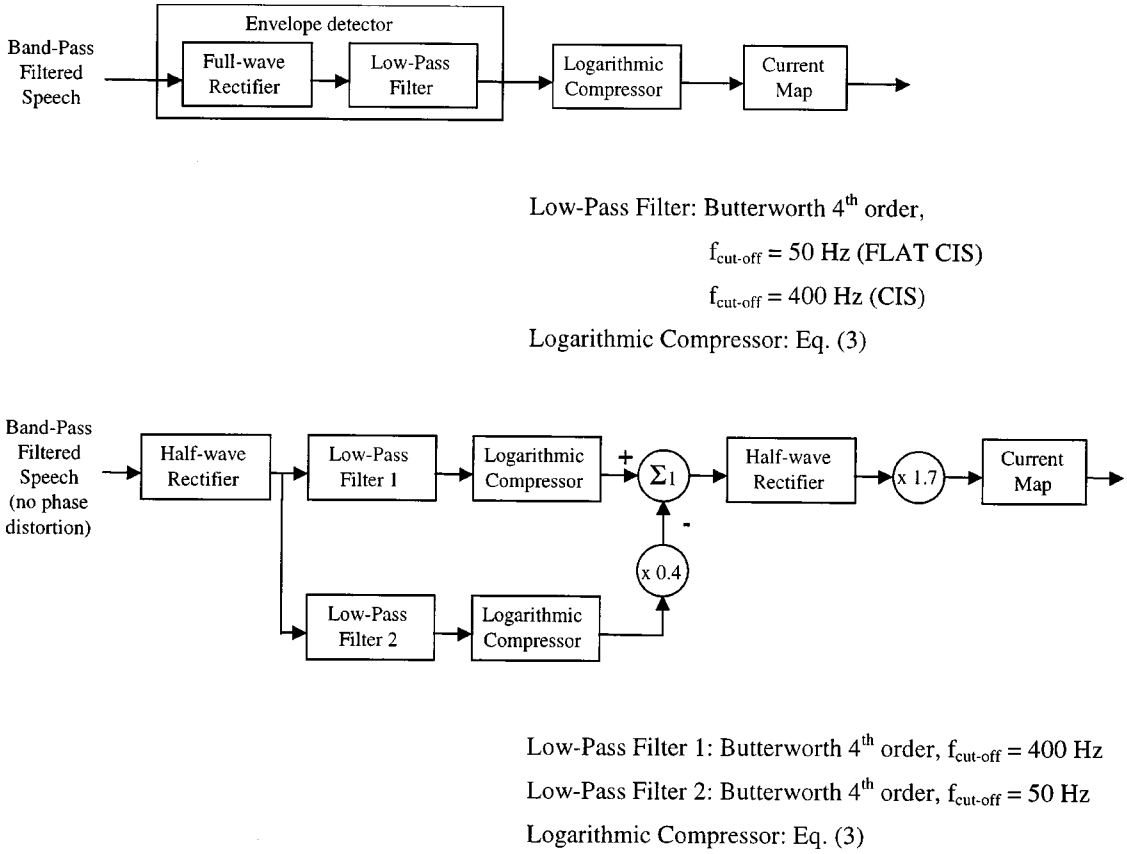


FIG. 2. Processing schemes for the three algorithms: the default CIS and FLAT CIS at the top, and $F0$ CIS at the bottom.

In the next step, the smoothed envelope is instantaneously compressed using the following formula [Eq. (3)], in which A and B are chosen such that the function does not show any discontinuity at the transition points ($x=128$ and $x=16\,384$):

$$\begin{aligned} y &= 0, \quad x \leqslant 128 \\ &= A\log(x) + B, \quad 128 < x < 16\,384 \\ &= 1000, \quad x \geqslant 16\,384. \end{aligned} \tag{3}$$

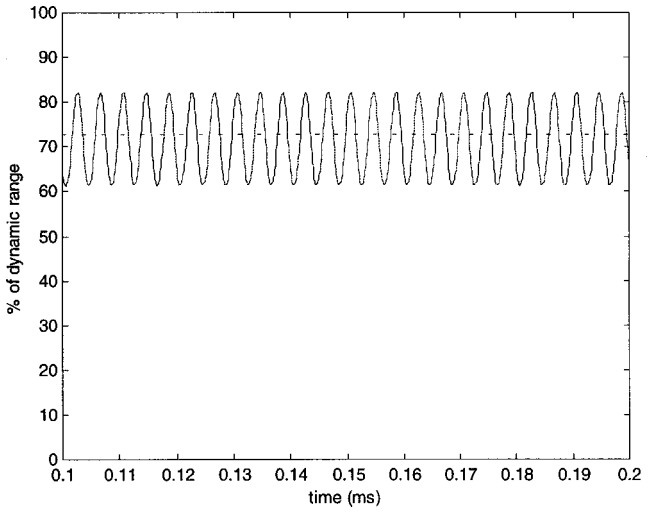


FIG. 3. Envelopes of synthetic vowel /a/ ($F0=250$ Hz) in channel 5 (frequency band between 1125 Hz and 1634 Hz) as a function of time. The solid line is the envelope of the low-pass filter of the CIS algorithm (cutoff frequency at 400 Hz). The dashed line corresponds to the envelope of the low-pass filter of the FLAT CIS algorithm (cutoff frequency at 50 Hz). In the latter case, the $F0$ fluctuations in the envelope are not present.

2. F0 CIS

The scheme of the $F0$ CIS algorithm is somewhat more complicated. There are three major differences compared with the other two algorithms: the band-pass filtering, the rectification and the calculation of the envelope.

- Instead of using one fourth-order Butterworth band-pass filter for each channel, a second order Butterworth filter is used through which the signal passes twice. First, the signal goes through the filter in the normal direction and next, the result is reversed in time, again passed through the same filter, and again reversed in time. This way, the phase distortion is zero, i.e., all frequency components are delayed with the same amount of samples.[6] If the normal procedure is followed, the maxima of the $F0$ fluctuations in adjacent channels do not coincide. In the lower two channels, this time difference can equal several milliseconds. This way, the periodicity in the response of neurons that are sensitive to stimulation on two or more channels, might be blurred. The cutoff frequencies on the other hand are identical in all algorithms.
- The filtered signal is half-wave instead of full-wave rectified. The difference is apparent mainly in the lower-

frequency channels, where individual harmonics can be resolved for high values of $F0$. Half-wave rectification reflects better the phase-locked behavior of auditory nerve fibers, and results in a larger modulation depth in the output signal.

- For the calculation of the envelope, two fourth order Butterworth low-pass filters are first employed with cutoff frequencies of 400 Hz and 50 Hz. In the output of the former filter, the $F0$ fluctuations are present in the envelope, while these are absent in the latter (see also Fig. 3). By subtracting an attenuated version of the ''flat'' envelope from the fluctuating envelope, the modulation depth is increased (summation in Fig. 2). Negative values have no meaning and are rejected (half-wave rectifier). To compensate for the resulting smaller amplitude, the signal in each channel is amplified with a factor of 1.7.

In the three algorithms, the envelope samples are mapped to current values (last block of each scheme in Fig. 2), taking the dynamic range of the channel of the subject under test into account. A linear function is used, which maps an envelope sample value of 1 to the current threshold I_{THR}, and a value of 1000 to the most comfortable current level I_{MCL}. The current envelopes in all channels are modulated with pulse trains with an overall rate of 10 000 pps. If 8 channels are used, the pulse rate per channel is 1250 pps, while this rate equals 1429 pps in the 7 channel case. The width of each biphasic pulse is constant and equal to 40 μs/phase. Following the CIS strategy, the pulse trains of all channels are interleaved. The channel order of stimulation is staggered: 1, 4, 7, 2, 5, 8, 3, and 6, where channel 1 corresponds to the most apical channel and channel 8 to the most basal channel. Stimulation on the nonactive electrode channel is skipped for the subject using only seven channels. Examples of the output of each algorithm for the synthetic vowel /a/ ($F0 = 200$ Hz) are shown in Fig. 4. The plot at the top is the output of the FLAT CIS strategy. Clearly, no $F0$ fluctuations are present in the envelopes. The plot in the middle is the output of the default CIS strategy, now showing a clear periodicity in the envelope corresponding to $F0$. Notice the frequency doubling of the $F0$ fluctuations in channel 1, due to the full-wave rectification (compare with channel 2). There also seems to be some asynchrony between channels 4,5, and 6, due to the phase distortion: the local maxima of the $F0$ fluctuations do not coincide in time. The plot at the bottom is the output of the $F0$ CIS algorithm. Clearly, the modulation depth of the $F0$ fluctuations is increased in all channels. Also, the synchronization between the channels is increased, and the periodicity of fluctuations in the output of channel 1 corresponds to $F0$.

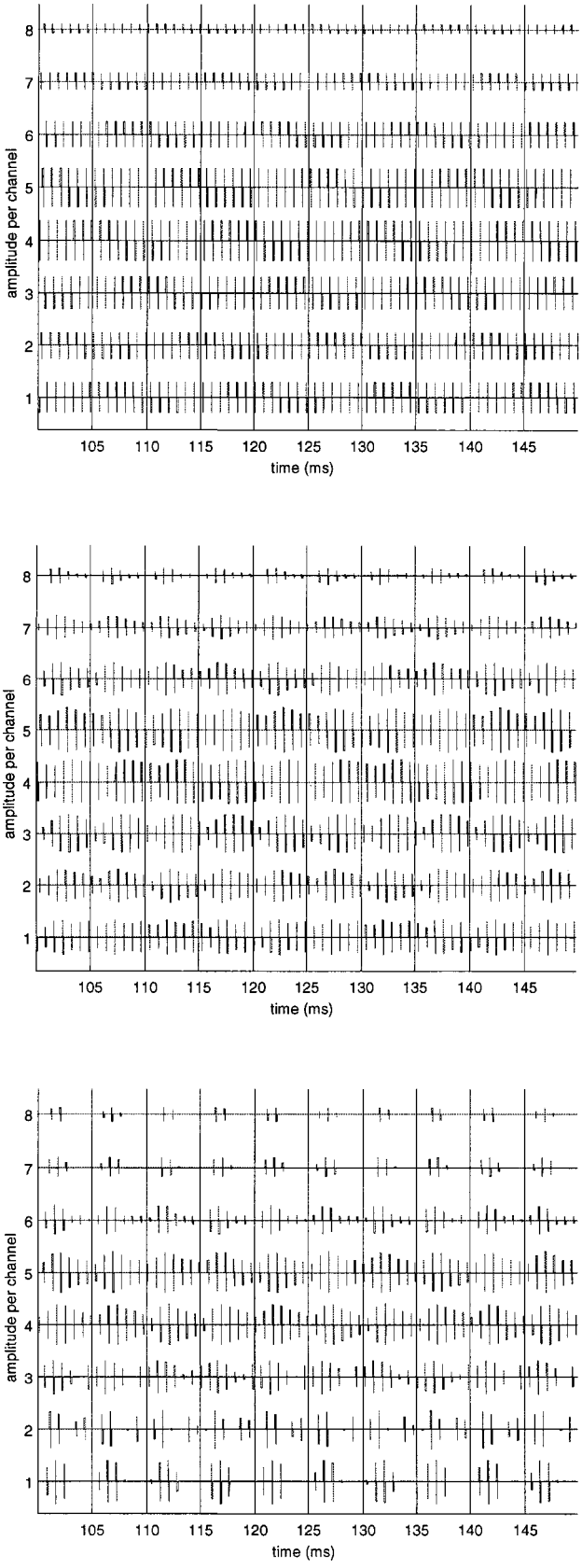


FIG. 4. The pulsatile output of the three CIS algorithms for the synthetic vowel /a/ ($F0 = 200$ Hz) as a function of time: FLAT CIS (upper plot), default CIS (middle plot), and $F0$ CIS (bottom plot). The channel numbers are indicated before each envelope trace on the left.

B. Stimuli

All stimuli were synthetic vowels, created using the Klatt synthesizer available in the CSRE software package.[7] The duration of the stimuli was 500 ms, and the highest

TABLE III. Formant frequencies (Fi) and bandwidths (Bi) in Hz of the synthetic vowels used in experiment II (Govaerts, 1978).

Vowel	$F1$	$B1$	$F2$	$B2$	$F3$	$B3$	$F4$	$B4$	$F5$	$B5$
/a/	890	110	1360	120	2475	160	3530	275	4830	325
/i/	225	40	2280	90	2820	120	3200	120	3500	150

sample rate (20 kHz) was chosen. The formant frequencies and bandwidths corresponded either to /a/ or /i/, and were kept constant for all fundamental frequencies. The exact values are typical for Flemish/Dutch spoken language and are given in Table III (Govaerts, 1978). When using the Klatt synthesizer, the value of the fundamental period has to equal an integer multiple of the sampling period. The fundamental period of the synthesized vowels ranged from 2.7 ms ($F0 = 370$ Hz) to 6.7 ms ($F0 = 149$ Hz).

After the synthesis, the stimuli were further processed using custom-made MATLAB[8] functions. The first 100 ms were truncated, since there was a large transition effect showing a dc shift at the output of the Klatt synthesizer. Second, the vowels were gated at the initial 50 ms and at the final 50 ms, using a cosine ramp. Finally, all vowels were balanced in root mean square level. Note that this does not guarantee that loudness can not be used as a cue when two vowels with different $F0$ must be discriminated. Within each channel, the mean amplitude of the pulses changes when $F0$ is varied. This is mainly due to the fact that the attenuation of the harmonics in the side bands of the band-pass filters depends on the frequency of the harmonic, and the latter changes along with $F0$. This change of average pulse amplitude in each channel can alter the perceived loudness and even the perceived timbre. This possible source of information corresponds to the aforementioned ''average amplitude cue.'' Figure 5 illustrates this ''average amplitude cue'' by showing the average pulse amplitude in each channel for two synthetic vowels (/i/) with different fundamental frequencies (250 Hz and 300 Hz). The vowel with the lower $F0$ seems to have more energy in the most apical channel, and less energy in the other channels.

To illustrate the effect of the different algorithms, the outputs of all vowels were further analyzed. For each synthesized vowel, all $F0$ fluctuations within the middle 200 ms were extracted, and both the minimum and the maximum pulse level within each fluctuation was determined. The average over all extracted fluctuations was calculated for both

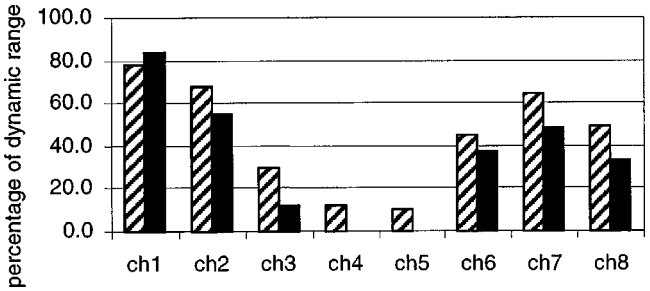


FIG. 5. Average pulse amplitude in each channel for two vowels. The dashed bars correspond to an /i/ with $F0 = 300$ Hz. The solid bars correspond to an /i/ with $F0 = 250$ Hz.

the minimum and the maximum pulse level. Figure 6 (for the /a/ and Fig. 7 /i/) show both values for the three algorithms as a function of $F0$. The data for each channel are shown in a separate panel. The vertical axis represents the relative dynamic range: ''0'' corresponds to the threshold level (I_{THR}), ''100'' to the most comfortable level (I_{MCL}). The dashed lines correspond to the FLAT CIS strategy, the thin solid lines to the default CIS strategy, and the thick solid lines to the $F0$ CIS strategy. For the FLAT CIS strategy, the mean minimum and mean maximum pulse levels almost completely overlap, illustrating that there is hardly any modulation according to $F0$ left in the signal. In the case that there is still a nonzero modulation depth, it is far below the lowest depth used in experiment I (at which scores were around chance level). The curves for the FLAT CIS strategy also illustrate the ''average amplitude cue'': the mean amplitude varies as a function of $F0$ in all channels. However, many of these curves show a clear nonmonotonic behavior, which can make this source of information very ambiguous. For the default CIS strategy (thin solid lines), there is a clear gap between both curves, which corresponds to the relative modulation depth. In the case of the vowel /a/, there is a large depth at the lowest $F0$'s (43.9% at $F0 = 149$ Hz, averaged across all channels), but it decreases rapidly for higher $F0$'s (17.5% at 250 Hz, 14.0% at 370 Hz). The same pattern is seen in the plots for the vowel /i/, but in this case the depth at the lowest $F0$'s is already smaller (31.5% at 150 Hz, 17.5% at 250 Hz, 16.2% at 370 Hz). This pattern is opposite to the pattern needed for good $F0$ discrimination; for example, the results of experiment I for subject SV clearly indicate that a higher modulation depth is needed for higher $F0$'s in order to achieve the same performance as for low $F0$'s. At the modulation depth for $F0$ equal to 250 Hz (17.5% for both vowels), he performed at chance. So, he might benefit from a new algorithm that increases the modulation depth.

The $F0$ CIS is successful in accomplishing this (thick solid lines in Figs. 6 and 7). The mean modulation depth varies from 53.8% ($F0 = 149$ Hz) to 38.6% (250 Hz) and 35.4% (370 Hz) for the vowel /a/. For the vowel /i/ these values are 48.9%, 38.4%, and 40.3%, respectively. Notice also that in many channels, the average minimum pulse is zero, i.e., at threshold level. In some cases, the minimum level shows a strong fluctuation when plotted as a function of $F0$. This effect is also visible for the default CIS algorithm, but to a lesser extent. It is inherent in the way the speech signal is analyzed in the frequency domain, namely by using band-pass filters with a flat response in a broad pass band, and steep slopes beyond the cutoff frequencies. Phase and attenuation of individual harmonics can differ substantially if, for example, the filter contains only one harmonic in the pass band close to the center frequency as opposed to the case where the filter contains two harmonics far from the center frequency. Another difference in the output between the $F0$ CIS and the default CIS algorithm is the frequency weighting. The maximum pulse levels in the two lowest frequency channels are higher for the $F0$ CIS strategy, and are higher in the remaining channels for the default strategy. To compensate this, one could apply a channel-specific gain in

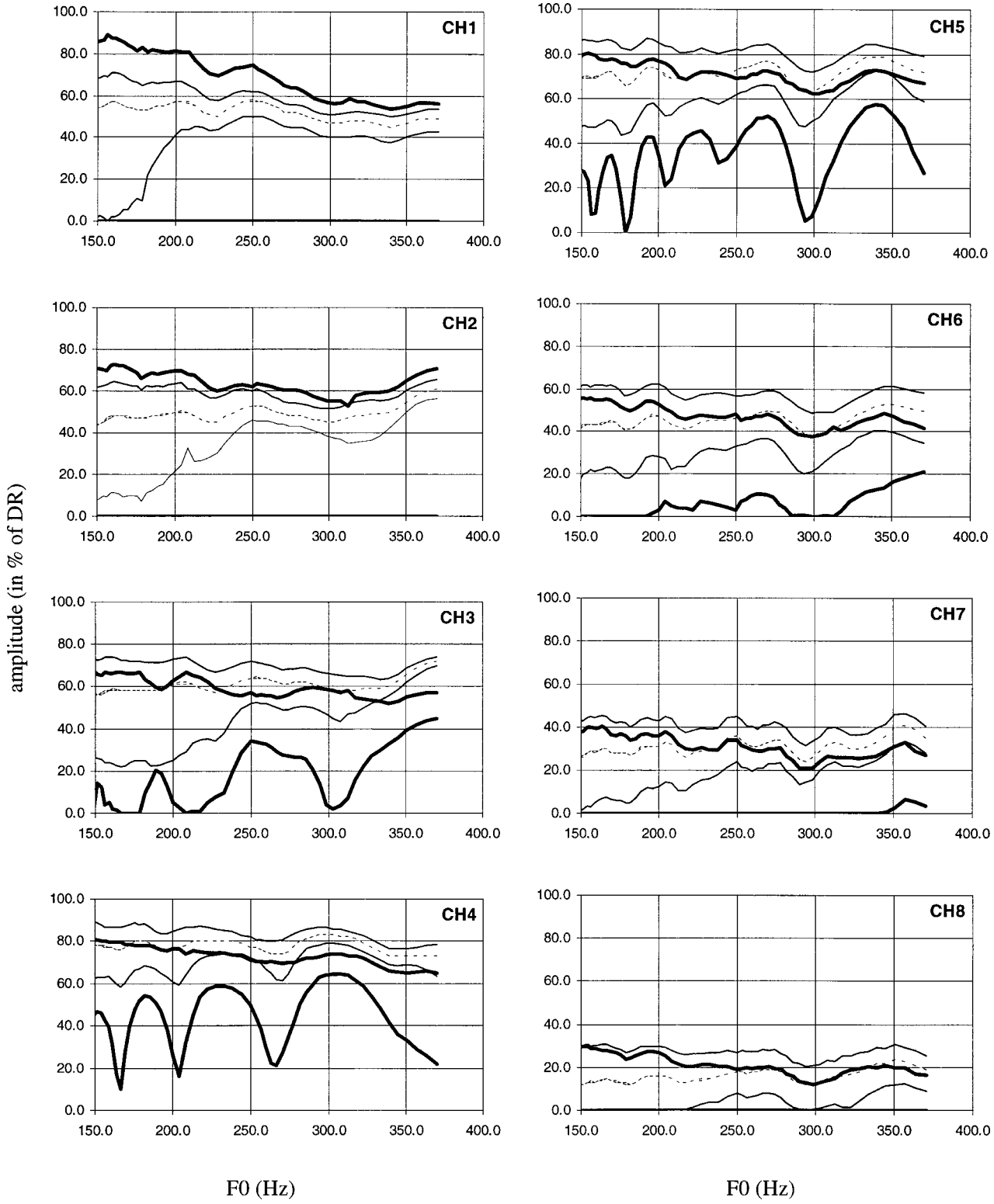


FIG. 6. Minimum and maximum pulse level within each fundamental period for the vowel /a/ for the three algorithms FLAT CIS (dashed lines), default CIS (solid thin lines) and *F*0 CIS (solid thick lines). The horizontal axis corresponds to *F*0 (in Hz); the vertical axis corresponds to the relative dynamic range (in %). Each plot shows the curves for the channel indicated at the top right.

the *F*0 CIS algorithm, instead of the factor 1.7 (see description of *F*0 CIS in Fig. 2).

C. Procedure

The same four subjects as for experiment I participated in this experiment. To estimate the smallest difference in *F*0 that could be discriminated, a 2-down 1-up adaptive procedure was used (Levitt, 1971). Two vowels with different *F*0, further called the *standard* and the *signal*, were presented and the subject had to indicate the stimulus with the higher pitch. Correct response feedback was always given. *F*0 of the standard stimulus was fixed to either 149 Hz or 250 Hz,

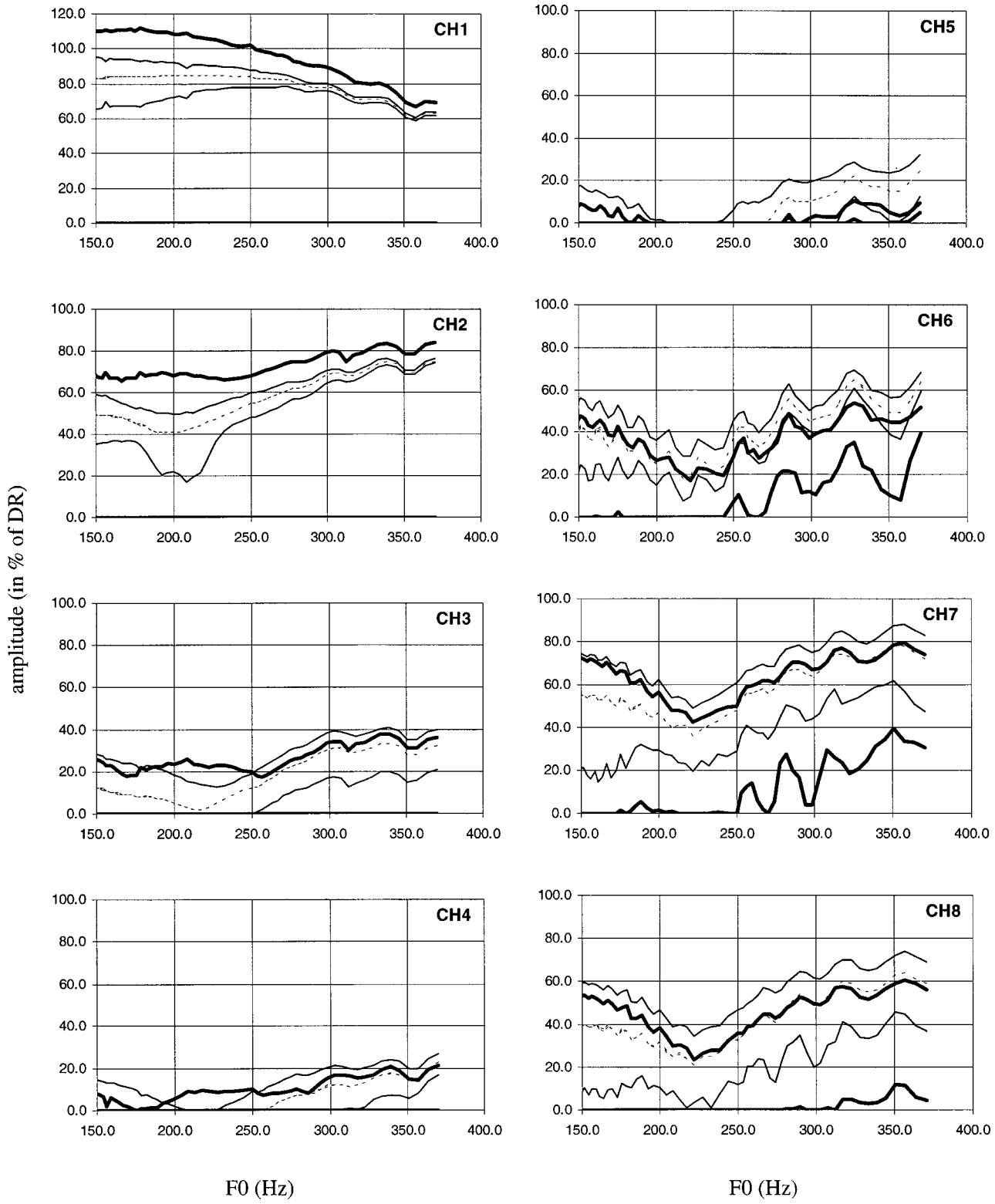


FIG. 7. Same as Fig. 6, but for the vowel /i/.

so there were 12 conditions in total: two phonemes (/a/ and /i/), two standard $F0$'s, and three algorithms. Initially, the signal had a $F0$ of 250 Hz or 370 Hz, respectively. After two successive correct answers, the $F0$ of the signal was lowered, while after one incorrect response, it was increased. Initially, the stepsize of the fundamental period was 0.4 ms (standard $F0 = 149$ Hz) or 0.2 ms (standard $F0 = 250$ Hz). After four reversals, the stepsize was reduced to 0.1 ms or 0.05 ms, respectively. The procedure stopped after 12 reversals, and the mean value of the fundamental period at the last 8 reversals was calculated. This value corresponds to the fundamental period of a signal that could be discriminated in approximately 71% of the cases (Levitt, 1971). For each condition, the whole procedure was repeated four times, and the mean and the standard deviation of the results were calculated. In some conditions, the subjects could not discriminate

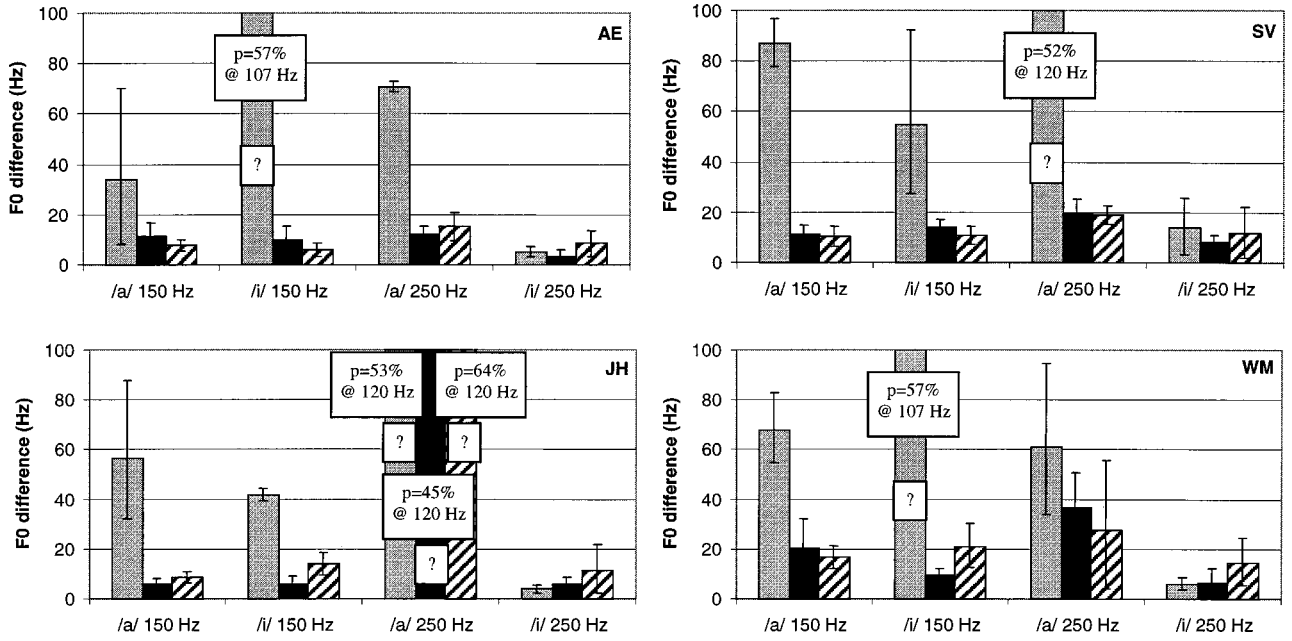


FIG. 8. Smallest discriminable $F0$ differences, as measured in experiment II. Each plot shows the results of a given subject. Vowel and $F0$ of the standard stimulus are indicated at the bottom. The three algorithms are: FLAT CIS (gray bars), default CIS (black) and $F0$ CIS (hashed). If the subject's results did not converge to some value, a large bar is displayed, and the percentage correct responses at the indicated $F0$ difference are given.

both stimuli, even at the largest difference in $F0$. In those cases, the percentage correct responses was calculated at this difference. Obviously, it is expected that these scores are lower than the aforementioned 71%.

D. Results

The results are shown in Fig. 8 for each subject separately. The horizontal axis corresponds to the conditions, the vertical indicates the $F0$ difference. In general, the largest $F0$ differences were obtained in the FLAT CIS algorithm. Still, in many cases the adaptive procedure converged to a value below the starting value. This indicates that the average amplitude cue can be used to distinguish between two vowels with different $F0$.

However, the fact that smaller $F0$ differences were obtained for the two other algorithms indicates that the inclusion of the envelope modulation cue increases the ability to make a judgment about the pitch. All subjects obtained strikingly lower $F0$ differences for those two algorithms at the lower frequency (150 Hz). The $F0$ difference that still could be heard varies between 6 Hz and 20 Hz across the subjects. Subjects SV and AE, who were sensitive to pitch changes above 250 Hz in experiment I, also benefit from the envelope modulation cue for the vowel /a/ at higher frequencies. Their results vary between 12 Hz and 19 Hz. The other two subjects do not obtain better results, since they do not have the necessary sensitivity. For all subjects, very good performance was obtained for the FLAT CIS algorithm in the ⟨/i/–250 Hz⟩ condition, so that there was no room for improvement. Underlying reasons for this unexpected result will be discussed in the next section. No significant differences are observed between the results of the default CIS and the $F0$ CIS algorithms. This indicates that there is a saturation effect, as also observed in experiment I. Further increase of the modulation depth does not improve the pitch discrimination ability.

E. Discussion

The very good performance in the ⟨/i/–250 Hz⟩ condition can be explained from Fig. 7. Above 250 Hz there is a clear quasi-monotonic relationship between the vowel's $F0$ and the average maximum pulse level in each fundamental period. This is the case for all three algorithms. Moreover, the slopes of the curves are rather steep, meaning that a small change in $F0$ results in a substantial change in pulse level in each channel. While this level increases for smaller $F0$'s in the lowest frequency channel, it is the other way around for the other channels. This can result in a pitch change according to tonotopic place coding, but also in a timbre or even loudness change. The latter was reported by subject JH, while the other three subjects still reported that they heard a change in pitch. For the lower frequencies (150–250 Hz), the slopes of the curves in Fig. 7 become very shallow or even change sign. This inconsistency probably explains why performance is a lot poorer for the ⟨/i/–150 Hz⟩ condition, when there are no envelope modulation cues (FLAT CIS). The same holds for the vowel /a/. The pulse level changes nonmonotonically with $F0$, which makes it hard to use only this source of information to identify the stimulus with the highest $F0$.

Using the algorithms that included a source of temporal information (default CIS and $F0$ CIS), the four subjects could hear $F0$ deviations of between 4% and 13% from 150 Hz. Two subjects were not sensitive to pitch changes above 250 Hz in experiment I, but the other two could hear $F0$ differences between 5% and 8% at 250 Hz in experiment II. These data are comparable with measurements of the rate

pitch difference limens (DL) obtained in other studies. McKay *et al.* (1994) found DLs between 2% and 15% at 150 Hz, and between 2% and 20% around 250 Hz, while one of the six subjects was not sensitive to rate changes from 250 Hz. Townshend *et al.* (1987) obtained rate pitch DLs between 10% and 30% at 150 Hz, and between 10% and 20% at 250 Hz, while one of the three subjects was not sensitive to rate changes from 250 Hz. It might also be interesting to compare the observed DLs for harmonic complexes with those obtained psychoacoustically with normal-hearing listeners. Houtsma and Smurzynski (1990) measured the DLs of complex tones ($F0$ equal to 200 Hz) from which the lower harmonics were removed. Averaged across four listeners, they found a DL of less then 1 Hz (0.5%), even when the first six harmonics were removed. Cullen and Long (1986) observed a similar value for unfiltered acoustic pulse trains with a rate of 200 Hz. In other words, normal-hearing listeners outperform implantees by an order of magnitude. The question of whether this is due to a better temporal or a better place coding (or the combination of both) still remains unresolved. However, the latter reason might be more favorable, given the facts that place coding is rather coarse in a cochlear implant and that the temporal representation of amplitude modulated stimuli is well preserved in the neural response for high carrier rates (Wilson, 1997).

In this experiment, synthetic vowels were used in order to be able to control all the parameters of the vowel. Formant frequencies and bandwidth were kept constant, and only $F0$ was varied. This is unlike real speech. Speakers with higher $F0$'s, such as women and children, typically have higher formant frequencies too, due to the smaller physical dimensions of the vocal tract. It is not unlikely that these possible sources of information would also be used if a comparable task as in this experiment was performed with real life speech.

The results of both experiments are consistent in most cases. In experiment I, subjects SV and AE showed a clear sensitivity to modulation frequency variations, both for the low- and the high-$F0$ frequencies. In experiment II, they obtained better results when the envelope modulation cue was added (FLAT CIS versus other two), except for the condition ⟨/i/−250 Hz⟩ for which there was no room for improvement. Subjects JH and WM could not detect modulation frequency variations from 250 Hz in experiment I, so no smaller $F0$ differences were obtained in experiment II, when the envelope modulation cue was added. Obviously, a cochlear implant user will not benefit from a speech processing algorithm that adds or enhances a source of information for which s/he is not sensitive.

However, there appear to be some inconsistencies. From the results of experiment I one would predict that some subjects should benefit from the $F0$ CIS algorithm in some conditions. For example, subject SV performed at chance for a modulation depth of 20% at the modulation frequency of 250 Hz. In the default CIS strategy, the modulation depth varies over all channels between 12% and 24% for the synthetic vowel /a/ ($F0=250$ Hz) used in this study. Also subject WM performed rather poorly in experiment I, and performed only significantly above chance in one condition, where modulation depth was 99%. However, no improvement is seen in their results of experiment II when $F0$ CIS is compared with the default CIS strategy. There are two possible reasons for this finding. First, not all channels were tested in experiment I, so it is still possible that there are other channels for which less modulation depth is needed in order to achieve maximum performance. However, for subject WM, two other channels were also tried in a pilot experiment, and no better results were obtained for those channels. Second, and more likely, the main difference between both experiments is that in experiment I only one channel is stimulated, while in experiment II all channels are stimulated. It is possible that less modulation depth is needed when the modulation according to $F0$ is presented on all channels. This can also explain the relatively good performance of subject WM in experiment II. At the low frequencies ($F0$ around 150 Hz), he clearly benefits from the inclusion of $F0$ fluctuations in the envelope, and can detect $F0$ differences of 10–20 Hz.

IV. EXPERIMENT III

An extra experiment was run to test the hypothesis that, if several channels are stimulated with synchronous fluctuations, the modulation depth needed to discriminate two stimuli with different envelope fluctuation rate is lower compared to when only one channel is stimulated. This also would imply that better performance is achieved for multiple-channel stimuli if the same modulation depth is used for both the single- and the multiple-channel stimuli.

A. Stimuli and procedure

The same four subjects participated in this experiment. For each subject, three adjacent channels were selected, on which SAM pulse trains with a carrier pulse rate of 1250 pps were presented. The central channel of the three (ch X) was one of the two channels used in experiment I: the basal channel for subject JH, and the apical channel for the other subjects. In addition, the adjacent channel towards the base (ch $X+1$) and the adjacent channel towards the apex (ch $X-1$) were selected. The SAM pulse trains were either presented on each channel separately, or concurrently on the three channels. In the latter case, the pulses on the three channels were interleaved. A modulation depth of 10% was chosen for subjects AE and JH, and a modulation depth of 40% for subjects SV and WM. The procedure was similar to experiment I. In each trial, two stimuli with different modulation frequencies, 150 Hz and 180 Hz, were presented in random order and the subjects had to indicate the stimulus with the higher pitch. The number of presentations was 100 in each condition. Each next trial consisted of a stimulus pair on any of the three channels, or on all three channels.

B. Results

The results of all subjects are shown in Fig. 9. The solid black bars represent the percentage correct scores for the three-channel stimuli, while the other bars represent the scores for the single-channel stimuli. It is clear that the best performance is achieved for the three-channel stimuli. The

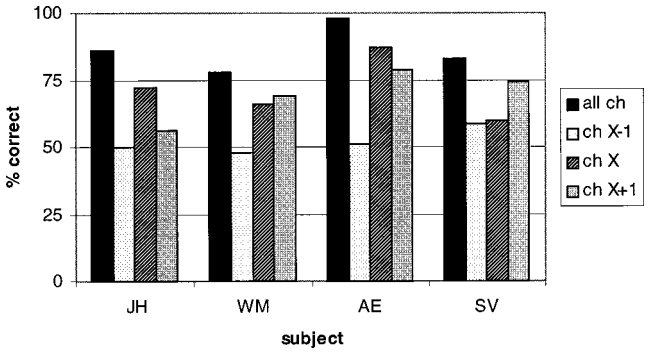


FIG. 9. Percentage correct scores for experiment III, grouped per subject (indicated at the bottom). The results for the three-channel stimuli are represented by black solid bars, while the other bars represent the scores for the single-channel stimuli. Channel *X* was different for each subject (see text).

average difference between the three-channel result and the result for the ''best'' channel is 11%. As shown with a paired-samples *t*-test, this difference is highly significant ($t[3]=9.1$, $p=0.003$).

C. Discussion

The results confirm that the subjects can combine the information on the three channels in some way, such that better performance can be achieved although the modulation depth is identical. An improvement of 11% is obtained, when comparing the results of the discrimination task for the three-channel stimuli with those for the single-channel stimuli. In the CIS strategies of experiment II, 7 or 8 channels were concurrently stimulated, which might enhance the performance even more.

The results of the experiments demonstrate the importance of the temporal coding of $F0$ for the perception of its corresponding pitch. As long as $F0$ appears as distinct temporal fluctuations in the envelope of the speech processor's output, tones with different $F0$ can be distinguished. In the SPEAK strategy, used with the Nucleus cochlear implant, the bandwidths of the extracted envelopes are limited to 200 Hz, which can fall below $F0$. However, the steepness of the smoothing filter is far from infinite, which means that enough $F0$ modulation might remain in the envelope. A greater concern is the use of the relatively low stimulation pulse rate of 300 pps or even lower. According to several physiological and psychophysical studies that recommended a carrier rate four times the modulation rate (McKay *et al.*, 1994; Wilson, 1997), only modulation frequencies up to 75 Hz can be encoded properly in the SPEAK strategy. However, in the aforementioned studies, single-channel stimuli were used, in contrast with the multiple-channel stimuli at the SPEAK processor output. Because less modulation depth is required when multiple channels are stimulated, it might also be true that a lower carrier rate suffices if all channels contain synchronous envelope modulations. This might explain why Jones *et al.* (1995) did not see any improvement in pitch related tasks if an explicit temporal code according to $F0$ was added to the SPEAK strategy.

V. SUMMARY AND CONCLUSIONS

Three experiments were conducted in this study. The first investigated how the modulation depth of a SAM pulse train on one electrode channel influences the ability to judge perceived pitch. Optimal performance in the pitch discrimination task depended on the subject, the channel (apical or basal) and the frequency range (150 Hz–180 Hz or 250 Hz–300 Hz). The relative modulation depths for which the highest scores were obtained varied between 20% and 99%. In the second experiment, the smallest discriminable $F0$ difference was measured for three speech processing algorithms using synthetic vowels. In the first algorithm (FLAT CIS), the temporal variation in the envelope according to $F0$ was removed, so subjects had to make a judgment on the basis of relative amplitude changes. They all performed very well in one condition (⟨/i/–250 Hz⟩), due to the unambiguous relation between $F0$ and the average amplitude in each channel. For the other conditions (⟨/i/–150 Hz⟩, ⟨/a/–150 Hz⟩, and ⟨/a/–250 Hz⟩), performance was much poorer, and in some cases at chance level. If the temporal source of information was added in the envelopes (default CIS and $F0$ CIS), significantly better results were obtained, provided that the subject was sensitive to modulation frequency changes in the corresponding frequency region. In the default CIS algorithm, modulation depths were obtained at which some subjects scored sub-optimally in experiment I. Therefore, the $F0$ CIS algorithm was designed in which the modulation depths were increased. However, this did not result in better performance. This indicates that less modulation depth in the envelope variation according to $F0$ is needed if all channels are stimulated. This hypothesis was confirmed in a third experiment where subjects performed better in a pitch discrimination task with SAM pulse trains, if three channels were stimulated concurrently, as opposed to only one.

ACKNOWLEDGMENTS

The authors wish to thank all participating subjects, who were implanted by Dr. E. Offeciers at the University ENT-Department of the A.Z. St. Augustinus Hospital in Antwerp, for taking part in the experiments. This study was supported by the Fund for Scientific Research—Flanders (Belgium).

[1]LAURA is a registered trade mark of Philips Hearing Implants, now Cochlear Technology Centre Europe.

[2]In contrast with this study, McKay *et al.* did not use SAM pulse trains. Only two pulse levels were used, with the highest one occurring only once in each modulation cycle.

[3]A length of 500 ms is used, while the test signals are 400 ms long. However, both time intervals are much larger than the temporal integration window for loudness estimation, so this should not influence the outcomes.

[4]For subject AE, the scores at $m'=40\%$ were not measured. The missing values were estimated by using the average scores at 20% and 99%. The possible error on this estimate will be small since the scores at both modulation depths do not differ substantially.

[5]The modulation gain is the ratio between the modulation depth in the response and the modulation depth in the stimulus.

[6]The advantage of this approach is that the amplitude characteristics of these band-pass filters are identical to those used in the other two algorithms. In order to achieve a zero-phase distortion in a real-time implementation, a FIR (Finite Impulse Response) filter should be used.

[7]Computerised Speech Research Environment, Version 4.5, AVAAZ Innovations, Inc. (http://www.avaaz.com).

[8]MATLAB is a registered trademark of The Mathworks, Inc. (http://www.mathworks.com).

Blamey, P. J., Dowell, R. C., Tong, Y. C., and Clark, G. M. (**1984**). ''An acoustic model of a multiple-channel cochlear implant,'' J. Acoust. Soc. Am. **76**, 97–103.

Blamey, P. J., Dowell, R. C., Clark, G. M., and Seligman, P. M. (**1987**). ''Acoustic parameters measured by a formant-estimating speech processor for a multiple-channel cochlear implant,'' J. Acoust. Soc. Am. **82**, 38–47.

Busby, P. A., Tong, Y. C., and Clark, G. M. (**1993**). ''The perception of temporal modulations by cochlear implant patients,'' J. Acoust. Soc. Am. **94**, 124–131.

Cariani, P. A., and Delgutte, B. (**1996**). ''Neural correlates of the pitch of complex tones. I. Pitch and pitch salience,'' J. Neurophysiol. **76**, 1698–1716.

Cullen, Jr., J. K., and Long, G. R. (**1986**). ''Rate discrimination of high-pass-filtered pulse trains,'' J. Acoust. Soc. Am. **79**, 114–119.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). ''Effect of temporal envelope smearing on speech perception,'' J. Acoust. Soc. Am. **95**, 1053–1064.

Eddington, D. K., Dobelle, W. H., Brackman, D. E., Mladejovsky, M. G., and Parkin, J. L. (**1978**). ''Auditory prostheses research with multiple channel intracochlear stimulation in man,'' Ann. Otol. Rhinol. Laryngol. **87**, Suppl. 53, 1–59.

Fu, Q.-J., and Shannon, R. V. (**2000**). ''Effect of stimulation rate on phoneme recognition by Nucleus-22 cochlear implant listeners,'' J. Acoust. Soc. Am. **107**, 589–597.

Govaerts, G. (**1978**). ''Perceptuele structuren van synthetische en natuurlijke klinkers'' [''Perceptual structures of synthetic and natural vowels''], Psychologica Belgica **18-1**, 27–67.

Houtsma, A. J. M., and Smurzynski, J. (**1990**). ''Pitch identification and discrimination for complex tones with many harmonics,'' J. Acoust. Soc. Am. **87**, 304–310.

Jones, P. A., McDermott, H. J., Seligman, P. M., and Millar, J. B. (**1995**). ''Coding of voice-source information in the Nucleus cochlear implant system,'' Ann. Otol. Rhinol. Laryngol. **104**, 363–365.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Litvak, L., Delgutte, B., Eddington, D. E., and Cariani, P. A. (**1999**). ''Auditory-nerve fiber responses to electric stimulation: Modulated and unmodulated pulse trains,'' in *Abstracts of the 1999 Conference on Implantable Auditory Prostheses, Asilomar, U.S.A.*, p. 62.

McDermott, H. J., McKay, C. M., and Vandali, A. E. (**1992**). ''A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant,'' J. Acoust. Soc. Am. **91**, 3367–3371.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). ''Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees,'' J. Acoust. Soc. Am. **96**, 2664–2673.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1995**). ''Pitch matching of amplitude-modulated current pulse trains by cochlear implantees: The effect of modulation depth,'' J. Acoust. Soc. Am. **97**, 1777–1784.

Miller, M. I., and Sachs, M. B. (**1984**). ''Representation of voice pitch in discharge patterns of auditory-nerve fibers,'' Hear. Res. **14**, 257–279.

Palmer, A. R. (**1990**). ''The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers,'' J. Acoust. Soc. Am. **88**, 1412–1426.

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). ''Pitch of the residue,'' J. Acoust. Soc. Am. **34**, 1418–1424.

Shannon, R. V. (**1992**). ''Temporal modulation transfer function in patients with cochlear implants,'' J. Acoust. Soc. Am. **91**, 2156–2164.

Shannon, R. V. (**1983**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.

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

Simmons, F., White, R., Walker, M., and Mathews, R. (**1981**). ''Pitch correlates of direct auditory nerve stimulations,'' Ann. Otol. Rhinol. Laryngol. **90**, 15–18.

Skinner, M. W., Holden, L. K., Holden, T. A., Dowell, R. C., Seligman, P. M., Brimacombe, J. A., and Beiter, A. L. (**1991**). ''Performance of postlingually deaf adults with the Wearable Speech Processor (WSP III) and mini speech processor (MSP) of the Nucleus multielectrode cochlear implant,'' Ear Hear. **12**, 3–22.

Tong, Y. C., and Clark, G. M. (**1985**). ''Absolute identification of electric pulse rates and electrode positions by cochlear implant patients,'' J. Acoust. Soc. Am. **77**, 1881–1888.

Townshend, B., Cotter, N., Van Compernolle, D., and White, R. L. (**1987**). ''Pitch perception by cochlear implant subjects,'' J. Acoust. Soc. Am. **82**, 106–115.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''Better speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.

Wilson, B. S. (**1997**). ''The future of cochlear implants,'' Br. J. Audiol. **31**, 205–225.

A comparison of threshold estimation methods in children 6–11 years of age

Emily Buss,[a] Joseph W. Hall, John H. Grose, and Madhu B. Dev
University of North Carolina School of Medicine, Department of Surgery, Division of Otolaryngology, Chapel Hill, North Carolina 27599

(Received 10 May 1999; revised 6 July 2000; accepted 3 November 2000)

Estimating detection threshold for auditory stimuli in children can be problematic because of lapses in attention and the time limits usually imposed by scheduling restrictions or fatigue. Data reported here were collected to compare the stability of threshold estimation procedures in testing children ages 6 to 11 in a three-alternative, forced-choice paradigm. Stimuli consisted of a 1-kHz tonal signal and a Gaussian noise masker, bandpass filtered between 500–2000 Hz and presented at 25-dB spectrum level. The signal was either presented for 400 ms in the presence of a continuous masker (simultaneous masking) or for 10 ms just prior to a 400-ms masker (backward masking). For each masking paradigm the 79% correct threshold was assessed via each of three procedures: 3-down, 1-up adaptive staircase (Levitt), maximum likelihood estimation (MLE), and method of constant stimuli. Percent correct was measured at the end of the study for a signal 10 dB above the previously determined threshold in order to estimate the most appropriate psychometric function asymptote for fitting data collected via the method of constant stimuli. Both the MLE and Levitt procedures produced equally stable threshold estimates for both conditions and age groups. This was the case despite considerable variability in backward-masking thresholds. [DOI: 10.1121/1.1337960]

PACS numbers: 43.66.Dc, 43.66.Sr, 43.66.Yw [SPB]

I. INTRODUCTION

Many methods for estimating detection threshold have been developed with the assumption that detection is determined by an underlying distribution that is relatively stable across trials and that sequential responses are statistically independent. These assumptions might be most problematic in cases where practice over the course of trials or fluctuation in attention are suspected to influence responses. Some studies have looked at the validity of these assumptions in infants (e.g., Bernstein and Gravel, 1990) and in preschool children (e.g., Wightman and Allen, 1992). Less is known about the appropriateness of standard psychophysical methods for young school-age children. Many studies on this age group have used methods developed for and tested on adults, such as the Levitt staircase (Levitt, 1971) or the maximum likelihood estimator (MLE; Green, 1993).

Previous research comparing Levitt and MLE tracking procedures with adult observers indicates a good correlation between detection thresholds for a tone in noise estimated using these two methods (Gu and Green, 1994). There are at least three inter-related nonsensory factors that might predict different results for young children, however. First is the duration of a track. The MLE procedure can require significantly fewer trials to obtain an estimate than a Levitt track, particularly compared to a typical 3-down, 1-up Levitt track (Shelton, Picardi and Green, 1982; Baker and Rosen, 1998), and this could lessen the impact of fatigue or boredom in a young listener. Second, inattention during the course of a track can have different consequences for threshold estimates produced by the MLE and Levitt procedures. Misses at the beginning of the MLE track have a much larger impact on threshold estimates than those occurring later in the track (Green, 1995). If children are more prone to problems in focusing attention or to confusion during the initial trials, then the MLE procedure could produce less accurate (higher) thresholds than the Levitt procedure. A third factor is the difference in exposure to the signal over the course of a typical track. A typical MLE track hones in on a likely threshold estimate fairly quickly, providing fairly few signal presentations at suprathreshold levels, while the typical Levitt track, on the other hand, provides more exposure to suprathreshold signals. Performance on tasks based on subtle or complex cues could benefit from the more extensive exposure to suprathreshold signals, and this effect could be greater for younger or less-experienced listeners.

The purpose of the study undertaken here was to compare Levitt and MLE tracking procedures for a three-alternative, forced-choice method in young school-aged children. Data were also collected using a method of constant stimuli and using a suprathreshold signal. Two masking paradigms were tested, simultaneous and backward masking. Simultaneous masking is thought to be based on a very simple cue. Backward masking, on the other hand, is associated with large individual differences and practice effects (Wilson and Carhart, 1970; Oxenham and Moore, 1995; Buss *et al.*, 1999), characteristics that could be interpreted as reflecting a complex or subtle detection cue. Based on this, it was hypothesized that backward masking might be particularly sensitive to differences in the amount and type of exposure to the signal that is provided during the course of the threshold estimation track. This would be reflected in lower

[a]Electronic mail: ebuss@med.unc.edu

thresholds measured using the Levitt procedure than the MLE in backward masking.

II. METHODS

A. Observers

A total of 34 observers took part in the study, 23 children (6.3 to 10.7, with a mean of 8.3 years) and 13 adults (21 to 49.5, with a mean of 34.6 years). All had normal hearing, defined as thresholds equal to or better than 15 dB HL at octave frequencies from 250 Hz to 8000 Hz (ANSI, 1989), and none of the observers in either age group had any known history of ear disease. Twelve children (5 male, 7 female) and 7 adults (2 male, 5 female) participated in the simultaneous-masking conditions. Fifteen children (9 male, 6 female) and 7 adults (1 male, 6 female) participated in the backward-masking conditions. The four children and three adults who participated in both conditions completed the simultaneous-masking conditions approximately 1 year prior to completing the backward-masking conditions.

B. Stimuli

In both the simultaneous- and backward-masking conditions, the signal was a 1-kHz pure tone ramped on and off with 10-ms $\cos^2$ ramps. The masker was Gaussian noise, filtered into a two-octave band centered on 1 kHz, and presented at a 25-dB spectrum level. In the simultaneous-masking conditions, the masker was played continuously and the signal was presented for 400 ms. In the backward-masking conditions, the masker was a 400-ms sample of the noise, ramped on and off with 10-ms $\cos^2$ ramps, and the signal was ramped on and off (no steady state) prior to the onset of the masker, with no delay between 0-voltage points of the signal and masker. The masking noise was generated continuously by a waveform generator (WG1, TDT), filtered digitally (PD1, TDT; 25-μs sampling rate), attenuated (PA4, TDT), and routed to an adder (SM3, TDT). In the backward-masking condition, the masker was passed through a switch (SW2, TDT) prior to attenuation, allowing the masker to be switched on and off. The signal, including onset and offset ramps, was generated by a second waveform generator (WG1, TDT), attenuated (PA4, TDT), and added to the masker (SM3, TDT). The combined signal and masker channels were played out over the left channel of the headphone buffer (HB6, TDT) to headphones (Sony, MDR-V6).

C. Procedures

Three procedures were used to estimate the 79% correct detection threshold. All employed a three-alternative, forced-choice presentation, with the signal appearing in one randomly selected interval. The interstimulus interval was approximately 400 ms for the simultaneous- and 1 s for the backward-masking conditions. Listening intervals and feedback were indicated via lights on a response box. All tracks began with a practice trial with a signal level 10 dB above expected detection threshold. If the observer missed this trial, then the experimenter paused the experiment, talked with the observer about the task, and then presented another practice trial. This continued until the observer responded correctly, at which point data collection began.

Each observer was randomly assigned to begin with either the Levitt or MLE tracking procedure. Three threshold estimates were obtained with that procedure, and a fourth was collected if the first three varied by more than 3 dB.[1] The observer was then tested using the second tracking procedure, with three or four estimates comprising the mean. Data for the method of constant stimuli were then collected, followed by suprathreshold detection. The thresholds reported here were calculated as the average of all threshold estimates obtained with each procedure, except where noted.

Levitt (1971) adaptive tracks followed a 3-down, 1-up stepping rule, estimating 79% correct. The initial step size was 8 dB, which was reduced to 4 and 2 dB after the first and second track reversals, respectively. Trials continued until a total of eight reversals was obtained, and the threshold estimate was calculated as the average signal level at the last six track reversals. In the MLE adaptive track the signal level presented on each trial was based on the 79% correct point of the psychometric function providing the best fit to the data collected in the track up to that point (see Green, 1995 for details). The psychometric function slope constant was set to $\omega = 0.4$ based on pilot data from a single adult observer. It has been shown that threshold estimates are relatively insensitive to mismatches in slope of the fitted psychometric function (Green, 1993). Asymptotic performance was set to 100% correct, with the assumption that inattention would be negligible. This decision could introduce a substantial bias in threshold estimates if the real underlying function reaches asymptote at a level well below this level of attentiveness (Green, 1995), so the consequences of this choice were assessed by *post hoc* fitting functions with asymptote estimated in the suprathreshold task.

The next five tracks employed a method of constant stimuli, resulting in estimates of percent correct. Five signal levels were chosen for each observer, distributed relative to that observer's estimated threshold (see Lam *et al.*, 1997). For the simultaneous-masking conditions the signal levels were $+2$, $+1$, $+0$, -1, and -2 dB relative to each observer's Levitt threshold for simultaneous masking. For the backward-masking conditions the signal levels were $+8$, $+4$, $+0$, -4, and -8 dB relative to each observer's Levitt threshold for backward masking. Each block was comprised of 30 trials, with signal levels interleaved.

The final experimental run was comprised of 30 trials, with the signal presented 10 dB above the threshold from the Levitt procedure for each observer. Again, percent correct was determined. In several cases, performance on this task was quite poor when compared to the previous data on that observer. In those cases, data were taken a second time, as discussed below.

III. RESULTS AND DISCUSSION

A. Levitt procedure

For the simultaneous-masking condition, average thresholds were 46.4 dB SPL (45.0 to 51.1) for children and 45.4 dB SPL (43.8 to 47.7) for adults. One track was omitted

from all calculations because of an unusually large standard deviation of the track reversal levels (9.8 dB) as compared to other child and adult data, which had average standard deviations of 2.3 and 2.2 dB, respectively. For the backward-masking conditions, average thresholds were 58.1 dB SPL (37.5 to 74.8) for children, and 46.2 dB SPL (35.5 to 53.2) for adults. Standard deviations of the track reversal levels were, on average, 3.3 and 4.2 dB for child and adult observers, respectively.

B. MLE procedure

For the simultaneous-masking condition, average thresholds were 46.8 dB SPL (44.6 to 49.9) for children and 45.6 dB SPL (43.8 to 48.6) for adults. The predictive value of the final psychometric function fit in the MLE track was assessed with the G^2 statistic, calculated as described in Allen *et al.* (1998). A criterion of $p<0.05$ was taken to indicate an acceptably good fit of the psychometric function to the data. One track from the child data and two from the adult data failed to reach this criterion and were omitted from all analyses.[2]

For the backward-masking conditions, average thresholds were 56.6 dB SPL (44.8 to 74.7) for children, and 47.0 dB SPL (33.8 to 54.4) for adults. All backward-masking tracks obtained an acceptably good fit, as indicated by a G^2 significance level of $p<0.05$.

C. Suprathreshold detection

For the simultaneous-masking conditions, most observers from both groups got all 30 trials correct for the detection of a suprathreshold signal, leading to an estimate of 100% accuracy at asymptote. Exceptions were two child observers who missed 1 and 2 trials, respectively, and an adult observer who missed 16 out of 30 trials. This observer admitted to a loss of concentration during the course of the run. On an immediate retest, 100% of the signals were correctly detected.

For the backward-masking conditions, more than half of the observers got all 30 trials correct for the detection of a suprathreshold signal, but there were notable exceptions. Two child observers initially performed very poorly on these trials, missing 7 and 17 out of 30 trials, respectively. Both of these observers were retested. The first child then performed at 100%, but the other failed to complete the track at all because he had fallen asleep. Based on this, it was decided to bring this observer back for a second test session on another day, and to then replace all the data.

Taken together, these data are interpreted as indicating that for both masking conditions children and adults have comparable difficulties maintaining attention in the final segment of a listening session. These results raise a question, however, about the utility of this measure because it highlights the extent to which attention to the task can fluctuate over the course of the experiment. Because of concern over this, psychometric functions (below) were fitted to the data with 100% asymptote as well as the asymptote estimated in the suprathreshold task.

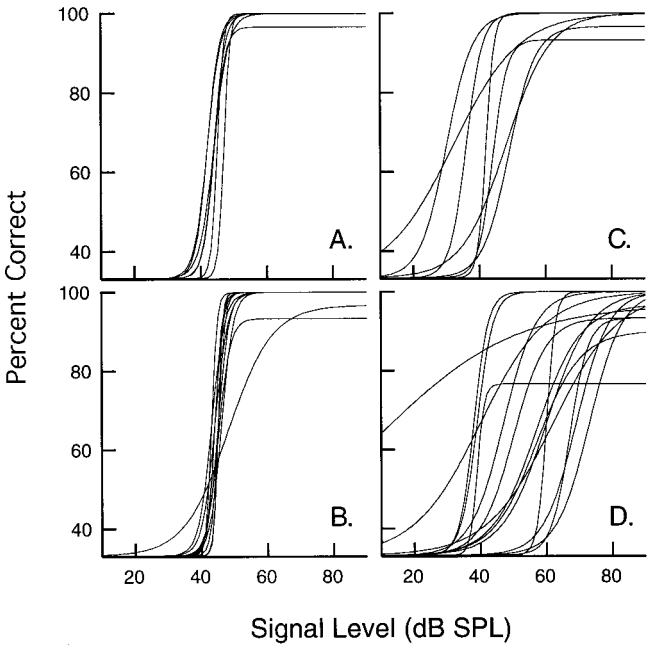


FIG. 1. Psychometric functions fitted to data obtained by the method of constant stimuli are plotted separately for subject groups and masking conditions. Functions fitted to simultaneous-masking data from adults are shown in panel A, and those from children in panel B. Functions fitted to backward-masking data from adults are shown in panel C, and those from children in panel D.

D. Method of constant stimuli

The data collected with the method of constant stimuli were fitted with psychometric functions using a maximum likelihood method with slope as a free parameter. The upper asymptote fitted to each observer's data was set as the estimate of asymptote based on the suprathreshold detection task. The function associated with the smallest sum-of-squares deviation from the data was chosen for each observer, and the 79% correct point of that function was taken as a threshold estimate.

For the simultaneous-masking conditions, average thresholds were 46.7 dB SPL (44.1 to 56.6) for children and 45.5 dB SPL (43.6 to 47.7) for adults. All fits were significant, as indicated by a G^2 significance level of $p<0.05$, and accounted for an average of 75% of the variance in the data. Slopes fitted to child and adult data were comparable, with the mean slope across subjects being $\omega=0.77$ and $\omega=0.71$, respectively. The left panels in Fig. 1 show all of the best-fitting functions, plotted separately for adult (panel A) and child (panel B) data.

For the backward-masking conditions, average thresholds were 56.9 dB SPL (39.6 to 76.9) for children and 44.1 dB SPL (32.6 to 53.8) for adults. All fits were significant, as indicated by a G^2 significance level of $p<0.05$, and accounted for an average of 78% of the variance in the data. Slopes fitted to child and adult data were comparable, with the mean slope across subjects being $\omega=0.28$ and $\omega=0.33$, respectively. The right panels of Fig. 1 show the best-fitting functions, plotted separately for adult (panel C) and child (panel D) data.

Threshold estimates for the data collected with the

method of constant stimuli employed individualized estimates of inattention and best-fitting values of slope (ω), whereas the MLE procedure assumed no effect of inattention (100% asymptote) and $\omega = 0.4$ for fitting all observers' data. The possible consequence of these discrepancies were evaluated in two ways. First, data from the method of constant stimuli were fitted again with the parameters used in the MLE procedure. The resulting threshold estimates did not vary substantially from those based on the original fits. Threshold changed by an average of 0.2 dB for the simultaneous-masking conditions and by 1.5 dB for the backward-masking condition. Second, the data collected with the MLE tracking method were refitted using the individualized parameters for slope and attention. Three of these fits failed to obtain a G^2 significance level of $p < 0.05$ and so were omitted when calculating the estimate of threshold. The resulting threshold estimates differed from the original MLE estimates by an average of 0.3 dB in the simultaneous-masking and 0.8 dB in the backward-masking conditions.

E. Stability as a function of number of trials

The issue of the number of trials necessary to obtain a stable threshold estimate with the Levitt and MLE procedures was evaluated by calculating threshold estimates based on a subset of data collected during a track. Because the MLE procedure sets the signal level for each trial as the 79% threshold estimate based on all the previous data, the average threshold estimate at any trial number can be determined by averaging the signal level across tracks. Levitt track estimates were calculated based on two, four, and six reversals. Levitt tracks based on two and four reversals were surprisingly stable as compared to those based on six reversals for both masking conditions. For the simultaneous-masking condition threshold estimates calculated at two and four reversals differed from the case of six reversals by 0.4 and 0.2 dB for child data, and by 0.6 and 0.3 dB for adult data. The two-reversal estimates could have been obtained after an average of 21 trials. Had the MLE tracks been stopped after 21 trials, thresholds would have deviated from the final estimate based on 30 trials by an average magnitude of 0.8 and 0.6 dB for children and adults, respectively. For the backward-masking conditions threshold estimates calculated at two and four reversals differed from the case of six reversals by 1.2 and 0.6 dB for child data, and by 0.7 and 0.6 dB for adult data. The two-reversal estimates could have been obtained after an average of 24 trials. Had the MLE tracks been stopped after 24 trials, thresholds would have deviated from the final estimate based on 30 trials by, on average, 1.9 and 0.9 dB for children and adults, respectively. These results suggest comparably stable threshold estimates for a small, fixed number of trials using these two tracking procedures.

F. Stability of estimates across conditions

Figure 2 shows threshold estimates based on the two tracking procedures (Levitt and MLE) plotted as a function of the threshold estimate obtained by fitting data obtained with the method of constant stimuli. Each MLE estimate (down-pointing triangles) was based on 30 trials. The Levitt estimate (up-pointing triangles) was computed as the average level at the last four reversals in each track, which were obtained after an average of 30 trials for the simultaneous-masking and 34 trials for the backward-masking conditions. Data for children (unfilled symbols) and data for adult (filled symbols) observers are plotted separately. The panel on the left (A) contains results from the simultaneous-masking conditions, and the panel on the right (B) contains results from the backward-masking conditions. The estimate based on the method of constant stimuli was chosen as a reference because those data were obtained after completion of the two tracking procedures, and so observers had many trials of previous exposure to the stimuli. In effect this was not a crucial decision because all thresholds were highly correlated. In the simultaneous-masking condition, correlation between pairs of estimates obtained for the three procedures ranged from $r = 0.82$ to 0.93. In the backward-masking condition, correlation between pairs of estimates obtained for the three procedures ranged from $r = 0.70$ to 0.99. Two analyses of variances (ANOVAs) were performed, one for each masking condition, with MLE and Levitt estimates and age group. For the simultaneous-masking condition there was no main effect of procedure ($F_{1,17} = 0.96$, $p = 0.3404$), no main effect of age group ($F_{1,17} = 2.11$, $p = 0.1644$), and no interaction ($F_{1,17} = 0.696$, $p = 0.4157$). For the backward-masking condition there was no effect of procedure ($F_{1,20} = 0.001$, $p = 0.991$), but there was a main effect of age ($F_{1,20} = 6.633$, $p = 0.0181$). The interaction was not significant ($F_{1,20} = 0.565$, $p = 0.4609$).

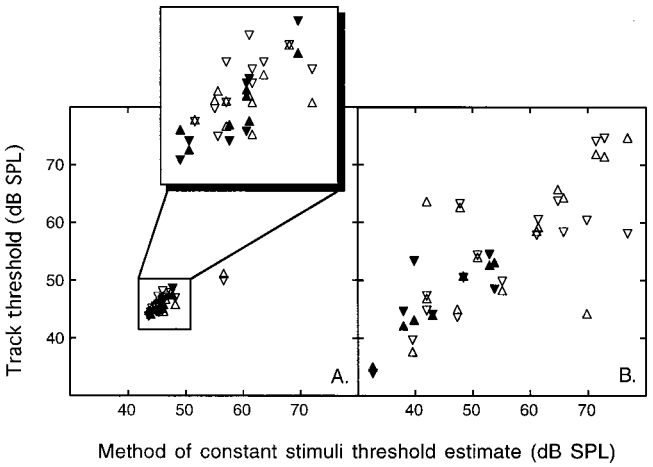


FIG. 2. Thresholds from the two tracking procedures are plotted as a function of the threshold estimate obtained by fitting data collected with the method of constant stimuli. One Levitt (△) and MLE (▽) estimate is plotted for each observer. Adult data are shown as filled symbols and child data as unfilled symbols. Panel A (left) shows simultaneous-masking threshold estimates, and panel B (right) shows backward-masking estimates. The insert in panel A does not include data for one observer. Data from the method of constant stimuli for this observer were not well-fitted by any of the functions.

IV. CONCLUSIONS

The intent of this study was to assess the relative stability of threshold estimates obtained with the MLE and Levitt tracking procedures. It was hypothesized that the small number of suprathreshold signal presentations in the typical MLE

track could introduce inaccuracies under some circumstances. In particular, it was hypothesized that child observers might benefit from more extensive exposure to suprathreshold examples of the signal, such as those available during the typical Levitt track, particularly when the cue was a subtle one that has been shown to produce practice effects. The results described above do not support this hypothesis.

Threshold estimates in the simultaneous-masking conditions were within 1–2 dB for the two groups, regardless of threshold estimation procedure. The lack of an age effect confirms previous findings that children of this age range perform on par with adults under conditions of simultaneous masking (e.g., Allen *et al.*, 1989; Hall and Grose, 1991). Thresholds obtained in the backward-masking paradigm were more variable than simultaneous-masking thresholds and differed by an average of 12.5 dB between child and adult groups. This difference is consistent with previous reports of age effects in backward masking (Buss *et al.*, 1999). Despite this variability, both Levitt and MLE tracking procedures produced comparable estimates of threshold when compared to the threshold estimates obtained with data from the method of constant stimuli. For both the simultaneous-masking and backward-masking conditions the Levitt estimates based on the average of four reversals were comparable to the MLE estimates, both in stability and in the number of trials to obtain threshold.

These results suggest that small numbers of track reversals using a Levitt procedure can produce stable results even for difficult tasks. The MLE paradigm may not be more efficient or accurate in standard listening tasks, such as simultaneous masking, but neither does it introduce inaccuracies for more complex listening tasks, such as backward masking.

ACKNOWLEDGMENTS

Helpful comments on this report were provided by Sid Bacon, Fred Wightman, and one anonymous reviewer. This work was supported by a grant from NIH, No. R01 DC00397-15.

[1]For one adult and two child observers only two threshold estimates were obtained, and a third was collected only if the range of first two estimates was more than 3 dB.

[2]This statistic was designed for assessing function fits with more uniformly spaced signal levels than were present in the MLE data. Despite this, the tracks identified in this way were unusual. In all three tracks that were singled out by this criterion there were several correct responses early in the track in response to stimuli well below the eventual threshold estimate, introducing a substantial bias towards low signal levels.

Allen, P., Jones, R., and Slaney, P. (**1998**). ''The role of level, spectral, and temporal cues in children's detection of masked signals,'' J. Acoust. Soc. Am. **104**, 2997–3005.

Allen, P., Wightman, F., Kistler, D., and Dolan, T. (**1989**). ''Frequency resolution in children,'' J. Speech Hear. Res. **32**, 317–322.

ANSI (**1989**). (S3.6-1989), ''Specification for Audiometers'' (American National Standards Institute, New York).

Baker, R. J., and Rosen, S. (**1998**). ''Minimizing boredom by maximizing likelihood—efficient estimation of masked thresholds,'' Br. J. Audiol. **32**, 104–105.

Bernstein, R. S., and Gravel, J. S. (**1990**). ''Method for determining hearing sensitivity in infants; The interweaving staircase procedure,'' J. Am. Acad. Audiol. **1**, 18–145.

Buss, E., Hall III, J. W., Grose, J. H., and Dev, M. B. (**1999**). ''Development of adult-like performance in backward, simultaneous and forward masking,'' J. Speech Lang. Hear. Res. **42**, 844–849.

Green, D. M. (**1993**). ''A maximum-likelihood method for estimating thresholds in a yes–no task,'' J. Acoust. Soc. Am. **93**, 2096–2105.

Green, D. M. (**1995**). ''Maximum-likelihood procedures and the inattentive observer,'' J. Acoust. Soc. Am. **97**, 3749–3760.

Gu, X., and Green, D. M. (**1994**). ''Further studies of a maximum-likelihood yes–no procedure,'' J. Acoust. Soc. Am. **96**, 93–101.

Hall III, J. W., and Grose, J. H. (**1991**). ''Notched-noise measures of frequency selectivity in adults and children using fixed-masker-level and fixed-signal-level presentation,'' J. Speech Hear. Res. **34**, 651–660.

Lam, C. F., Dubno, J. R., Ahlstrom, J. B., He, N. J., and Mills, J. H. (**1997**). ''Estimating parameters for psychometric functions using the four-point sampling method,'' J. Acoust. Soc. Am. **102**, 3697–3703.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Oxenham, A. J., and Moore, B. C. J. (**1995**). ''Additivity of masking in normally hearing and hearing-impaired subjects,'' J. Acoust. Soc. Am. **98**, 1921–1934.

Shelton, B. R., Picardi, M. C., and Green, D. M. (**1982**). ''Comparison of there adaptive psychophysical procedures,'' J. Acoust. Soc. Am. **71**, 1527–1533.

Wightman, F., and Allen, P. (**1992**). ''Individual differences in auditory capability among preschool children,'' in *Developmental Psychoacoustics*, edited by L. A. Werner and E. W. Rubel (American Psychological Association, Washington, DC), pp. 113–363.

Wilson, R. H., and Carhart, R. (**1970**). ''Forward and backward masking: Interactions and additivity,'' J. Acoust. Soc. Am. **4**, 1254–1263.

Forward masking: Adaptation or integration?

Andrew J. Oxenham[a)]
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 1 October 1999; revised 1 October 2000; accepted 30 October 2000)

The aim of this study was to attempt to distinguish between neural adaptation and persistence (or temporal integration) as possible explanations of forward masking. Thresholds were measured for a sinusoidal signal as a function of signal duration for conditions where the delay between the masker offset and the signal offset (the offset–offset interval) was fixed. The masker was a 200-ms broadband noise, presented at a spectrum level of 40 dB (*re*: 20 μPa), and the signal was a 4-kHz sinusoid, gated with 2-ms ramps. The offset–offset interval was fixed at various durations between 4 and 102 ms and signal thresholds were measured for a range of signal durations at each interval. A substantial decrease in thresholds was observed with increasing duration for signal durations up to about 20 ms. At short offset–offset intervals, the amount of temporal integration exceeded that normally found in quiet. The results were simulated using models of temporal integration (the temporal-window model) and adaptation. For both models, the inclusion of a peripheral nonlinearity, similar to that observed physiologically in studies of the basilar membrane, was essential in producing a good fit to the data. Both models were about equally successful in accounting for the present data. However, the temporal-window model provided a somewhat better account of similar data from a simultaneous-masking experiment, using the same parameters. This suggests that the linear, time-invariant properties of the temporal-window approach are appropriate for modeling forward masking. Overall the results confirm that forward masking can be described in terms of peripheral nonlinearity followed by linear temporal integration at higher levels in the auditory system. However, the difference in predictions between the adaptation and integration models is relatively small, meaning that influence of adaptation cannot be ruled out. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1336501]

PACS numbers: 43.66.Dc, 43.66.Mk, 43.66.Ba [RVS]

I. INTRODUCTION

Forward masking, where the threshold of a signal is elevated by a masker preceding it in time, has been the subject of intense study over a number of decades (e.g., de Maré, 1940; Lüscher and Zwislocki, 1947; Munson and Gardner, 1950; Zwislocki *et al.*, 1959; Plomp, 1964; Elliott, 1971; Widin and Viemeister, 1979; Jesteadt *et al.*, 1982; Moore and Glasberg, 1983; Nelson, 1991; Plack and Oxenham, 1998). Despite the continued interest in the phenomenon, and the many empirical facts known about it, the mechanisms underlying forward masking remain poorly understood and a matter of debate.

In situations where the masker is a brief impulse, Duifhuis (1973) has shown that peripheral frequency selectivity can play a role in producing forward (and backward) masking, because of the finite response times of the auditory filters. In such situations, the responses to the masker impulse and a brief signal may physically overlap in the auditory periphery, even if the acoustic stimuli do not, producing what is essentially simultaneous masking. However, subsequent work has indicated that for maskers longer than a single impulse, the role of peripheral interaction in forward masking is generally negligible, at least for signal frequencies of 1 kHz and above (Vogten, 1978; Gorga *et al.*, 1980; Carlyon, 1988).

Adaptation in the auditory nerve has been proposed as a candidate for the neural site of forward masking (Smith, 1977, 1979). Certainly, a number of aspects of auditory-nerve adaptation resemble psychophysical forward masking. For instance, the growth of adaptation with increasing stimulus level is nonlinear and eventually saturates, just as the growth of forward masking is generally nonlinear, with signal threshold often increasing only slowly as a function of masker level (Jesteadt *et al.*, 1982; Moore and Glasberg, 1983). Such resemblances have led many psychophysicists to refer to forward masking in terms of neural adaptation (Duifhuis, 1973; Kidd and Feth, 1981; Jesteadt *et al.*, 1982; Bacon, 1996; Nelson and Swain, 1996). However, quantitative studies of forward masking in the auditory nerve, using detection theoretic analysis techniques, have demonstrated much less masking in individual auditory-nerve fibers than is measured psychophysically (Relkin and Turner, 1988), as well as different dependencies on parameters such as signal duration and rise time (Turner *et al.*, 1994). Also, the fact that cochlear implant patients show forward masking over a similar time scale as normal-hearing listeners (Shannon, 1990) suggests that forward masking occurs at a higher stage of processing than the inner hair cells. No physiological studies using signal-detection methods have been carried out at centers higher than the auditory nerve. Thus it is possible that neural adaptation at higher levels of the auditory pathways mediates forward masking.

An alternative view of forward masking is that it is due to a continuation, or persistence, of neural activity, after the

a)Electronic mail: oxenham@mit.edu

physical offset of the masker (Plomp, 1964; Penner, 1975; Zwicker, 1984; Moore *et al.*, 1988; Oxenham and Moore, 1994). The site of such masking is hypothesized to be higher than the auditory nerve, but no specific mechanisms have been proposed, other than that it could reflect stimulus integration in neurons with relatively long time constants.

Quantitative models of forward masking have generally been in the persistence category, with the exception of the model of Dau *et al.* (1996a,b), which uses feedback loops to produce an effect similar to adaptation, although the output of the model also shows persistence, due to a low-pass filter with a cutoff frequency of 8 Hz. In the model of Moore *et al.* (1988), after peripheral filtering, stimuli are squared (giving a quantity proportional to intensity) and then passed through a linear temporal integrator, which smooths the temporal representation of the stimuli such that the response to the masker overlaps with, and hence can mask, the response to the signal. The disadvantage of using a linear temporal integrator with a square-law nonlinearity, such as that proposed by Moore *et al.* (1988), is that many nonlinear aspects of forward masking, such as its growth with masker level, can only be predicted by assuming that the integrator changes shape with level (Plack and Moore, 1990).

More recently, it has been suggested that the nonlinear aspects of forward masking may be due to mechanical nonlinearities observed in the response of the basilar membrane (BM) to sound, and that once those nonlinearities are taken into account, forward masking can be treated as linear (Oxenham and Moore, 1995, 1997). This account relies on the finding that the response of the BM to tones at characteristic frequency (CF) is linear at very low levels, but becomes highly compressive at higher levels, with the transition occurring at sound levels of somewhere between 20 and 50 dB SPL (Rhode, 1971; Sellick *et al.*, 1982; Ruggero *et al.*, 1997). At very short masker-signal intervals, when the masker and signal are in the same level region, growth of masking should be nearly linear. As the delay between the masker and signal is increased, the growth of masking should become increasingly nonlinear as the masker level falls into the compressive region of the BM response, while the signal remains in the low-level, linear region. Further data and model predictions by Plack and Oxenham (1998) support the idea that the nonlinear growth of forward masking is due to BM nonlinearity rather than saturating neural adaptation.

Based on studies to date, a model with a BM-like compressive nonlinearity, followed by a linear sliding temporal integrator, or temporal window, provides a quantitative framework for describing the nonlinear growth of masking for different masker-signal delays (Plack and Oxenham, 1998); the effect of masker duration (Oxenham and Moore, 1994); and the effect of cochlear hearing impairment (Oxenham and Moore, 1997) in forward masking. However, as noted by Plack and Oxenham (1998), the fact that BM nonlinearity can account for many of the nonlinearities observed in forward masking does not rule out the possibility that forward masking is due to neural adaptation following BM nonlinearity. It could be, for instance, that the responsible adaptation behaves in a quasi-linear way, such that the amount of adaptation is approximately proportional to the excitation produced by the masker.

The aim of this study was to distinguish between two possible mechanisms, adaptation and temporal integration, which have both been hypothesized to underlie forward masking. In this attempt it is assumed that the integration mechanism is linear and time-invariant, as assumed in most models (Moore *et al.*, 1988; Oxenham and Moore, 1994). Adaptation, on the other hand, is by definition time-variant; that is, the response to a given stimulus depends on any previous stimulation.

In an adaptation-based explanation, it seems likely that the portion of the signal furthest from the masker is most important in determining threshold, and that portions of the signal closer in time to the masker should contribute less to detection. This is because the response to portions closer to the masker will be more adapted, and hence less detectable, than later portions of the signal. It follows, therefore, that thresholds in forward masking should be less dependent on signal duration than in simultaneous masking or in quiet, as long as the time interval between the forward masker offset and signal offset is kept constant.

The temporal window is hypothesized to be a peaked function with a shorter time constant for times after the peak than before, in order to account for the fact that backward masking decays much more rapidly than does forward masking. Because of this, the greatest signal output from the window is achieved when the peak of the window is near the offset of the signal. It is generally assumed that threshold is determined by the point in time at which the signal-to-masker ratio at the output of the temporal window is greatest, and that threshold corresponds to a fixed signal-to-masker ratio. Because an increase in signal duration results in a greater signal area under the temporal window at this point in time, the model predicts that signal thresholds will decrease as signal duration is increased, at least up to durations of 20 ms or so (Oxenham *et al.*, 1997). This is true whether or not a forward masker is present; because the window is linear, the presence of the masker has no effect on the response to the signal. Thus the temporal-window model predicts at least as much temporal integration in forward masking as in quiet, or in simultaneous masking. The prediction that adaptation should lead to less dependence on signal duration in forward masking, while an integration-based account should lead to at least as much integration as is observed in quiet, forms the basis of this study.

Zwislocki *et al.* (1959) found no effect of signal duration if the interval between masker and signal offsets was held constant, thus supporting the adaptation hypothesis. However, they used smoothly gated 1-kHz tones for both masker and signal. It is possible that the signal was perceived simply as a continuation of the masker at short gaps between the masker offset and the signal onset, thus confounding the effect of signal duration (Neff, 1986). This view is supported by another experiment in that study. Initially, Zwislocki *et al.* (1959) found a nonmonotonic relationship between signal threshold and *masker* duration. The relationship reverted to being monotonic when the masker was gated abruptly, providing a salient cue for the masker offset. Unfortunately,

the experiment varying *signal* duration was not repeated with an abruptly gated masker. Elliott (1962) used a broadband noise forward masker and compared thresholds for 5- and 10-ms sinusoidal signals (gated with 1-ms ramps) at 2 kHz. She concluded that her results were consistent with Zwislocki *et al.*'s hypothesis that thresholds did not depend on signal duration. However, as only two short signal durations were tested, it is difficult to draw strong conclusions from her results.

More recent studies have suggested that signal duration may affect thresholds in forward masking, even when the offset–offset interval is held constant. Thornton (1972), using a 1170-Hz forward masker and a 1753-Hz signal, found that signal thresholds decreased with increasing duration for durations between 10 and 20 ms in roughly the same way as they did in quiet. However, the effect is small (less than 5 dB) and in the only figure of untransformed data (his Fig. 3), there appears to be a nonmonotonic relationship, with thresholds *increasing* again as the signal duration is increased from 30 to 250 ms. Fastl, using himself as the only observer, measured integration for a signal with a fixed offset–offset interval of 20 ms for a broadband masker (Fastl, 1976a), 10 ms for a critical-band masker (Fastl, 1976b), and 10 ms for a pure-tone masker (Fastl, 1979). In all these studies (Fig. 7 in all three papers), Fastl found a decrease in threshold with increasing signal duration, except in the case of the 4-kHz pure-tone masker when the signal was at the same frequency as the masker, which he also ascribed to a ''confusion'' effect. For the two noise maskers, the decrease in thresholds with increasing duration was similar to that observed in quiet. Finally, Gralla (1992) performed a similar study also using a 4-kHz signal and offset–offset times ranging from 1.5 to 100 ms. In broad agreement with Fastl's data, he found that thresholds with a broadband forward masker decreased with increasing signal duration for the first 10 ms, and then remained constant. Results with a tonal and critical-band masker showed nearly no effect, perhaps also due to confusion.

In summary, there seems to be disagreement in the literature regarding the effects of signal duration in forward masking, which makes it difficult to support either adaptation or integration theories. Also, none of the studies discussed above used a forced-choice measurement procedure, using instead variations of a Bekesy tracking method. The purpose of the present experiment was to study the effects of signal duration in forward masking using an adaptative forced-choice procedure with the aim of helping to distinguish between a theory of adaptation or a theory of linear temporal integration to account for forward masking.

II. TEMPORAL INTEGRATION IN FORWARD MASKING

A. Stimuli

A broadband Gaussian noise (0–7000 Hz) with a half-amplitude duration of 200 ms, gated with 2-ms raised-cosine ramps, and a spectrum level of 40 dB (*re*: 20 μPa) was used as a forward masker. The signal was a 4-kHz sinusoid, gated with 2-ms ramps. A broadband noise was used to avoid possible confusion effects. A relatively high signal frequency was chosen to allow the use of very brief signals (4-ms total duration) without the signal bandwidth exceeding the estimated bandwidth of the auditory filter at 4 kHz (Glasberg and Moore, 1990). Also, at such high frequencies, it is unlikely that peripheral interaction due to filter ringing limits performance. The masker-signal offset–offset interval was set to 4, 6, 9, 12, 22, 52, or 102 ms. At each of these intervals, thresholds were measured for the same range of signal durations, provided the masker and signal did not overlap in time. For example, for an offset–offset interval of 22 ms, thresholds were measured for total signal durations of 4, 6, 9, 12, and 22 ms; for an offset–offset interval of 4 ms, thresholds were only measured for a signal duration of 4 ms. When discussing the data, the signals are referred to by their half-amplitude durations, which were 2, 4, 7, 10, 20, 50, and 100 ms.

The masking noise was created at the beginning of each run by generating a 2-s circular buffer of Gaussian noise, performing a discrete Fourier transform, setting the components above 7 kHz to 0, and applying an inverse Fourier transform. A random starting point within the resulting noise buffer was selected for each presentation. Stimuli were generated digitally at a sampling rate of 32 kHz and were played out via the built-in 16-bit DAC and reconstruction filter of a Silicon Graphics workstation. Stimuli were then passed through a programmable attenuator (TDT PA4) and a headphone buffer (TDT HB6) before being presented to the left earpiece of a Beyer DT 990 headset.

B. Procedure

Thresholds were measured using a three-interval forced-choice method with a two-down one-up adaptive procedure that tracks the 70.7%-correct point of the psychometric function. Each interval contained the noise masker, and one interval, chosen at random, also contained the signal. The interstimulus interval, measured from the offset of one masker to the onset of the next, was 500 ms. Listeners were required to select the interval containing the signal. The signal level was initially adjusted in steps of 8 dB. After every two reversals, the step size was halved until the minimum step size of 2 dB was reached. The run was terminated after another eight reversals. Threshold was defined as the median level at the last eight reversals. For every listener, four such estimates were made for each condition, and the mean and standard deviation of the four estimates were recorded. Listeners were given at least two hours practice before data were recorded. Responses were made via a computer keyboard, and feedback was provided via a computer monitor, which was placed outside the listening booth and was visible through a booth window. Listeners were tested in a single-walled sound-attenuating booth, which was situated in a sound-attenuating room.

C. Subjects

Four normal-hearing listeners participated as subjects. S1 was the author, S2 and S3 were female and male univer-

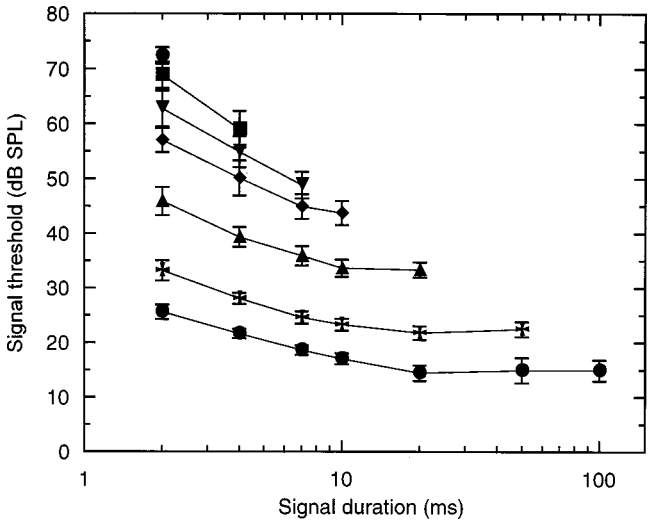


FIG. 1. Mean thresholds for a 4-kHz signal after a 40-dB spectrum level (*re*: 20 μPa) broadband forward masker, as a function of signal duration. The parameter is the time interval between the masker offset and the signal offset. These offset–offset intervals were (from top to bottom on the graph): 4, 6, 9, 12, 22, 52, and 102 ms. Error bars represent ± 1 standard error of the mean across the four listeners.

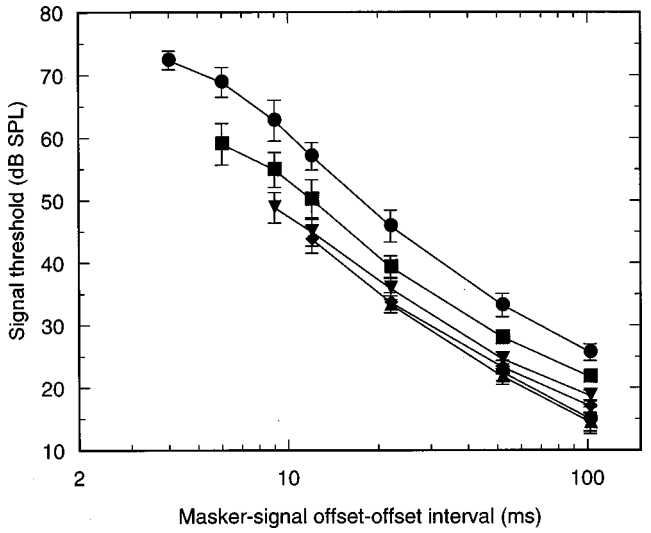


FIG. 2. Mean signal thresholds in forward masking, replotted from Fig. 1, as a function of masker-signal interval (measured from the masker offset to the signal offset). The parameter is signal duration. Durations were (from top to bottom on the graph): 2, 4, 7, 10, 20, 50, and 100 ms.

sity students, and S4 was a female university employee. S2 through S4 were paid for their participation. All had thresholds at octave frequencies between 250 and 8000 Hz of 15 dB HL or less. The ages of the listeners were 27, 19, 20, and 55 years for S1–S4, respectively.

D. Results

All four listeners showed a very similar pattern of results, and so only the mean data are shown in Fig. 1. Thresholds are plotted as a function of signal duration, with the masker-signal offset–offset interval as the parameter. Standard deviations across the four repetitions for each listener were generally less than 2.5, and always less than 5 dB. The error bars in Fig. 1 denote ± 1 standard error of the mean across the four listeners. All listeners showed considerable temporal integration for short offset–offset intervals. For instance, at an offset of 9 ms (downward-pointing triangles) mean thresholds decreased by nearly 14 dB as the signal duration increased from 2 to 7 ms. This is a *greater* change than would be expected for thresholds in quiet or in simultaneous masking, where a decrease of between 2.5 and 3 dB per doubling of duration is expected (e.g., Florentine *et al.*, 1988). At larger offset–offset intervals, the amount of temporal integration is less. Even so, for an offset–offset interval of 102 ms, the 11-dB decrease in mean thresholds as the signal duration is increased from 2 to 20 ms at least matches that normally found in quiet and in simultaneous masking. Little or no temporal integration was observed for signal durations greater than 20 ms.

Figure 2 depicts the mean data in a more conventional way, with thresholds plotted as a function of offset–offset interval, and signal duration as the parameter. The decay of forward masking is very similar to that observed in previous studies. However, in contrast to the studies of Zwislocki *et al.* (1959) and Elliott (1962), the data show a clear effect of signal duration for durations of 20 ms and less. The next section examines how well the results are quantitatively described by models of both adaptation and integration, with and without peripheral nonlinearity.

III. MODEL PREDICTIONS

A. Temporal-window model

The mean data from the experiment were fitted using the temporal-window model, as described in previous papers (Oxenham and Moore, 1994; Plack and Oxenham, 1998). Stimuli, including the noise maskers, were represented by envelopes which were flat during the steady-state periods, as it is assumed that the random fluctuations in the broadband noise do not play a decisive role in determining thresholds. The level of the noise was set at the expected level passing through an auditory filter centered at 4 kHz with an equivalent rectangular bandwidth (ERB) of 456 Hz, as derived by Glasberg and Moore (1990). The amplitude envelopes of the stimuli were then passed through a nonlinearity, designed to simulate the effect of BM compression. The compressive nonlinearity is described by the equations:

$$\begin{aligned} L_{\text{out}} &= 0.78 L_{\text{in}}, & L_{\text{in}} &< 35 \text{ dB SPL}, \\ L_{\text{out}} &= 0.16 L_{\text{in}} + 21.7, & L_{\text{in}} &\geqslant 35 \text{ dB SPL}. \end{aligned} \tag{1}$$

The slopes were determined by the psychophysical estimate of BM nonlinearity found by Oxenham and Plack (1997). The breakpoint of 35 dB is somewhat lower than that found in their study, but is the same as that used by Plack and Oxenham (1998). This breakpoint is also well within the range of physiological estimates, which range from 20 to 55 dB SPL (Murugasu and Russell, 1995; Ruggero *et al.*, 1997). The output of this stage is expressed in units of intensity (i.e., the values are squared). This is consistent both with psychophysical (Oxenham and Moore, 1995; van de Par and Kohlrausch, 1998) and physiological (Yates *et al.*, 1990) data.

TABLE I. Parameter values for the temporal-window model. Values from the three different fits are given, along with the sum of squared errors (SSQ). The total number of data points fitted was 29. Fit 1 was the original fit using the nonlinearity described by Eq. (1) and the window parameters from subject AO in Oxenham and Moore (1994). Fit 2 used the same nonlinearity as Fit 1, but allowed the window parameters to vary so as to provide the best fit (lowest sum of squared errors, SSQ). Fit 3 used a less compressive nonlinearity, with a slope of unity (in dB/dB coordinates) below 35 dB SPL and a slope of 0.25 above. The decision ratio was always allowed to vary to best fit the data.

	Parameters					
	T_{b1} (ms)	T_{b2} (ms)	w	T_a (ms)	Decision ratio	SSQ
Fit 1	4.0	29.0	0.0251	3.5	2.75	475.8
Fit 2	3.1	21.0	0.206	3.5	1.62	113.0
Fit 3	4.6	16.6	0.170	3.5	1.67	50.7

The stimuli were then passed through a sliding temporal integrator, or temporal window. The window comprises three exponential functions, one to describe backward masking (making it largely irrelevant for the present study), and two to describe the initial and later portions of forward masking. The window is described by the following equations:

$$W(t) = (1-w)\exp(t/T_{b1}) + w\exp(t/T_{b2}), \quad t<0,$$
$$W(t) = \exp(-t/T_a), \quad t\geq 0, \tag{2}$$

where t is time relative to the peak of the window, T_{b1} and T_{b2} are the time constants describing the decay of forward masking, w is the weighting factor determining the relative contributions of these two time constants, and T_a is the time constant describing the decay of backward masking. The parameters of the window were set in advance to those derived for subject AO in the study by Oxenham and Moore (1994), and are given in the first line of Table I.[1] The decision device was based on the ratio of the output due to the masker and signal, and output due to the masker alone. The point in time at which this value is greatest is assumed to be the time at which the decision is made. The signal-to-masker ratio provided the only free parameter, and was assumed to remain constant for all conditions. A constant ''internal noise'' was added to simulate threshold in quiet (see Oxenham and Moore, 1994, for details). This was set so as to correctly predict the mean threshold in quiet (20.9 dB SPL) for the 2-ms (half-amplitude duration) tone for listeners S1–S3. (This threshold was not measured in S4.)

Results from these simulations are shown in the right panel of Fig. 3. The left panel replots the mean data from Fig. 1 for comparison. The overall fit is not good, with the sum of squared errors for the 29 data points of 476, giving an rms error of 4.05 dB. This is caused primarily by very poor predictions of one or two data points, such as the threshold for the 4-ms offset–offset interval. The poor fit is perhaps not surprising, given that the window was derived using a simple power-law nonlinearity (Oxenham and Moore, 1994), rather than the more complex nonlinearity used here. It is known that changes in the input nonlinearity change the effective window shape (Penner, 1978; Oxenham *et al.*, 1997), meaning that the old window was probably not appropriate for the current nonlinearity. Nevertheless, the model manages to capture the main trends of the data rather well. In particular, the model shows an increase in the slope of integration at shorter offset–offset intervals, as in the data. This increased integration at short offset–offset intervals in the model predictions can be explained as follows. At the longer intervals, the overall signal level is low, falling into the more linear region of the BM input–output function. At shorter intervals, the overall signal level is above 40 dB SPL and would therefore be expected to fall within the more compressive region of the BM input–output function (Oxenham and Plack, 1997). As greater compression leads to a steeper integration slope (Penner, 1978; Oxenham *et al.*, 1997), greater integration for the higher signal levels is expected.

The lack of temporal integration for signal durations greater than 20 ms is also reflected in the predictions of the temporal-window model. Oxenham *et al.* (1997) proposed that integration for signal durations up to between 10 and 20 ms may be governed by a short-term temporal integrator. For durations longer than that, temporal integration in quiet and in simultaneous masking may be due to a multiple-looks mechanism, whereby information from discrete samples is combined over time (Viemeister and Wakefield, 1991). Such multiple looks would not provide much advantage in a forward-masking situation, as the ''looks'' closer to the masker would be less detectable and would hence provide

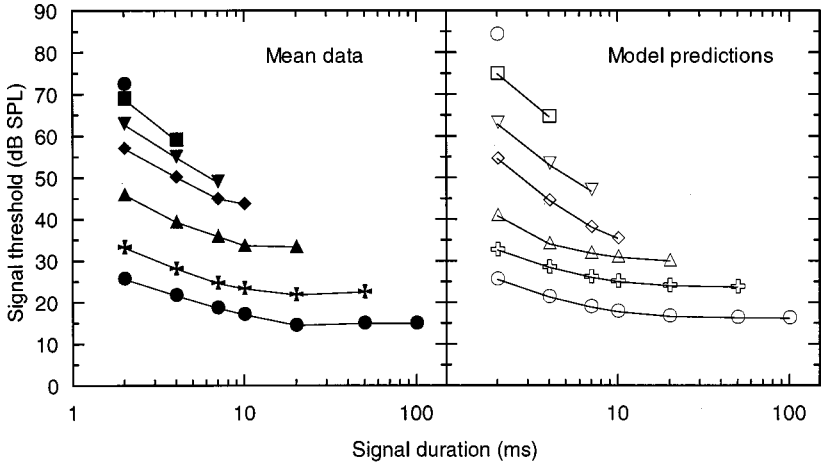


FIG. 3. Predictions of the temporal-window model (right panel) compared with the mean data (left panel). Thresholds are plotted as a function of signal duration, as in Fig. 1.

less information than looks further from the masker. Thus, temporal integration beyond 10 or 20 ms would not be expected.

The fact that the model captures the main trends of the data is encouraging, especially as all parameters except the signal-to-masker ratio were fixed in advance. However, as mentioned above, the overall fit is rather poor. This could be due to a nonoptimal choice of model parameters, or simply because it is not possible to accurately predict both the decay of forward masking and signal integration in forward masking using a simple integrator model. This is an important question: If the latter explanation is correct, it suggests that additional mechanisms, such as adaptation, may be required to provide a full account of forward masking.

To answer this question, the model was used to fit the data again, but this time certain parameters were either optimized using a minimization routine, or were fixed to different values. First, the nonlinearity remained the same [see Eq. (1)], but the three window parameters defining forward masking, T_{b1}, T_{b2}, and w, were allowed to vary so as to produce the best fit to the data (defined using a least-squares criterion). The best-fitting values, given in the second line of Table I, produced a considerably improved fit, with a total sum of squares of only 113, as opposed to 476.

Next, the nonlinearity was changed. Instead of using slopes of 0.78 and 0.16 for lower and higher signal levels, respectively, the values were changed to 1.0 and 0.25. These values are consistent with some previous studies (Oxenham and Moore, 1994; Oxenham *et al.*, 1997) and imply linear growth at low levels and compressive growth with a compression ratio of 4:1 (as opposed to 6:1) at levels above 35 dB SPL. Thus these values imply less compression overall, but are still within reasonable limits of what is known physiologically about BM nonlinearity. Again, the three window parameters were allowed to vary, and the resulting values are given in the third line of Table I. With these parameters, the overall fit was excellent, with a total sum of squares of only 50.7. These predictions are shown as curves along with the mean data in Fig. 4.

The additional modeling shows that it is possible to account accurately for both the decay of forward masking and temporal integration in forward masking using a simple linear integrator. As there is considerable interaction between the model parameters, it is not trivial to attach a meaning to the window changes necessary to provide a good fit to the data. It seems, however, that the data are best described by a somewhat less compressive nonlinearity than has been used in some previous studies.

It should be noted that the compressive nonlinearity within the model is essential in producing good fits to the data. This is demonstrated by the poor performance of a linear version of the model which integrates intensity, as described by Moore *et al.* (1988) and Plack and Moore (1990). With four free parameters (as with the nonlinear model), the best fit produced a sum of squared errors of 285—considerably worse than the nonlinear model. As expected from a linear model, the predicted amount of temporal integration was the same for all masker-signal offset intervals and for the signal in quiet. This is in contrast to the data, where more integration was observed at shorter masker-signal offset intervals. Also, the window shape necessary to predict a sufficiently rapid decay of forward masking was too narrow to predict the observed amount of temporal integration in any condition. For instance, the linear model predicted a decrease in threshold of 5.2 dB as the signal duration increased from 2 to 10 ms, while for the same increase in duration, the mean data show a decrease of 13.3 dB in the 12-ms offset–offset condition and 8.6 dB in the 102-ms condition. The present data therefore provide further evidence for the importance of including peripheral nonlinearity in models of temporal processing.

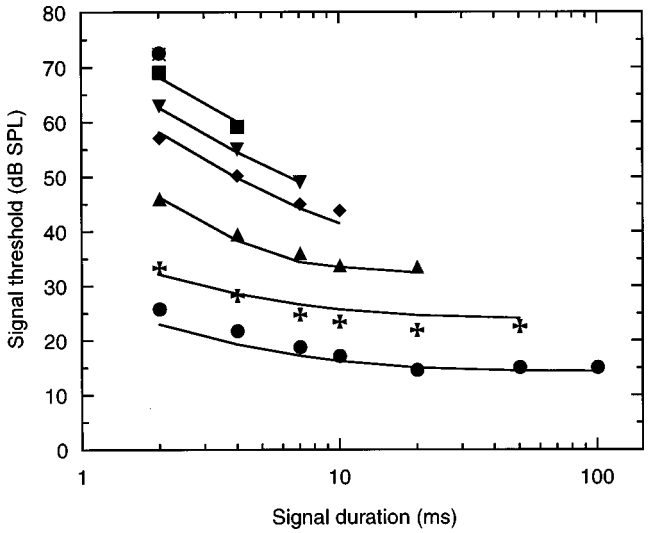


FIG. 4. Predictions of the modified temporal-window model (solid curves, and cross for the 4-ms offset condition) compared with the mean data (symbols). See text for model details. Thresholds are plotted as a function of signal duration, as in Fig. 1.

In summary, the results appear to be quantitatively consistent with the idea that forward masking is mediated by BM nonlinearity followed by linear temporal integration. However, as with all previous modeling studies (e.g., Plack and Oxenham, 1998), the success of the temporal-window approach does not rule out the possibility that an adaptation mechanism could account for the data equally well. The following section introduces a functional model of adaptation, to provide a quantitative comparison with the temporal-window model.

B. Adaptation model

In modeling adaptation, it is assumed that a signal is masked because the response of the auditory system to the signal is reduced to a level at or below threshold in quiet. Thus adaptation due to a forward masker is modeled as a gain function, which is lowest at the masker offset and increases as the time from the masker offset increases, such that the gain is unity (0 dB) at times of about 200 ms or greater. It is further assumed that the auditory system can integrate over the duration of the signal, with a rectangular window, and that the threshold for a given signal is determined by the integral of the signal intensity, modified by the continuously varying gain function. In this way, when no masker is present, the gain is always 0 dB and the model acts

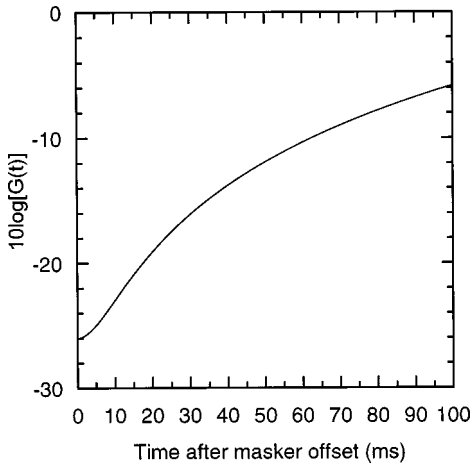


FIG. 5. Third-order polynomial function from Eq. (3), plotted as $10\log[G(t)]$, which was used to determine the instantaneous gain in the adaptation model.

as an energy detector at low levels, predicting thresholds in quiet as a function of duration reasonably well. Note that this model is in fact also a model of temporal integration. What distinguishes the two models is whether the masker is partly integrated with the signal (as in the temporal-window model) to produce masking, or whether (as in the adaptation model) it is assumed that the integration window can be optimally shaped and positioned by the auditory system so that no masker energy is integrated with the signal.

In this model, the stimuli were first passed through the same static nonlinearity that was used with the final, most successful, temporal-window model: for levels at and below 35 dB SPL the stimuli were passed linearly; for levels above 35 dB SPL the stimuli were compressed using a power-law nonlinearity with a slope of 0.25 in dB coordinates (i.e., 4:1 compression). As in the temporal-window model, the output of the nonlinearity was squared to provide units of intensity. Following this, the stimuli were multiplied by the gain function over the duration of the signal, to produce an ''adapted'' representation of the signal. At threshold, the integral of this adapted representation should be constant for all combinations of signal duration and masker-signal interval.

A number of mathematical functions were tried in attempting to describe the adaptation gain function. The most successful of these was a third-order polynomial. The values of the polynomial's four coefficients were fitted using a multidimensional minimization routine, incorporating all the mean data from the experiment. The best-fitting equation is given below, where $G(t)$ is the gain at time t in ms following masker offset:

$$G(t) = 2.83\times10^{-8}t^3 + 2.29\times10^{-5}t^2 + 1.96\times10^{-5}t + 2.5\times10^{-3}. \quad (3)$$

The threshold output was set by integrating the 2-ms signal at a level corresponding to threshold in quiet (20.9 dB SPL). Using this function, which is plotted in Fig. 5, the fit to the data was very good. The sum of squared errors was 68.7, which is comparable to the error of 50.7 from the best-fitting temporal window using the same number of free parameters. The pattern of predictions was also very similar to the data, predicting increasing amounts of integration at higher signal levels.

As with the temporal-window model, a linear version of the adaptation model, using a gain function derived from Plomp's (1964) exponential equation, was not successful in predicting the data, producing an overall sum of squared errors of 534.4. As with the linear temporal-window model, insufficient temporal integration was observed. However, whereas the linear temporal-window model predicted equal integration at all masker-signal offsets, the linear adaptation model predicted more integration at longer offset–offset intervals than at shorter intervals. This is contrary to the data, but is expected because the adaptation gain function becomes shallower (and so has less influence) at longer intervals when plotted as a function of linear time; at shorter offset intervals, the gain function changes more rapidly, leading to less integration being predicted.

In summary, an adaptation model was almost as successful at predicting the data as the temporal-window model. For both categories of model, the initial nonlinearity was crucial in producing good predictions. Thus, taken on their own, the data from the present study do not distinguish between integration and adaptation as mechanisms underlying forward masking.

C. Accounting for temporal integration in simultaneous masking with the temporal-window model and the adaptation model

The previous two sections showed that both integration and adaptation models could account reasonably well for the data from the present study, given sufficient free parameters. At this point it is worth considering whether the two mechanisms could ever be distinguished, or whether they are in some way mathematically equivalent. In the adaptation model, a rectangular integration window was used. At levels above 35 dB SPL, the 4:1 compression of the nonlinearity would result in an integration slope four times steeper than at low levels for a long integration window (Penner, 1978), which is much more than actually measured. This ''over-integration'' is offset by the effects of adaptation, which attenuate early portions of the signal in much the same way that a temporal window would. In this way, the effect of adaptation in forward masking can be thought of as a temporal weighting function and a trade can be established between the recovery from adaptation (zero in the case of the temporal-window model) and the attenuation due to the temporal weighting function (zero in the case of the rectangular-window adaptation model).

This apparent equivalence of adaptation and integration only holds in the presence of a forward masker. In the absence of recovery from adaptation, the two models become different, with the temporal-window model retaining the same weighting function and the adaptation model changing its effective weighting function (to become rectangular in the present model). To illustrate this point, data were taken from a study of temporal integration in simultaneous masking (Oxenham *et al.*, 1997). The signal was a 6-kHz sinusoid, gated with 1-ms ramps and the masker was a broadband Gaussian noise at a spectrum level of 20 dB (*re*: 20 μPa).

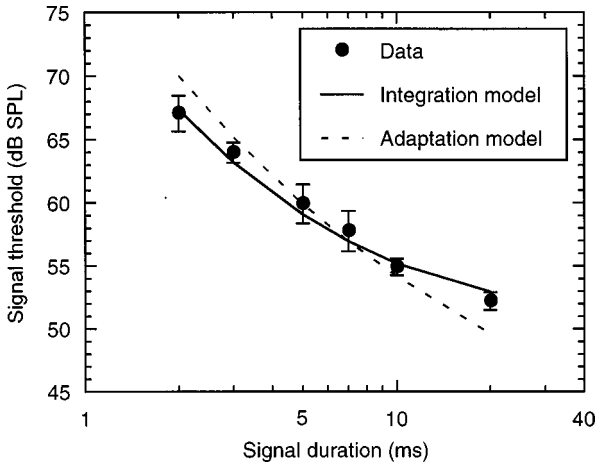


FIG. 6. Signal integration in the presence of a simultaneous masker, taken from Oxenham *et al.* (1997) (circles) together with the predictions from the temporal-window model (solid curves) and the adaptation model (dashed curves).

This masker level was chosen here as the threshold levels most closely match the levels found for short masker-signal offset intervals in the present study. Data, representing the mean of four normal-hearing listeners, are shown in Fig. 6 for signal durations from 2 to 20 ms; see Oxenham *et al.* (1997) for further details. Both the temporal-window and the adaptation models were used to simulate the data. For the adaptation model, it was assumed that recovery from adaptation plays no role in simultaneous masking, and so only the rectangular weighting function (20-ms duration) was used. The window shapes and the nonlinearity were kept as they were for the forward-masking data—the only free parameter was the threshold signal-to-masker ratio. The predictions of the temporal-window model and the adaptation model are shown as a solid curve and dashed curve, respectively.

It can be seen that the temporal-window model provides a better description of the simultaneous-masking data than does the adaptation model, which predicts too much integration. The difference in thresholds between the 2-ms and 20-ms signal is 14.8 dB, compared with 14.4 dB predicted by the temporal-window model and 20.4 dB predicted by the adaptation model. Thus the adaptation model overestimates the amount of temporal integration between 2 and 20 ms by almost 6 dB. This suggests that, while both models can provide a reasonable account of signal integration in forward masking, the temporal-window model produces predictions that are more consistent with both simultaneous and forward masking. In other words, the assumption of time invariance, inherent in the temporal-window approach, seems to be justified.

The support for the temporal-window approach relies on the assumption that the integration window shape remains the same for both simultaneous and nonsimultaneous masking. If one is prepared to allow the window shape to vary between these two conditions, then the adaptation model could be altered to bring it into line with the simultaneous-masking data also. At present it appears more parsimonious to assume that the window shape remains constant.

D. Why ignore adaptation?

The emphasis so far has been on distinguishing between integration (or persistence) and adaptation on an either/or basis. Clearly, it is not currently possible to rule out some complex interaction between both mechanisms. Also, given that adaptation is clearly observable at the level of the auditory nerve, it may seem perverse to ignore it completely. On the other hand, it may be that adaptation acts somewhat like an automatic gain control, which equally affects the response to stimuli *and* the spontaneous rate, or noise floor. If this is so, the net effect of adaptation on forward masking may approach zero. Some support for this view can be found in the work of Relkin and Turner (1988). They argued that, because spontaneous neural firing, as well as the response to the signal, is reduced by a forward masker, the physiological measure of forward masking should not simply be the reduction in the response to the signal, but rather the discriminability of the response to the signal from the response to no signal. When they did this, they found many auditory nerve fibers that exhibited little or no forward masking. These arguments and findings can be interpreted as support for an approach that ignores the effects of neural adaptation in the periphery.

E. Limitations of the models

While the temporal-window model has proved useful, especially in its ability to elucidate the role of peripheral compression, there are many aspects of it which are unsatisfactory. For instance, the decision device uses only one instant in time, rather than combining information over time (Oxenham and Moore, 1994); it ignores statistical fluctuations, taking into account only average power; and it cannot make use of any fine-structure cues. The window itself provides a reasonable description of forward masking, decrement detection (Oxenham and Moore, 1994), and short-term signal integration (Oxenham *et al.*, 1997), but it is not capable of describing modulation detection (Dau *et al.*, 1997) or of capturing the asymmetry between increment and decrement detection (Oxenham, 1997). Perhaps it could be viewed as the low-pass filter in a bank of modulation filters (Dau *et al.*, 1997).

Unlike the model proposed by Dau *et al.* (1996a,1997), no set of parameters have so far been proposed which can provide a reasonable fit to a wide variety of data. Indeed, the often *ad hoc* alteration of window shape for specific data sets (Alcántara *et al.*, 1996; Carlyon and Datta, 1997), including the ones presented here, suggests that no such set may be possible. Nevertheless, the simplicity of the model, in terms of both its assumptions and implementation, make it a useful tool for testing other hypotheses, such as the one addressed in the present study. While the data presented here and elsewhere support the idea of linear temporal integration following peripheral nonlinearity, many other peripheral and central factors, which are not currently captured by the temporal-window model, may also play a role.

Another inconsistency between the temporal-window model predictions and the data is that the model does not predict elevated signal thresholds at the onset of a long

masker, relative to thresholds in the temporal center. This effect is known as overshoot (Zwicker, 1965). Interestingly, the effect requires a relatively wide-band or off-frequency masker with a narrow-band, high-frequency signal. This suggests that the effect will never be captured satisfactorily with a single-channel model, at least with a linear filter. In contrast, the temporal-window model predicts ''undershoot,'' i.e., that thresholds at the onset of a masker will be somewhat lower than thresholds in the temporal center. It is possible that the mechanism responsible for overshoot produces an effect that is sufficiently large to counteract the undershoot that would otherwise be observed. However, providing a quantitative account of overshoot will be challenging, as individual differences can be very large (Bacon and Liu, 2000).

The adaptation model presented here is simply an *ad hoc* construction designed to test the feasibility of an adaptation mechanism in principle. It is therefore unlikely that it would be useful in its current form as a more general model of auditory processing.

Alterations in either model may change the predictions in many ways. The apparently poorer fits of the adaptation model to the simultaneous-masking data might change if different assumptions about the influence of, for instance, noise variability were made. While the present results lean in favor of the temporal-window model, further work will be necessary to provide a definitive answer to the question of whether integration or adaptation is more responsible for forward masking.

IV. SUMMARY

Temporal integration in forward masking was studied as a way to try to distinguish between explanations of forward masking in terms of adaptation and integration. Signal thresholds in the presence of a broadband forward masker were measured as a function of signal duration, with the time interval between the masker offset and the signal offset (offset–offset interval) held constant at a value between 4 and 102 ms.

The experiment showed that substantial temporal integration was observed for signal durations between 2 and 20 ms. The amount of temporal integration varied with offset–offset interval: at the shorter offset–offset intervals, the amount of temporal integration exceeded that normally found in quiet, while at longer offset–offset intervals, where the signal level fell below about 40 dB SPL, the amount of integration matched that found in quiet. No further improvements in threshold were found for signal durations longer than 20 ms.

The data were used to test models of temporal integration and adaptation. For both models, the inclusion of peripheral nonlinearity, resembling the input–output function of the basilar membrane, was essential in predicting the data. Using such a nonlinearity, both models provided similarly good fits to the data. However, the temporal-integration model provided a better fit to data from simultaneous masking with the same parameters. If a constant window shape is assumed for both forward and simultaneous masking, the results favor the view that forward masking is better described as a persistence in neural activity than as neural adaptation. However, the weight of evidence is at present still not overwhelming.

ACKNOWLEDGMENTS

This work was supported mainly by NIH/NIDCD Grant No. R01 DC 03909. The data were collected while the author was a Wellcome Trust Research Fellow (Grant No. 0044215/Z/95/Z) at the Institute for Perception Research (IPO), Eindhoven, The Netherlands. I thank Neal Viemeister for helpful exchanges on the modeling of adaptation. Comments by Chris Plack and two anonymous reviewers resulted in a substantial improvement of this paper.

[1]An error in Table I of Oxenham and Moore (1994) gives the value of $10\log(w)$ for the AO exponential, instead of $20\log(w)$; all other entries in that table are correct. This incorrect value was also used in the study of Plack and Oxenham (1998), although that paper actually cited the corrected value. The correct value, which was used here, is -32 and not -16 dB, or 0.025 in linear units.

Alcántara, J. I., Holube, I., and Moore, B. C. J. (**1996**). ''Effects of phase and level on vowel identification: Data and predictions based on a nonlinear basilar-membrane model,'' J. Acoust. Soc. Am. **100**, 2382–2392.

Bacon, S. P. (**1996**). ''Comments on ''Manipulations of the duration and relative onsets of two-tone forward maskers'' [J. Acoust. Soc. Am. 94, 1269–1274 (1993)],'' J. Acoust. Soc. Am. **99**, 3246–3248 (L).

Bacon, S. P., and Liu, L. (**2000**). ''Effects of ipsilateral and contralateral precursors on overshoot,'' J. Acoust. Soc. Am. **108**, 1811–1818.

Carlyon, R. P. (**1988**). ''The development and decline of forward masking,'' Hear. Res. **32**, 65–80.

Carlyon, R. P., and Datta, A. J. (**1997**). ''Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,'' J. Acoust. Soc. Am. **101**, 3636–3647.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). ''Modeling auditory processing of amplitude modulation. I. Detection and masking with narrowband carriers,'' J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996a**). ''A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure,'' J. Acoust. Soc. Am. **99**, 3615–3622.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996b**). ''A quantitative model of the 'effective' signal processing in the auditory system. II. Simulations and measurements,'' J. Acoust. Soc. Am. **99**, 3623–3631.

de Maré, G. (**1940**). ''Fresh observations on the so-called masking effect of the ear and its possible diagnostic significance,'' Acta Oto-Laryngol. **28**, 314–316.

Duifhuis, H. (**1973**). ''Consequences of peripheral frequency selectivity for nonsimultaneous masking,'' J. Acoust. Soc. Am. **54**, 1471–1488.

Elliott, L. L. (**1962**). ''Backward masking: Monotic and dichotic conditions,'' J. Acoust. Soc. Am. **34**, 1108–1115.

Elliott, L. L. (**1971**). ''Backward and forward masking,'' Audiology **10**, 65–76.

Fastl, H. (**1976a**). ''Temporal masking effects: I. Broadband noise masker,'' Acustica **35**, 287–302.

Fastl, H. (**1976b**). ''Temporal masking effects: II. Critical band noise masker,'' Acustica **36**, 317—331.

Fastl, H. (**1979**). ''Temporal masking effects: III. Pure tone masker,'' Acustica **43**, 283–294.

Florentine, M., Fastl, H., and Buus, S. (**1988**). ''Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking,'' J. Acoust. Soc. Am. **84**, 195–203.

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

Gorga, M. P., Stelmachowicz, P. G., Abbas, P. J., and Small, A. M. J. (**1980**). ''Some observations on simultaneous and nonsimultaneous masking,'' J. Acoust. Soc. Am. **67**, 1821–1822.

Gralla, G. (**1992**). ''Wahrnehmungskriterien bei Simultan- und Nachhörschwellenmessungen (Detection cues in simultaneous and post masking experiments),'' Acustica **77**, 243–251.

Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). ''Forward masking as a function of frequency, masker level, and signal delay,'' J. Acoust. Soc. Am. **71**, 950–962.
Kidd, G., and Feth, L. L. (**1981**). ''Patterns of residual masking,'' Hear. Res. **5**, 49–67.
Lüscher, E., and Zwislocki, J. J. (**1947**). ''The delay of sensation and the remainder of adaptation after short pure-tone impulses on the ear,'' Acta Oto-Laryngol. **35**, 428–455.
Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Growth of forward masking for sinusoidal and noise maskers as a function of signal delay: Implications for suppression in noise,'' J. Acoust. Soc. Am. **73**, 1249–1259.
Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). ''The shape of the ear's temporal window,'' J. Acoust. Soc. Am. **83**, 1102–1116.
Munson, W. A., and Gardner, M. B. (**1950**). ''Loudness patterns—a new approach,'' J. Acoust. Soc. Am. **22**, 177–190.
Murugasu, E., and Russell, I. J. (**1995**). ''Salicylate ototoxicity: The effects on basilar membrane displacement, cochlear microphonics, and neural responses in the basal turn of the guinea pig cochlea,'' Aud. Neurosci. **1**, 139–150.
Neff, D. L. (**1986**). ''Confusion effects with sinusoidal and narrowband-noise forward maskers,'' J. Acoust. Soc. Am. **79**, 1519–1529.
Nelson, D. A. (**1991**). ''High-level psychophysical tuning curves: Forward masking in normal-hearing and hearing-impaired listeners,'' J. Speech Hear. Res. **34**, 1233–1249.
Nelson, D. A., and Swain, A. C. (**1996**). ''Temporal resolution within the upper accessory excitation of a masker,'' Acust. Acta Acust. **82**, 328–334.
Oxenham, A. J. (**1997**). ''Increment and decrement detection in sinusoids as a measure of temporal resolution,'' J. Acoust. Soc. Am. **102**, 1779–1790.
Oxenham, A. J., and Moore, B. C. J. (**1994**). ''Modeling the additivity of nonsimultaneous masking,'' Hear. Res. **80**, 105–118.
Oxenham, A. J., and Moore, B. C. J. (**1995**). ''Additivity of masking in normally hearing and hearing-impaired subjects,'' J. Acoust. Soc. Am. **98**, 1921–1934.
Oxenham, A. J., and Moore, B. C. J. (**1997**). ''Modeling the effects of peripheral nonlinearity in normal and impaired hearing,'' in *Modeling Sesnsorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ).
Oxenham, A. J., Moore, B. C. J., and Vickers, D. A. (**1997**). ''Short-term temporal integration: Evidence for the influence of peripheral compression,'' J. Acoust. Soc. Am. **101**, 3676–3687.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Penner, M. J. (**1975**). ''Persistence and integration: Two consequences of a sliding integrator,'' Percept. Psychophys. **18**, 114–120.
Penner, M. J. (**1978**). ''A power law transformation resulting in a class of short-term integrators that produce time-intensity trades for noise bursts,'' J. Acoust. Soc. Am. **63**, 195–201.
Plack, C. J., and Moore, B. C. J. (**1990**). ''Temporal window shape as a function of frequency and level,'' J. Acoust. Soc. Am. **87**, 2178–2187.
Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.
Plomp, R. (**1964**). ''The rate of decay of auditory sensation,'' J. Acoust. Soc. Am. **36**, 277–282.
Relkin, E. M., and Turner, C. W. (**1988**). ''A reexamination of forward masking in the auditory nerve,'' J. Acoust. Soc. Am. **84**, 584–591.
Rhode, W. S. (**1871**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.
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.
Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). ''Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique,'' J. Acoust. Soc. Am. **72**, 131–141.
Shannon, R. V. (**1990**). ''Forward masking in patients with cochlear implants,'' J. Acoust. Soc. Am. **88**, 741–744.
Smith, R. L. (**1977**). ''Short-term adaptation in single auditory nerve fibers: Some poststimulatory effects,'' J. Neurophysiol. **40**, 1098–1112.
Smith, R. L. (**1979**). ''Adaptation, saturation, and physiological masking in single auditory-nerve fibers,'' J. Acoust. Soc. Am. **65**, 166–178.
Thornton, A. R. D. (**1972**). ''PSM studies II. Post-stimulatory masked thresholds as a function of probe tone duration,'' J. Sound Vib. **22**, 183–191.
Turner, C. W., Relkin, E. M., and Doucet, J. (**1994**). ''Psychophysical and physiological forward masking studies: Probe duration and rise-time effects,'' J. Acoust. Soc. Am. **96**, 795–800.
van de Par, S., and Kohlrausch, A. (**1998**). ''Diotic and dichotic detection using multiplied-noise maskers,'' J. Acoust. Soc. Am. **103**, 2100–2110.
Viemeister, N. F., and Wakefield, G. H. (**1991**). ''Temporal integration and multiple looks,'' J. Acoust. Soc. Am. **90**, 858–865.
Vogten, L. L. M. (**1978**). ''Low-level pure-tone masking: A comparison of ''tuning curves'' obtained with simultaneous and forward masking,'' J. Acoust. Soc. Am. **63**, 1520–1527.
Widin, G. P., and Viemeister, N. F. (**1979**). ''Intensive and temporal effects in pure-tone forward masking,'' J. Acoust. Soc. Am. **66**, 388–395.
Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). ''Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range,'' Hear. Res. **45**, 203–220.
Zwicker, E. (**1965**). ''Temporal effects in simultaneous masking by white-noise bursts,'' J. Acoust. Soc. Am. **37**, 653–663.
Zwicker, E. (**1984**). ''Dependence of post-masking on masker duration and its relation to temporal effects in loudness,'' J. Acoust. Soc. Am. **75**, 219–223.
Zwislocki, J., Pirodda, E., and Rubin, H. (**1959**). ''On some poststimulatory effects at the threshold of audibility,'' J. Acoust. Soc. Am. **31**, 9–14.

Three-dimensional vocal tract imaging and formant structure: Varying vocal register, pitch, and loudness

Kenneth Tom
Department of Speech Communication, California State University Fullerton, Fullerton, California 92831

Ingo R. Titze
Department of Speech Pathology and Audiology, National Center for Voice and Speech, University of Iowa, Iowa City, Iowa 52242

Eric A. Hoffman
Division of Physiologic Imaging, Department of Radiology, University of Iowa College of Medicine, Iowa City, Iowa 52242

Brad H. Story
Department of Speech and Hearing Sciences, University of Arizona, Tucson, Arizona 85721

(Received 26 July 1999; accepted for publication 17 October 2000)

Although advances in techniques for image acquisition and analysis have facilitated the direct measurement of three-dimensional vocal tract air space shapes associated with specific speech phonemes, little information is available with regard to changes in three-dimensional (3-D) vocal tract shape as a function of vocal register, pitch, and loudness. In this study, 3-D images of the vocal tract during falsetto and chest register phonations at various pitch and loudness conditions were obtained using electron beam computed tomography (EBCT). Detailed measurements and differences in vocal tract configuration and formant characteristics derived from the eight measured vocal tract shapes are reported. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1332380]

PACS numbers: 43.70.Aj, 43.70.Jt, 43.75.Rs [AL]

I. INTRODUCTION

In the adult male voice, the highest fundamental frequencies of the speaking or singing voice are usually produced in falsetto register. Loud, high-pitched phonation in falsetto register can be thought of as one extreme of the range of capabilities of the vocal mechanism. Physiologically, however, falsetto phonation can be produced throughout the upper third to two-thirds of the fundamental frequency range at various degrees of vocal intensity. For most voices, there is also some degree of overlap in fundamental frequencies that can be produced in either chest or falsetto register (Colton and Hollien, 1972; Hollien, 1974, 1977; Titze, 1988, 1994; Welch *et al.*, 1988). Although falsetto phonation is an intrinsic part of human vocal production, and its use as a phonatory setting in speech (Laver, 1980) and numerous styles of singing recognized (Malm, 1967; Giles, 1982), there are few studies examining systematic changes in three-dimensional vocal tract morphology and corresponding changes in vocal tract resonances in falsetto versus chest (modal) register phonation. Detailed measurements of 3-D vocal tract dimensions in falsetto and chest register phonations would (1) facilitate more natural computer synthesis/simulation of phonation in falsetto mode, (2) provide insights into the nature of articulatory changes in vocal tract shape as a function of register, pitch and loudness, and (3) allow for the derivation of corresponding formant structure. Formant and vocal tract shape information for falsetto phonations in a singing voice at various pitch/loudness conditions versus normal speech phonation in chest register would also allow one to examine vocal tract changes for the presence of articulatory manipulations to increase either vocal intensity (e.g., use of a singer's formant) or consistency in vocal timbre (vowel modification).

Three-dimensional imaging and measurement techniques have been used to characterize the vocal tract configurations of numerous phonetic structures and to calculate relevant formant information. Magnetic resonance imaging (MRI) has been used to acquire 3-D images of vocal tract shapes for vowels and continuants (Baer *et al.*, 1991; Moore, 1992; Sulter *et al.*, 1992), sustained dark and light allophones of /l/ (Narayanan *et al.*, 1997), and rhotics (Alwan *et al.*, 1997), as well as a larger inventory of speaker-specific vowels and consonants (Story *et al.*, 1996). High-resolution 3-D images of differences in vocal tract shape as a function of register, fundamental frequency, and vocal intensity have not been available.

Most acoustically based studies of formant structure in adult speech production have excluded high-pitched phonation (Hillenbrand *et al.*, 1995; Huber *et al.*, 1999; Lienard and Di Benedetto, 1999; Peterson and Barney, 1952). This is due, at least in part, to practical and theoretical limitations of currently available acoustic analysis techniques. Because spectral harmonics are so widely spaced in phonations with high fundamental frequencies, accurate estimates of formant frequencies and bandwidths in high-pitched falsetto phonations are difficult, if not impossible, to obtain using standard acoustic techniques, e.g., linear predictive coding (LPC) analysis (Markel and Gray, 1976).

The objectives of the current study were threefold: (1) to

acquire high-resolution three-dimensional images of the vocal tract during phonation at various pitch and loudness conditions in falsetto and chest register for a single subject, a trained singer, (2) to obtain the corresponding vocal tract length and cross-sectional area functions, and (3) to specify formant frequencies and bandwidths associated with each phonatory condition. As an alternative to LPC techniques for estimating formants, an analysis by synthesis approach was used. Vocal tract area functions derived from high-resolution 3-D vocal tract image sets were used to calculate corresponding formant data (Titze *et al.*, 1994; Tom, 1996).

Three-dimensional imaging of the vocal tract can be accomplished by obtaining a series of contiguous image slices through the portion of the body encompassing the vocal tract, segmenting the airway shape from its bordering tissues and reconstructing it in three dimensions. Images can be acquired with either x-ray computed tomography (x-ray CT) or magnetic resonance imaging (MRI). Each technique has its advantages and disadvantages. In terms of reducing the risk of any adverse side effects, MRI has the clear advantage. No hazardous effects have been observed from short term exposures to the magnetic fields currently used in MRI scanning systems. For imaging airways, however, MRI techniques have a number of disadvantages (Baer *et al.*, 1991; Moore, 1992; Sulter *et al.*, 1992). Image resolution and accuracy are limited. Air-to-tissue boundaries can be distorted due to MRI artifacts, effectively blurring the edges of the vocal tract slightly. Tissues that are low in hydrogen content, such as bony structures and teeth, are captured poorly and appear to be the same gray scale density as air. Using MRI technology available at the time this study was performed, the scanner activation time required to scan an entire vocal tract was approximately 4 to 5 min (Story *et al.*, 1996), depending on the desired resolution and scanning parameters being used. The addition of pauses required for breathing when imaging the vocal tract during actual phonation increased total image acquisition time to 10 to 15 min per vocal tract shape. Under such circumstances subject fatigue and movement artifact become an important consideration. Because the present study included phonations at high effort conditions that could not be sustained over the total image acquisition time required by MRI techniques, the use of MRI was not feasible.

At the time these studies were performed, the scanner of choice amongst the x-ray CT technologies was electron beam computed tomography (EBCT) because of the high speed of image acquisition (100 ms per slice). For imaging airways, electron beam computed tomography techniques yield images of higher resolution than MRI images. The air–tissue boundary is captured with greater accuracy and bony structures and teeth are clearly imaged. Using EBCT scanners, a high-resolution volumetric study encompassing the entire vocal tract can be scanned relatively quickly (12 to 18 s). This comparatively brief image acquisition time greatly reduces the potential for subject fatigue and associated movement artifact, which can blur resultant images. The chief disadvantage associated with EBCT is its use of ionizing radiation, which limits the number of scans considered safe (International Commission on Radiological Protection, 1977; National Council on Radiation Protection and Measurements, 1987).

TABLE I. Phonatory conditions for vocal tract imaging.

Pitch/Condition	Vowel for all phonations=/ɑ/ Register	Pitch (F_0)	Loudness
B-flat$_4$ ff	Falsetto	B-flat$_4$ (466 Hz)	very loud (ff)
B-flat$_4$ mp	Falsetto	B-flat$_4$ (466 Hz)	moderately soft (mp)
F_4 ff	Falsetto	F_4 (349 Hz)	very loud (ff)
F_4 mp	Falsetto	F_4 (349 Hz)	moderately soft (mp)
C_4 ff	Falsetto	C_4 (262 Hz)	very loud (ff)
C_4 mp	Falsetto	C_4 (262 Hz)	moderately soft (mp)
Speech loud	Chest	D_3 (147 Hz)	very loud
Speech comfortable	Chest	B-flat$_2$ (117 Hz)	comfortable loudness

II. IMAGE ACQUISITION AND ANALYSIS

A. Image acquisition protocol

Volumetric images of the vocal tract were scanned from a single male subject for sustained phonation of the vowel /ɑ/ under eight phonatory conditions, varying voice register, pitch, and loudness levels. The conditions, summarized in Table I, included three sung falsetto register pitch levels (low pitch, 262 Hz; medium pitch, 349 Hz; high pitch, 466 Hz) and chest register speech phonation at a self-selected comfortable pitch level. Each of these four pitch levels was scanned at two intensity levels, moderately low intensity (*mezzo piano, mp*) and very loud intensity (*fortissimo, ff*).

For the purposes of this study, a trained singer who could readily and consistently produce falsetto as well as chest register phonations throughout his fundamental frequency and intensity ranges was recruited. The subject for this study was a 45-year-old adult male who has had extensive singing training in the Western classical tradition, including 12 years of vocal study as a baritone and 6 years of study as a countertenor. An active performer, he has sung as a countertenor for the past 11 years performing early music using a falsetto-based singing technique to vocalize in the alto range. The subject's medical and recent health history were unremarkable and there was no history of speech or hearing disorders. The subject is a native speaker of General North American English and his speaking fundamental frequency was within normal limits. The subject's fundamental frequency range spanned from 69 Hz ($C\sharp_2$) to 392 Hz (G_4) in chest register and from 147 Hz (D_3) to 587 Hz (D_5) in falsetto register.

For each phonatory condition to be scanned, the subject was positioned comfortably in a supine position on the imaging table. His lower neck was supported and stabilized with a rolled towel. Head positioning was aligned before each set of scans such that the Frankfort plane was perpendicular to the imaging table and the anatomic midline centered. Each phonatory condition required approximately 20 s of actual scanner activation time. Total acquisition time depended on how long the subject could repeatedly prolong the target phoneme /ɑ/ at a particular phonatory condition without significant movement of vocal tract structures. The following system was devised to time scanning interruptions, which allowed the subject to rest and breathe between vowel

repetitions. Before each condition, the subject produced several trial utterances to gauge how many seconds he could produce steady, consistent phonations without introducing noticeable movement artifact in the images. When this utterance length was determined, the radiology technician timed pauses such that he stopped before the end of a vowel reiteration and reinitiated imaging as soon as the subject began the next reiteration, as monitored over the intercom. Including these pauses, the total time required to image the vocal tract for each condition ranged from approximately 60 to 90 s.

B. EBCT scanning parameters

The EBCT images were acquired with an Imatron C-150 scanner (Boyd and Lipton, 1983). Each volume set consisted of 60 contiguous, parallel, axial slices. Slice thickness was 3 mm. These scanned images encompassed the hard palate superiorly, the first tracheal ring inferiorly, the lips anteriorly, the posterior pharyngeal wall posteriorly, and the buccal walls to the left and right of vocal tract air space. Slice scan aperture was 100 ms. The field of view (FOV) for each slice was 21 cm and the image matrix was 512×512 pixels. The resolution in the plane of imaging (axial) was 0.410 mm, which is near the theoretical limit of the scanner's resolution.

The accuracy of the image acquisition and analysis procedures using the Imatron C-150 scanner has been assessed with a tubular phantom of known dimensions (Story, 1995). The phantom consisted of three connected sections of air-filled tubing placed in a closed water-filled plastic enclosure. Known and measured cross-sectional areas of the phantom differed by 1.8% to 2.0%.

C. Image analysis

Image analysis was accomplished in three stages, i.e., image segmentation, 3-D airway reconstruction, and airway measurement. These procedures were performed using UNIX-based image display and measurement software called VIDA™ (volumetric image display and analysis), which was developed by Hoffman and colleagues (Hoffman *et al.*, 1992). Further information regarding all VIDA modules can be accessed at http://everest.radiology.uiowa.edu. These image analysis techniques, as applied to vocal tract airway analysis, have been described in detail by Story (1995) and Story *et al.* (1996).

The vocal tract was segmented, i.e., differentiated from surrounding tissue, using a seeded region growing technique whereby all airway voxels (3-D pixels) were assigned a unique gray scale value (Hoffman *et al.*, 1983; Udupa, 1991). Reconstruction of the vocal tract in three dimensions was accomplished using a process called shape-based interpolation (Raya and Udupa, 1990; Udupa, 1991) on the segmented image set, yielding a stack of slices with the same voxel dimension (0.410 mm) along all three axes. The reconstructed 3-D image data from the shape-based interpolated vocal tract was the basis for subsequent cross-sectional area measurements. The edges of the interpolated airway shape were also used to perform 3-D surface renderings of the vocal tract. Graphically represented as a 3-D object with the use of shading, surface renderings can be displayed at any number of angles or magnification levels. The display itself cannot be measured directly, but allows the user to assess the quality of the segmentation procedure and to observe 3-D views of the vocal tract's outer shape. To measure cross-sectional areas from the shape-based interpolated data set, an algorithm designed to study the upper airway (Hoffman and Gefter, 1990; Hoffman *et al.*, 1992) was used. Tube length (in this case, vocal tract length) was quantified using methods described by Story *et al.* (1996). Formants were obtained with a wave-reflection vocal tract model (Kelly and Lochbaum, 1962; Liljencrants, 1985) using area functions from the 3-D image data as input, and calculating its response to an impulse excitation.

III. RESULTS AND DISCUSSION

A. Quantitative area functions

The ''raw'' area functions measured from the volumetric image data for the eight phonatory conditions were discretized for use in speech simulation (Story *et al.*, 1996; Tom, 1996). The process of discretization involved choosing the discretized vocal tract length (even multiples of vocal tract sections 0.396 825 cm in length) that best fit the measured vocal tract lengths, normalizing measured data to this length, fitting the data to a cubic spline curve, and sampling the cubic spline curve at equally spaced intervals of 0.396 825 cm. The discretized vocal tract length that best fit all eight phonatory conditions was 17.46 cm. Detailed numerical area functions for the eight phonatory conditions based on the discretized vocal tracts can be accessed at http://www.ncvs.org/rescol/articles/vocaltract.html. Measured vocal tract lengths for the eight phonatory conditions can also be accessed at this website. The change from soft to loud intensity (higher effort phonations) in the medium- and high-pitch falsetto phonations was consistently associated with an increase in measured vocal tract length, while the opposite pattern occurred for speech and low falsetto pitch conditions.

Numerical area functions for the piriform sinuses used in acoustic modeling are listed in Table II. The main trend, with regard to changes in the 3-D shape of the piriform sinuses from soft to loud intensity within each pitch condition, was an increase in both length and cross-sectional areas. The measured lengths of the piriform sinuses are presented in Table III. Some slight left–right asymmetries in piriform sinus length occurred when the superior–inferior alignment of

TABLE II. Piriform sinus area functions in square centimeters, at equal intervals of 0.396 825 cm, expressed as a single branch for acoustic modeling purposes. Section 1 represents the area function of the superior-most portion of the piriform sinuses. Piriform sinus branch length for all conditions was 1.59 cm (four sections at 0.396 825 cm/section).

Section No.	B-flat$_4$ ff	B-flat$_4$ mp	F$_4$ ff	F$_4$ mp	C$_4$ ff	C$_4$ mp	Speech loud	Speech comfortable
1	2.39	1.34	2.16	1.82	2.05	2.17	2.58	2.26
2	2.13	1.13	2.07	1.60	2.24	1.94	2.41	2.21
3	1.88	1.02	2.19	1.50	2.28	1.36	2.60	1.95
4	1.26	0.83	1.83	1.04	1.74	0.11	1.29	0.83

TABLE III. Piriform sinus lengths associated with variations in vocal register, pitch, and loudness levels.

Condition	Left piriform sinus (cm)	Right piriform sinus (cm)
B-flat$_4$ ff	1.76	1.52
B-flat$_4$ mp	1.68	1.68
F_4 ff	1.76	1.76
F_4 mp	1.60	1.60
C_4 ff	1.88	1.88
C_4 mp	1.15	1.11
Speech loud	1.48	1.48
Speech comfortable	1.19	1.19

the piriform sinuses was not completely perpendicular to the transverse imaging plane of the EBCT scanner.

B. Comparing changes in area functions due to variations in register, pitch, and loudness

Changes in 3-D vocal tract configuration as a function of changes in vocal register, pitch, and loudness can be assessed by comparing differences in area functions along the length of the vocal tract. The area functions associated with the vocal tract shape for the /ɑ/ prolongation in the comfortable speech condition are compared to those of the other phonatory conditions in Fig. 1. Beginning at the origin, measured units on the *x* axis represent distance in centimeters above the glottis. On the *y* axis, measured units represent cross-sectional area in square centimeters.

A characteristic common to the vocal tract shapes across all eight phonatory conditions was a widening of the vocal tract airway above the glottis that begins about 2 cm past the glottis, expands to its widest point at about 4 cm past the glottis, and begins to narrow again at about 5 cm past the glottis. For the most part, this is a consequence of the changes in cross-sectional area that occur as the piriform sinuses converge with the main vocal tract tube. This finding concurs with Story *et al.* (1996), who found that the location of this widening was consistent across all vowels. They suggested that this location's uniformity served to point out the consistency of the image analysis procedures in terms of defining the glottal termination. In the current study, the extent of this supraglottal widening varied with phonatory conditions. The absolute cross-sectional area at its widest point was greater (in some cases, significantly so) than those found by Story *et al.* (1996) or in previous research, to which they compared their findings (Baer *et al.*, 1991; Fant, 1960; Yang and Kasuya, 1994). The widest cross-sectional area of the supraglottal widening for the /ɑ/ vowel in these studies were approximately 1.1 cm^2 (Story *et al.*, 1996), 2.2 and 2.9 cm^2 (Baer *et al.*, 1991), 4.1 cm^2 (Fant, 1960), and 1.9 cm^2 (Yang and Kasuya, 1994). In the current study this measure ranged from approximately 3.8 cm^2 for the loud speech condition in chest register to 6.2 cm^2 for the falsetto register, high-pitched, very loud condition. For the speech condition and low-pitch falsetto condition, the supraglottal widening reduced slightly in area as intensity increased. For the medium- and high-pitch falsetto conditions, the supraglottal widening increased in dimension.

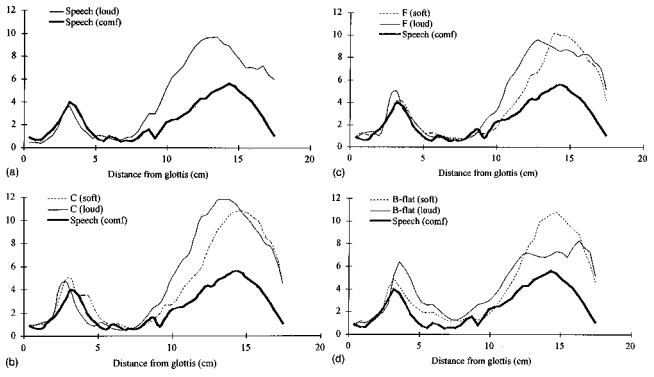


FIG. 1. Comparisons of the area functions for chest register comfortable speech (bold line) and (a) chest register loud speech (narrow line); (b) falsetto register, low pitch (C_4, 262 Hz), moderately soft (dashed line) and very loud (narrow line); (c) falsetto register, medium pitch (F_4 349 Hz), moderately soft (dashed line) and very loud (narrow line); and (d) falsetto register, high pitch (B-flat$_4$, 466 Hz), moderately soft (dashed line) and very loud (narrow line).

In Fig. 1(a), a comparison of comfortable (bold line) and very loud speech (narrow line) in chest register, the vocal tract gesture accompanying the increase in vocal intensity, was an overall increase in oral cavity volume that begins approximately 7 cm from the glottis, just anterior to the vowel constriction in the oropharynx. At its widest point in the oral cavity, the cross-sectional area almost doubled in the very loud speech condition (from about 6 to 10 cm^2) and the mouth opening increased significantly (from about 1 to 6 cm^2). This occurred concurrently with a slight reduction in dimension of the supraglottal widening that occurred about 3.5 cm past the glottis.

The dimensions of the oral cavity also increased significantly in changing from comfortable speech to either soft/loud low-pitch falsetto phonations [Fig. 1(b)]. Within low-pitch falsetto, increased intensity was associated with an increase in the size of the oral cavity and a simultaneous decrease in the size of the supraglottal widening, similar to the pattern of changes in vocal tract configuration with increased loudness for the speech (chest register) conditions. For falsetto phonations sung at the medium- and high-pitch levels [Figs. 1(c) and (d)], the oral cavity is again significantly larger than for comfortable speech. The changes in vocal tract shape from moderately soft (dashed line) to very loud intensity levels (narrow line) at medium and high pitch in falsetto, however, reversed the pattern found in the transition from lower to higher intensity for the speech and low-pitch falsetto conditions. Rather than increasing the volume of the oral cavity, the subject reduced the oral cavity volume, while simultaneously increasing the volume of the supraglottal widening in the lower pharynx. This anterior/posterior shift in relative volumes or cross-sectional areas may be a strategy to balance the need for maintaining approximate vowel quality (relatively stable $F1$ and $F2$) while simultaneously preserving vocal timbre (stable $F3$ and $F4$).

The greatest contrasts in vocal tract configuration due to register, pitch, and loudness were observed in the area functions for comfortable speech phonation in chest register and for high-pitched very loud sung phonation in falsetto register [Fig. 1(d)]. In comfortable speech phonation in chest register

TABLE IV. Formant frequencies associated with variations in vocal register, pitch and loudness levels.

Pitch and loudness	$F1$ (Hz)	$F2$ (Hz)	$F3$ (Hz)	$F4$ (Hz)
Falsetto register				
B-flat$_4$ ff	601	1230	2751	3553
B-flat$_4$ mp	612	1139	2735	3611
F_4 ff	592	1030	2858	3728
F_4 mp	599	1102	2764	3643
C_4 ff	604	939	2729	3833
C_4 mp	582	1062	2738	3850
Chest register				
Speech loud (D_3)	682	1058	2740	3851
Speech comfortable (B-flat$_2$)	543	993	2585	3747

(bold line), the cross-sectional area peaked at 3.7 cm^2 at the widest point of the supraglottal widening, decreased to 1 cm^2 or less for the vowel constriction, increased to 5.8 cm^2 at the widest point in the oral cavity, and then gradually reduced to about 1 cm^2 at the mouth termination. For high-pitched very loud sung phonation in falsetto register (narrow line), the cross-sectional area peaked at 6.4 cm^2 (a 73% increase over moderately soft) at the widest point of the supraglottal widening, decreased to less than 1.5 cm^2 (a 50% increase) for the vowel constriction, increased to 8.3 cm^2 (a 43% increase) at the widest point in the oral cavity, then closed down to an area of approximately 5 cm^2 (a 50% increase) at the mouth termination.

C. Formant structure

The first four formants for each of the eight vocal tract shapes are summarized in Table IV. The change from soft to loud intensity in chest register speech phonations resulted in an increase in all formant frequencies, and was associated with increases in mouth opening (Pickett, 1999). As such, the subject followed the previously reported tendency to raise $F1$ when increasing vocal intensity (Huber *et al.*, 1999; Sundberg *et al.*, 1993): the $F1$ increased from 583 to 642 Hz.

Changes from comfortable speech in chest register to sung phonations in falsetto followed a similar pattern: increases in mouth opening yielded higher formant frequencies overall. For the high-pitched, very loud sung falsetto tone, $F1$ (601 Hz) and $F2$ (1230 Hz) values approached a neutral vowel (Pickett, 1999), not unexpected for a vocal tract configuration with such a large supraglottal widening, oral cavity volume, and mouth opening.

In general, the subject's vowel formants tended towards those of a phonetically more neutral vowel, when compared to norms for spoken /ɑ/, especially in the sung phonations. $F1$ values for the subject ranged from 543 to 682 Hz, and $F2$ ranged from 939 to 1230 Hz. Although the vowel formants ($F1$,$F2$) produced by the subject were below reported means for the spoken /ɑ/ vowel, they fall within the range of values produced by speakers in normative studies for spoken /ɑ/, /ɔ/, or /ʌ/ (Hillenbrand *et al.*, 1995; Peterson and Barney, 1952). This may be part of the subject's vocal technique to retain a perceptually darker vowel quality and warmer vocal timbre, even as effort is increased, by using a phonetically more rounded and/or centralized allophone (Titze, 1994). The tendency for untrained speakers/singers is to raise the larynx with increased effort, and to thereby shorten the vocal tract, increase formant frequencies, and produce a more strident vocal timbre.

A singer's formant in the 2800–3000-Hz region was not observed in any of the phonatory conditions. This finding was not unexpected, since the phenomenon is primarily associated with classically trained male singers phonating in chest register (Sundberg, 1987), and not with countertenor falsetto-based singing.

IV. CONCLUSION

Volumetric imaging of the vocal tract using EBCT was used to document 3-D changes in vocal tract configuration during phonation, which occurred as a function of vocal register (chest or falsetto), pitch (low, medium, and high in falsetto register, and speech), and loudness (soft versus very loud) for a single male subject. The 3-D images are of higher resolution and corresponding measures of greater accuracy than those acquired in previous studies using MRI techniques. EBCT techniques (using 3 mm slices) allowed for greater image resolution (0.410 mm in the plane of imaging), reduced image acquisition time, and, thus, minimization of movement artifacts.

Although the results of the entire study are from one subject, a trained countertenor, the data may not be atypical of other male singers or speakers, who produce falsetto similarly. For the population of singers in the classical, commercial, and world music arenas, and theatrical and voice-over artists who utilize falsetto phonation more often and more intensively than many speakers, additional research towards a comprehensive characterization of the physiology of falsetto phonation is vital. The current data, elicited from a subject who has developed an optimal technique for vocalizing in the falsetto register across a range of frequencies and intensity levels provides insight into a falsetto type that has been shown to be reliable and associated with low risk for injury to the tissues of the vocal folds.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Dr. William Stanford, Dr. Brad Thompson, and the members of the Division of Physiologic Imaging, Department of Radiology at the University of Iowa College of Medicine for their contributions to this study. This research was supported in part by research grant No. P60 DC 00976 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

Alwan, A., Narayanan, S., and Haker, K. (**1997**). ''Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data, Part II. The rhotics,'' J. Acoust. Soc. Am. **101**, 1078–1089.

Baer, T., Gore, J. C., Gracco, L. C., and Nye, P. W. (**1991**). ''Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels,'' J. Acoust. Soc. Am. **90**, 799–828.

Boyd, D. P., and Lipton, J. J. (**1983**). ''Cardiac computed tomography,'' Proc. IEEE **71**, 298–307.
Colton, R. H., and Hollien, H. (**1972**). ''Phonational range in the modal and falsetto registers,'' J. Speech Hear. Res. **15**, 708–713.
Fant, G. (**1960**). *The Acoustic Theory of Speech Production* (Moulton, The Hague).
Giles, P. (**1982**). *The Countertenor* (Muller, London).
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.
Hoffman, E. A., and Gefter, W. B. (**1990**). ''Multimodality imaging of the upper airway: MRI, MR spectroscopy, and ultrafast x-ray CT,'' in *Sleep and Respiration*, edited by F. G. Issa, P. M. Suratt, and J. E. Remmers (Wiley-Liss, New York), pp. 291–301.
Hoffman, E. A., Sinak, L. J., Robb, R. A., and Ritman, E. L. (**1983**). ''Noninvasive quantitative imaging of shape and volume of lungs,'' Am. Physiol. Soc. 1414–1421.
Hoffman, E. A., Gnanaprakasam, D., Gupta, K. B., Hoford, J. D., Kugelmass, S. D., and Kulawiec, R. S. (**1992**). ''VIDA: an environment for multidimensional image display and analysis,'' SPIE Proc. Biomed. Image Proc. and 3-D Microscopy, Vol. 1660, San Jose, CA, 10–13 February.
Hollien, H. (**1974**). ''On vocal registers,'' J. Phonetics **2**, 125–143.
Hollien, H. (**1977**). ''The registers and ranges of the voice,'' in *Approaches to Vocal Rehabilitation*, edited by M. Cooper and H. C. Cooper (Thomas, Springfield, IL), pp. 76–121.
Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., and Johnson, K. (**1999**). ''Formants of children, women and men: The effects of vocal intensity variation,'' J. Acoust. Soc. Am. **106**, 1532–1542.
International Commission on Radiological Protection. (**1977**). *Recommendations of the International Commission on Radiological Protection*, ICRP Publication 26 (Pergamon, Oxford).
Kelly, J. L., and Lochbaum, C. C. (**1962**). ''Speech synthesis,'' in Proceedings 4th International Congress on Acoustics, paper 642, pp. 1–4.
Laver, J. (**1980**). *The Phonetic Description of Voice Quality* (Cambridge U. P., New York).
Lienard, J. S., and Di Benedetto, M. G. (**1999**). ''Effect of vocal effort on spectral properties of vowels,'' J. Acoust. Soc. Am. **106**, 411–422.
Liljencrants, J. (**1985**). ''Speech Synthesis with a Reflection-Type Line Analog,'' DS dissertation, Department of Speech Communication and Music Acoustics, Royal Institute of Tech., Stockholm, Sweden.
Malm, W. P. (**1967**). *Music Cultures of the Pacific, the Near East and Asia* (Prentice–Hall, Englewood Cliffs, NJ).
Markel, J. D., and Gray, A. H. (**1976**). *Linear Prediction of Speech* (Springer-Verlag, New York).
Moore, C. A. (**1992**). ''The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images,'' J. Speech Hear. Res. **35**, 1009–1023.
Narayanan, S. S., Alwan, A. A., and Haker, K. (**1997**). ''Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals,'' J. Acoust. Soc. Am. **101**, 1064–1077.
National Council on Radiation Protection and Measurements. (**1987**). *Recommendations on Limits for Exposure to Ionizing Radiation*, NCRP Report No. 91 (National Council on Radiation Protection and Measurements, Bethesda, MD).
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Pickett, J. M. (**1999**). *The Acoustics of Speech Production* (Allyn and Bacon, Boston).
Raya, S. P., and Udupa, J. K. (**1990**). ''Shape-based interpolation of multidimensional objects,'' IEEE Trans. Med. Imaging **9**, 32–42.
Story, B. H. (**1995**). ''Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract,'' Ph.D. dissertation, University of Iowa.
Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). ''Vocal tract area functions from magnetic resonance imaging,'' J. Acoust. Soc. Am. **100**, 537–554.
Sulter, A. M., Miller, D. G., Wolf, R. F., Schutte, H. K., Wit, H. P., and Mooyaart, E. L. (**1992**). ''On the relation between the dimensions and resonance characteristics of the vocal tract: a study with MRI,'' Magn. Reson. Imaging **10**, 365–373.
Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois U. P., DeKalb, IL).
Sundberg, J., Titze, I. R., and Scherer, R. C. (**1993**). ''Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source,'' J. Voice **7**, 15–29.
Titze, I. R. (**1988**). ''A framework for the study of vocal registers,'' J. Voice **2**, 183–194.
Titze, I. R. (**1994**). Principles of voice production (Prentice–Hall, Englewood Cliffs, NJ).
Titze, I. R., Mapes, S., and Story, B. (**1994**). ''Acoustics of the tenor high voice,'' J. Acoust. Soc. Am. **95**, 1133–1142.
Tom, K. (**1996**). ''Intensity control in male falsetto phonation: An analysis by synthesis approach,'' Ph.D. dissertation, University of Iowa.
Udupa, J. K. (**1991**). ''Computer aspects of 3-D imaging in medicine: A tutorial,'' in *3D Imaging in Medicine*, edited by J. K. Udupa and G. T. Herman (CRC, Boca Raton, FL).
Welch, G. F., Sergeant, D. C., and MacCurtain, F. (**1988**). ''Some physical characteristics of the male falsetto voice,'' J. Voice **2**, 151–163.
Yang, C-S., and Kasuya, H. (**1994**). ''Accurate measurement of vocal tract shapes from magnetic resonance images of child, female, and male subjects,'' Proc. ICSLP **94**, 623–626, Yokohama, Japan.

Effects of consonant environment on vowel formant patterns

James M. Hillenbrand[a] and Michael J. Clark
Speech Pathology and Audiology, Western Michigan University, Kalamazoo, Michigan 49008

Terrance M. Nearey
Department of Linguistics, University of Alberta, Edmonton, Alberta T6G 2E7, Canada

(Received 21 June 2000; accepted for publication 7 November 2000)

A significant body of evidence has accumulated indicating that vowel identification is influenced by spectral change patterns. For example, a large-scale study of vowel formant patterns showed substantial improvements in category separability when a pattern classifier was trained on multiple samples of the formant pattern rather than a single sample at steady state [J. Hillenbrand *et al.*, J. Acoust. Soc. Am. **97**, 3099–3111 (1995)]. However, in the earlier study all utterances were recorded in a constant /hVd/ environment. The main purpose of the present study was to determine whether a close relationship between vowel identity and spectral change patterns is maintained when the consonant environment is allowed to vary. Recordings were made of six men and six women producing eight vowels (/i,ɪ,ɛ,æ,ɑ,ʊ,u,ʌ/) in isolation and in CVC syllables. The CVC utterances consisted of all combinations of seven initial consonants (/h,b,d,g,p,t,k/) and six final consonants (/b,d,g,p,t,k/). Formant frequencies for F_1-F_3 were measured every 5 ms during the vowel using an interactive editing tool. Results showed highly significant effects of phonetic environment. As with an earlier study of this type, particularly large shifts in formant patterns were seen for rounded vowels in alveolar environments [K. Stevens and A. House, J. Speech Hear. Res. **6**, 111–128 (1963)]. Despite these context effects, substantial improvements in category separability were observed when a pattern classifier incorporated spectral change information. Modeling work showed that many aspects of listener behavior could be accounted for by a fairly simple pattern classifier incorporating $F0$, duration, and two discrete samples of the formant pattern. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1337959]

PACS numbers: 43.70.Bk, 43.71.An, 43.72.Ne, 43.70.Fq [KRK]

I. INTRODUCTION

A major focus of recent vowel perception research has been an examination of the relationship between formant-frequency movements and vowel identity. A good deal of evidence has accumulated implicating a secondary but quite important role for spectral change in vowel recognition. Reviews of this work can be found in Nearey (1989) and Strange (1989). Briefly, the evidence favoring this view includes the work of Strange, Jenkins, and Johnson (1983) and Nearey and Assmann (1986) showing high identification rates for ''silent-center'' stimuli in which vowel centers were gated out, leaving only brief onglides and offglides. Nearey and Assmann also reported a sharp decrease in identification rates for silent center signals in which onglides and offglides were played in reverse order (see also Jenkins, Strange, and Edman, 1983; Parker and Diehl, 1984; Andruski and Nearey, 1992; Jenkins and Strange, 1999). Further, several studies have reported relatively high identification error rates for both natural and synthetic vowels with static formant patterns (Fairbanks and Grubb, 1961; Hillenbrand and Gayvert, 1993a; Hillenbrand and Nearey, 1999). For example, Hillenbrand and Nearey asked listeners to identify naturally produced /hVd/ syllables and two different formant-synthesized versions. An ''original formant'' (OF) set of synthetic signals was generated using the measured formant contours, and a second set of ''flat formant'' (FF) signals was synthesized with formant frequencies fixed at the values measured at the steadiest portion of the vowel. The OF synthetic signals were identified with substantially greater accuracy than the FF signals. Finally, a number of pattern recognition studies have reported better classification accuracy and/or improved prediction of listener error patterns for pattern recognition models that incorporate spectral change as opposed to models that are driven by spectral measurements sampled at a single cross section of the vowel (Assmann, Nearey, and Hogan, 1982; Nearey and Assmann, 1986; Parker and Diehl, 1984; Zahorian and Jagharghi, 1993; Hillenbrand *et al.*, 1995). For example, Hillenbrand *et al.* trained a discriminant classifier on various combinations of fundamental frequency and formant measurements from /hVd/ syllables spoken by 45 men, 48 women, and 46 children. The pattern classifier was substantially more accurate when it was trained on two samples of the formant pattern (taken at 20% and 80% of vowel duration) than a single sample taken at the steadiest portion of the vowel.

An important limitation of the work conducted on this problem to date is the exclusive reliance on either isolated vowels or /hVd/ syllables. It is firmly established that vowel formant patterns are affected not only by the identity of the vowel, but also by consonant environment. In a classic study, Stevens and House (1963) reported formant measurements for eight vowels (/i,ɪ,ɛ,æ,ɑ,ʌ,ʊ,u/) spoken by three men. The vowels were produced in isolation, in /hVd/ syllables, and in

[a]Electronic mail: james.hillenbrand@wmich.edu

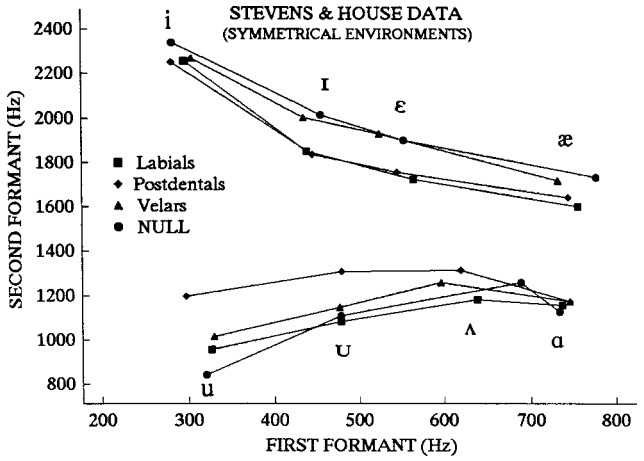


FIG. 1. Stevens and House (1963) data showing the effects of consonant environment on steady-state vowel formants.

symmetrical CVC syllables with 14 consonants (/p,b,f,v,θ,ð,s,z,t,d,ʧ,ʤ,k,g/). Effects of consonant context were examined by comparing the formant values in these 14 environments to formant values for the same vowels in isolation or /hVd/ context, which the authors referred to as ''null'' environments. Formant frequencies and bandwidths were measured for F_1-F_3 at the center of the vowel using a spectrum-matching technique. The most general summary of the Stevens and House findings is that the non-null consonant environments typically had the effect of shifting the formant frequencies—particularly F_2—toward more centralized values. Systematic effects were seen for the manner, voicing, and place of articulation of the flanking consonants. The place effects, which were easily the most important, are reproduced in Fig. 1. The effects of place on F_1 values tended to be small and rather consistent in magnitude from one vowel to the next. Place effects on F_2, on the other hand, were sometimes quite large and varied considerably in magnitude from one vowel to the next. The largest effect by far was an upward shift averaging about 350 Hz in F_2 for /u/ in the environment of postdental consonants; a shift averaging about 200 Hz was also seen for /ʊ/ in the environment of postdentals. There were also downward shifts in F_2 of some 100–200 Hz for front vowels (with the exception of /i/) in the environment of labial and postdental consonants. The effects of manner class and voicing were typically rather small. Vowels flanked by voiced consonants tended to be produced with slightly lower F_1 values as compared to the same vowels in the context of unvoiced consonants. Manner class had little effect on F_1 values, but vowels in stop consonant environments tended to have slightly higher F_2 values.

Stevens and House interpreted these varied findings in terms of a production undershoot model. The production system was assumed to be driven by targets corresponding to articulatory postures in null environments, but these idealized targets were purportedly not realized due to inertial constraints. Stevens and House (1963) also suggested that listeners make tacit use of knowledge of these context effects in recognizing vowels: ''The rules governing these deviations in the acoustic signal must, of course, be invoked in some way by the listener in order to make an identification of the signal'' (p. 122).

A problem that is presented by the findings discussed above is that there are clearly multiple influences on the detailed formant contours of even relatively simple citation-form CVC utterances. The two influences that are of interest in the present study are the consonant context effects described above and the ''vowel inherent spectral change'' patterns that have been observed in studies such as Nearey and Assmann (1986) and Hillenbrand *et al.* (1995) using isolated vowels or /hVd/ syllables. The primary question that is to be addressed is whether context effects such as those described by Stevens and House act to obscure or complicate the relationships between vowel identity and spectral change patterns that have been observed in previous studies using neutral contexts. Some preliminary evidence on this question comes from Zahorian and Jagharghi's (1993) study of 11 vowels in CVC context with 7 initial consonants and 6 final consonants. Zahorian and Jagharghi reported better pattern classification accuracy for feature sets incorporating spectral change than for static feature sets. However, no acoustic measurements were made of the coarticulatory patterns, making it impossible to relate either the pattern classification results or their listener data to specific context-conditioned effects.

The present study consisted of a replication and extension of Stevens and House, but with several differences in method. The most important of these were: (1) since consonant context effects are nearly certain to be more complex in the nonsymmetrical environments that typically prevail in natural speech, CVCs were recorded in both symmetrical and nonsymmetrical environments, and (2) since we were interested in studying the spectral change patterns for vowels, full format contours were measured rather than sampling the formant pattern once at steady state.

II. METHODS

A. Test signals

Talkers consisted of six men and six women between the ages of 25 and 64. Seven of the speakers were raised in Michigan; the others were from northern Illinois (2), upstate New York (1), Nebraska (1), and northern Ohio (1). All of the speakers were phonetically trained. The speech material consisted of isolated vowels and CVC syllables, only a subset of which was analyzed for the present study. The initial consonants consisted of /h,b,d,g,p,t,k,r,l,w/, the vowels consisted of /i,ɪ,ɛ,æ,ɑ,ʌ,ɔ,ʊ,u,ɝ/), and the final consonants consisted of /b,d,g,p,t,k,r,l/. The initial consonants, vowels, and final consonants were recorded in all combinations. Each of the ten vowels was also recorded in isolation, for a total of 9516 utterances (12 talkers×10 initial consonants×10 vowels×8 final consonants+10 isolated vowels, less 17 unpronounceable combinations, such as /rɝr/). For purely practical reasons, a subset of these recordings was selected for use in the present experiment. Selected for analysis were the eight vowels studied by Stevens and House (/i,ɪ,ɛ,æ,ɑ,ʌ,ʊ,u/) in isolation and in combination with seven initial consonants (/h,b,d,g,p,t,k/) and six final consonants (/b,d,g,p,t,k/).

Recordings were made in a small sound-attenuated booth using a Shure SM58 dynamic microphone. Signals were preamplified, low-pass filtered at 4.3 kHz, and directly digitized at a 10-kHz sample rate using a Tucker & Davis 16-bit A/D. Subjects read from word lists containing the phonetic transcriptions of the utterances to be read. The word lists were blocked by vowel and proceeded in an orderly way through the consonant environments. Subjects were given no special instructions regarding duration or intonation contour, except that they were urged try to avoid a drop in pitch at the end of each page of transcriptions. Individual syllables were later excised from the longer recordings. The signals were auditioned at this time, and if the experimenter noticed an obvious production error, the talker was brought back for another session to re-record the syllables that had been mispronounced. In a few cases the talker was no longer available, and the mispronounced utterances were simply deleted. Twenty-three utterances were deleted in this way, leaving a total of 4105 utterances.

B. Acoustic measurements

The formant estimation methods were similar to those described in Hillenbrand *et al.* (1995). Formant analysis began with the extraction of peaks from 14-pole, 128-point linear predictive coding (LPC) spectra every 5 ms using 25.6-ms Hamming-windowed segments. A graphical display of the spectral peaks was then overlaid on a gray-scale LPC spectrogram. Formant tracks for F_1-F_3 were determined by hand editing the spectral peaks, deleting spurious peaks in some cases, and interpolating through ''holes'' in the formant track in other cases. The number of LPC poles was occasionally increased to separate merged formants. In some cases—many of them involving formant mergers—it was judged that a formant could not be measured with confidence. In these cases, zeros were written into the formant slot, and the values for that formant were simply omitted from all subsequent analyses. Formants were edited only between the starting and ending times of the vowel, which were determined by visual inspection of the LPC spectrograms. Measures of vowel duration included the vocalic segment only and not the initial burst associated with consonant release.

Vowel ''steady-state'' times were determined automatically. We experimented with a number of algorithms and settled on a simple technique that seemed to show the best agreement with the visual inspection method that has been used in many previous studies. The vowel formant contour was reduced to an array of $\log F_2 - \log F_1$ values (Miller, 1989). The sum of differences between adjacent frames was then calculated for every sequence of five frames (35 ms) throughout the first 60% of the vowel.[1] Steady-state time was defined as the middle of the sequence of five frames showing the smallest absolute summed difference.

Fundamental-frequency contours were measured using a conventional autocorrelation pitch tracker (Hillenbrand, 1988), followed by hand editing using the tool described above. If there was any uncertainty about the F_0 contour, the experimenter examined the time waveform and a narrow band spectrogram.

C. Listening test

The test signals were presented for identification to 24 phonetically trained listeners. The listeners were first- and second-year graduate students in speech-language pathology. The listeners spoke the same Northern Cities dialect as the speakers, with roughly 80% of the listeners from Michigan, and the remainder from other areas of the upper Midwest, such as the northern parts of Indiana, Illinois, and Ohio. Subjects were tested individually in a quiet room in four sessions of about an hour each. Stimuli were scaled to maximum peak amplitude, low-pass filtered at 4.3 kHz at the output of a 16-bit D/A converter, amplified, and delivered to a single loudspeaker positioned about 1 m from the listener's head at an average intensity of approximately 77 dBA. Over the course of the four sessions, each listener identified one presentation of each of the 4105 signals. The order of presentation was fully randomized (i.e., not blocked by talker or context), and the presentation order was shuffled separately for each listener. Listeners responded by pressing one of eight keys on a computer keyboard that had been labeled with phonetic symbols for the vowels. The listening test was self-paced, and subjects could repeat a stimulus as many times as they wished before entering a response. Each listening test was preceded by a brief practice session to ensure that listeners understood the task and interpreted the phonetic symbols appropriately.

III. RESULTS

A. Listening test

The average identification rate for the test signals was 94.6%, with nearly identical rates for the male (94.5%) and female (94.8%) talker groups. The majority (61.7%) of the individual tokens were correctly identified by all 24 listeners, and 86% of the signals were identified as the intended vowel by at least 90% of the listeners. For 78 signals (1.9%) the label assigned by a plurality of the panel was a vowel other than that intended by the talker. The most common of these misidentifications consisted of tokens that were intended as /æ/ but heard as /ɛ/ (40% of the signals misidentified by a plurality of the listeners) and tokens that were intended as /ɛ/ but heard as /ɪ/ or /æ/ (23% of the signals misidentified by a plurality of the listeners). Average intelligibility for individual talkers varied from 88.7% to 98.0% (s.d.$=2.6$).

As seen in Fig. 2, intelligibility was higher for some vowels than others. As we will show below, the differences in identification rates across vowels are highly significant. Despite this, however, it is important to note that all individual vowels were well identified. Average rates varied from about 90% to 99%, with a standard deviation across vowels of only 3.4%.

Vowel intelligibility was also affected somewhat by consonant environment. Figure 3 shows identification rates, averaged across all vowels, as a function of both initial consonant and final consonant. (Labeling data for the isolated vowels are not shown in the figure.) It can be seen that the consonant environment effects are quite small in magnitude, with only 3.3% separating the most intelligible contexts from the least intelligible. The standard deviation computed across

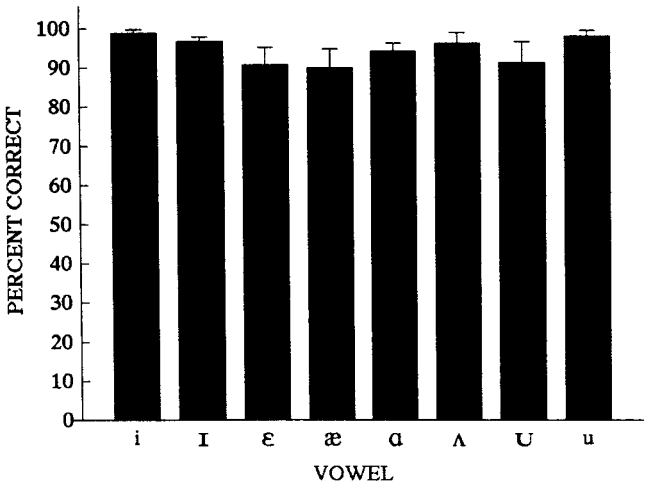


FIG. 2. Vowel intelligibility averaged across all consonant environments. Error bars show one standard deviation.

the seven average identification rates for initial consonants (i.e., the black bars in Fig. 3) is only 0.8%, and the standard deviation computed across the six average identification rates for final consonants (the open bars in Fig. 3) is only 1.3%.

1. Statistical treatment of labeling data

The two sources of individual variability in the perceptual responses that are reasonably viewed as random are speakers and listeners. Accordingly, repeated measures analyses of variance (ANOVAs) were run by listener (i.e., pooling syllable scores over talkers and treating listeners as random effect) and by talker (i.e., pooling syllable scores over listeners and treating talker as a random effect).[2] In each case, Studebaker's (1985) rationalized arcsine transformation was applied to the percent-correct values after pooling. Following a practice common in the psycholinguistics literature, we will report F ratios separately by listener (F_L) and by talker (F_T). We will also report the F'_{min} (Clark, 1973), which we will use to determine significance levels. In both cases, there are five treatment factors: initial consonant place, initial consonant voicing, vowel, final consonant place, and final consonant voicing.

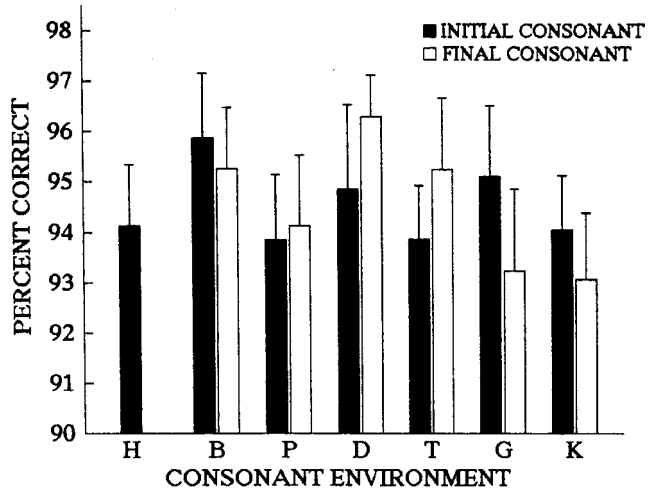


FIG. 3. Percent-correct vowel identification for each of seven initial consonant environments and six final consonant environments.

The usual procedures for *post hoc* comparison of means are clumsy in higher-order factorial analyses of variance. Consequently, we have chosen to follow up significant F tests with t tests derived from effect estimates in the linear model underlying the ANOVA. A significant t value for a specific coefficient of a main or interaction effect indicates that the coefficient in question was significantly different from the average of the entire family of coefficients for that effect. A significant t value can be also be interpreted as indicating that the cell mean associated with the coefficient in question is significantly different from the grand mean and any main effect's lower-order interaction terms. Significance levels of t tests are estimated using the Sidak approach to multiple comparisons with a family size equal to the number of effect coefficients for the main effect or interaction term in question.

The main effect for vowel was highly significant [$F_L(7,161)=30.1$, $F_T(7,77)=3.9$, $F'_{min}(7,97)=3.4$, $p<0.01$]. It is useful to have an indication of the relative contribution of phonetic factors. We will use *percent of total phonetic variation accounted for*, defined as the ratio of the sum of squares for a given main or interaction effect to the sum of squares of all phonetic factors. We base this measure on the *by-talkers* analysis. The vowel main effect was easily the most important of any of the phonetic effects that were observed, accounting for 41.2% of total phonetic variation (TPV). Sidak-corrected planned comparisons revealed that /u/ and /i/ were identified significantly better than average: /u/ by 1.1 percentage points [$t'_{min}(16)=4$, $p_{min}<0.01$], and /i/ by 1.2 percentage points [$t'_{min}(16)=7.1$, $p_{min}<0.000\,01$]. The main effect for final place of articulation was also highly significant [$F_L(2,46)=79.7$, $F_T(2,22)=12.2$, $F'_{min}(2,29)=10.6$, $p<0.001$] and accounted for about 5.7% of TPV. Sidak-corrected planned comparisons showed that vowels in the environment of final velars were identified less well than average by about 0.9 percentage points [$t'_{min}(14)=-3.4$, $p_{min}<0.01$], while those in final alveolar contexts were identified about 0.6 percentage points better than average [$t'_{min}(14)=4.0$, $p_{min}<0.001$]. No other main effects were significant.

Although they were all small in absolute magnitude, several interactions reached significance. The interaction pattern is displayed in Fig. 4, which shows percent correct as a function of the initial consonant, with the final consonant as the parameter. The vowel by final place interaction was highly significant [$F_L(14,322)=20.4$, $F_T(14,154)=4.4$, $F'_{min}(14,222)=3.6$, $p<0.0001$], accounting for about 9.6% of TPV. Sidak-corrected planned comparisons showed that syllables ending in /æ/+labial stops were identified about 2.1 percentage points better than average [$t'_{min}(18)=3.9$, $p_{min}<0.05$]. The vowel by final voicing interaction was also significant [$F_L(7,161)=36.8$, $F_T(7,77)=2.9$, $F'_{min}(7,89)=2.7$, $p<0.01$], accounting for about 4.9% of TPV. However, Sidak-adjusted comparisons failed to identify any specific effect as significantly above average.

The three-way initial voicing by initial place by vowel

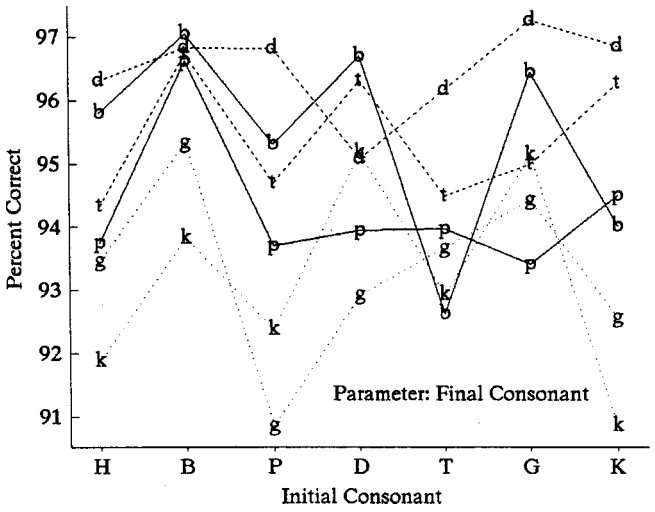


FIG. 4. Vowel intelligibility as a function of the initial consonant and final consonant.

interaction was highly significant [$F_L(14,322)=14.2$, $F_T(14,154)=3.8$, $F'_{min}(14,240)=3.0$, $p<0.001$], accounting for about 5.9% of TPV. The three-way interaction of initial voicing by final voicing by final place [$F_L(2,46)=21.6$, $F_T(2,22)=5.0$, $F'_{min}(2,32)=4.0$, $p<0.05$] and the four-way interaction of initial place by initial voicing by final place by final voicing [$F_L(4,92)=18.4$, $F_T(4,44)=3.4$, $F'_{min}(4,61)=2.8$, $p<0.05$] were also significant. These interactions accounted for less than 1% of the TPV. In no case for three- or four-way interactions did Sidak-adjusted comparisons identify specific factor combinations as significantly different from average for the family in question.

An important point which we hope does not get lost in the details of the ANOVA results reported above is that the influences of consonant environment on average vowel recognition rates are rather small in absolute magnitude. For example, as can be seen in Fig. 4, the full range of variation separating the most intelligible from the least intelligible contexts is only about 6%, and the standard deviation in average recognition rates computed over all 42 phonetic environments displayed in Fig. 4 (e.g., /hVb/, /hVd/, /hVp/,..., /kVk/) is a very modest 1.7%.

A final note on the listening test results concerns labeling data for /u/ and /ʊ/ in alveolar contexts. Recall that the largest context effects observed by Stevens and House (1963) consisted of a raising of F_2 for /u/ and /ʊ/ in the environment of alveolar consonants relative to the same vowels in null environments (Fig. 1). As will be discussed below, our formant measurements showed an even larger effect, averaging about 500–600 Hz for /u/ and about 200–300 Hz for /ʊ/. Our results further showed that the effect is conditioned primarily by the presence of a *syllable-initial* alveolar. It was therefore of some interest to determine whether there is any evidence that these context-conditioned shifts in formant values have an adverse effect on vowel intelligibility. The recognition rates for /u/ and /ʊ/ in the environment of syllable-initial alveolars turn out to be unremarkable. Average recognition rates, pooled across all contexts except initial alveolar, are 98.1% for /u/ and 91.5% for /ʊ/. These figures compare with nearly identical rates in initial alveolar contexts of 98.3% for /u/ and 92.5% for /ʊ/. There is, in short, no evidence that the largest of the context-conditioned shifts in vowel formants caused any difficulty for the listeners.

B. Acoustic measurements

1. Vowel duration

Vowel durations in various consonant voicing environments are shown in Table I. To simplify the examination of voicing effects, initial /h/ environments were excluded from the computation of the means reported in the first four columns of the table. The average durations associated with the eight vowels, pooled across all consonant environments, are strongly correlated with average durations from the /hVd/ data of Hillenbrand *et al.* (1995), and with the /tVp/ data of Black (1949). The widely observed increase in duration for vowels preceding voiced versus unvoiced stops is quite evident in our data (i.e., compare V–V with V–U and U–V with U–U). Also apparent in Table I is evidence for systematically longer vowels when *preceded by* voiced versus unvoiced stops (i.e., compare V–V with U–V and V–U with U–U in Table I). This effect, which averages some 20–40 ms, was confirmed by two separate ANOVAs for vowel and initial consonant voicing, one comparing durations for V–V environments with U–V environments (i.e., C_1=voiced/C_2=voiced versus C_1=unvoiced/C_2=voiced) and the second

TABLE I. Vowel durations in ms in different stop-consonant voicing environments, and in all consonant environments. Measurements for isolated vowels were excluded. Standard deviations shown in parentheses. (V–V=voiced initial consonant, voiced final consonant; U–V=unvoiced initial consonant, voiced final consonant; V–U=voiced initial consonant, unvoiced final consonant; U–U=unvoiced initial consonant, unvoiced final consonant.)

Vowel	V–V	U–V	V–U	U–U	All consonant environments
/i/	255.9 (46.8)	233.6 (48.3)	169.8 (32.8)	144.1 (34.1)	198.7 (61.3)
/ɪ/	190.6 (29.3)	174.2 (30.7)	137.3 (32.6)	116.7 (28.1)	153.1 (41.7)
/ɛ/	218.2 (28.3)	191.1 (29.8)	160.3 (31.6)	127.8 (27.9)	176.1 (44.7)
/æ/	331.8 (50.5)	286.0 (43.1)	254.0 (50.5)	214.0 (40.6)	266.6 (65.6)
/ɑ/	328.9 (57.3)	290.3 (41.9)	235.7 (50.1)	194.2 (45.2)	255.9 (73.8)
/ʌ/	215.0 (35.4)	178.8 (30.2)	146.6 (30.9)	118.6 (24.3)	162.9 (49.2)
/ʊ/	208.8 (35.2)	189.7 (31.2)	152.9 (36.0)	124.7 (31.9)	166.2 (46.6)
/u/	261.2 (46.5)	241.9 (44.5)	171.7 (37.1)	147.3 (31.3)	203.6 (66.6)
All vowels	251.8 (65.7)	223.5 (58.8)	177.2 (54.2)	146.3 (46.3)	198.0 (69.3)

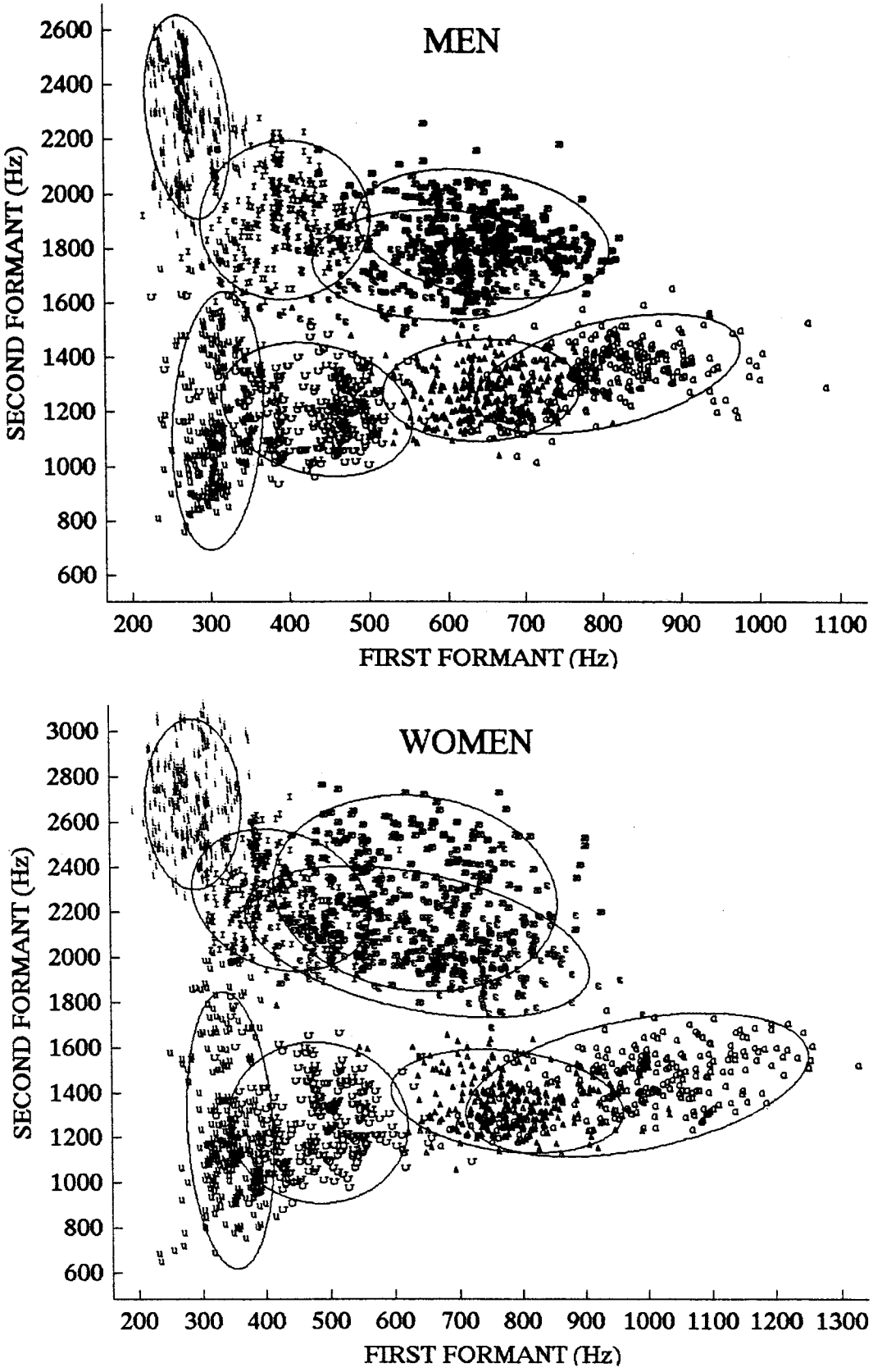


FIG. 5. Values of F_1 and F_2 measured at steady state for men (top panel) and women (bottom panel). Formant measurements are plotted for isolated vowels and for all consonant environments. Not displayed are measurements for signals with identification error rates of 15% or higher. To improve the clarity of the display the database was thinned by removing redundant data points, resulting in the display of measurements from about two-thirds of the well-identified tokens.

comparing V–U with U–U environments. Both ANOVAs showed highly significant effects for vowel and initial consonant voicing. This effect is consistent with Fischer-Jorgensen (1964) and Crystal and House (1988), but differs from the conclusions reached by Peterson and Lehiste (1960).

2. Average formant values

Values of F_1 and F_2 measured at steady state are displayed in Fig. 5. Formant measurements are plotted for isolated vowels and for all consonant environments. Not displayed in Fig. 5 are measurements for signals with identification error rates of 15% or higher. To improve the clarity of the display the database was thinned by removing redundant data points, resulting in the display of measurements from about two-thirds of the well-identified tokens. There is, of course, considerable variability in formant values within each vowel category, and a good deal of overlap among vowels.[3] A major goal of the present study was to determine what aspects of this variability were associated with consonant environment.

3. Effects of place of articulation

Figure 6 shows the effects of place of articulation on the frequencies of F_1 and F_2 for syllables that are symmetrical with respect to place of production (e.g., /bVb/, /bVp/, /pVp/, /pVb, /dVd/, /dVt/, etc.). Also plotted are formant values for isolated vowels and /hVd/ syllables, environments referred to as ''null'' by Stevens and House (1963). The general look of this figure is similar to the Stevens and House (SH) data (Fig. 1), which were based on strictly symmetrical syllables (/bVb/, /dVd/, /gVg/, etc.). As with SH, the largest effect by far is a raising of F_2 for /u/ in the environment of alveolar consonants. At about 500 Hz for the men and nearly 600 Hz for the women (relative to null environments), this upward shift is even larger than the roughly 350 Hz effect reported by SH.[4] Sizable upward shifts in F_2 for alveolar environments are seen for the remaining back/central vowels, especially /ʊ/, with shifts averaging 214 Hz for men and 281 Hz for women. Also seen for the back/central vowels was a fairly uniform upward shift in F_2 averaging 98 Hz for men and 117 Hz for women for the velar environments. For front vowels, the most consistent effect is a downward shift in F_2 of some 85–100 Hz for labial environments. As in the SH results, the effects of place on F_1 values tend to be rather small. The only moderately sizable effect that appears to be consistent across men and women is a downward shift in F_1 averaging some 50 Hz for /ɛ/ and /æ/ in the environment of alveolar and velar consonants.

The effects of place on formant values for syllables that are either symmetrical or nonsymmetrical with respect to place of articulation are shown in Figs. 7 and 8. Figure 7 shows the effects of initial consonant environment, while Fig. 8 shows the effects of final consonant environment. In Fig. 7, showing the effects of initial consonant place, averages that are plotted with the square symbols for labials, for example, were pooled over all syllables with C_1=/b,p/, regardless of the final consonant. Similarly, in Fig. 8, the

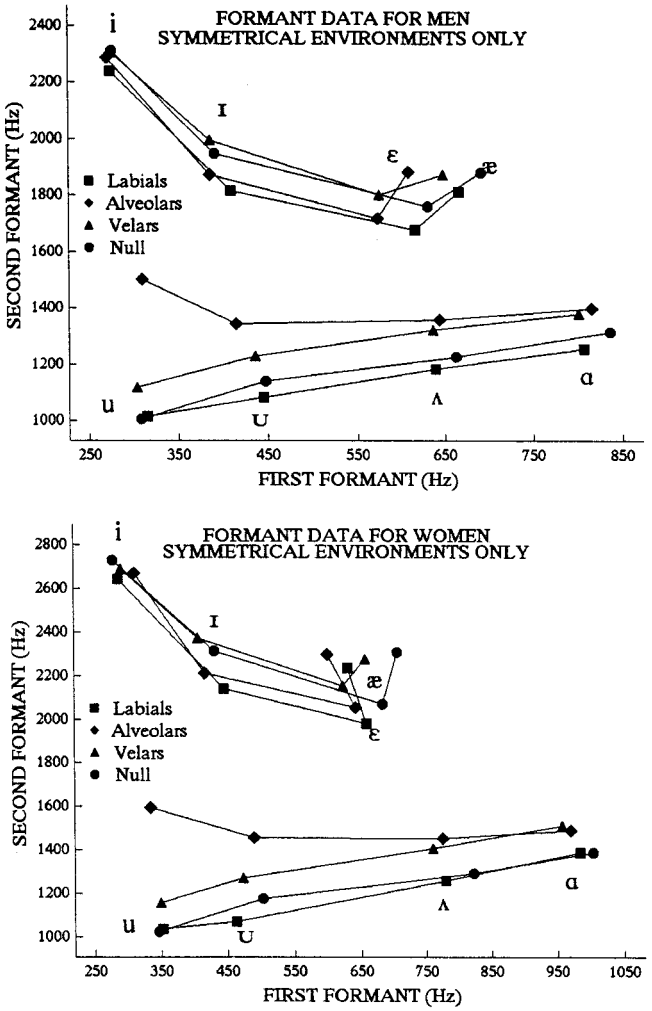


FIG. 6. Average formant frequencies at steady state as a function of the place of articulation of the surrounding consonants for men (top panel) and women (bottom panel). Data are shown for symmetrical environments only.

square symbols represent averages pooled over all syllables with C_2=/b,p/, regardless of the initial consonant. Although there are many minor differences, the general look of Fig. 7 (initial environments) is quite similar to that of Fig. 6 (symmetrical environments only). However, there are some very important differences between Fig. 8 (final environments) and both Fig. 6 and Fig. 7. It can be seen, for example, that the upward shifts in F_2 for /u/ and /ʊ/ in alveolar environments are much smaller for final alveolars than initial alveolars. Similarly, the upward shifts in F_2 for /ʌ/ and /ɑ/ in alveolar and velar environments and the downward shifts in F_2 for front vowels in labial environments are much more pronounced when the relevant environments are initial rather than final. The conclusion from these comparisons is that the place-dependent effects for symmetrical environments seen in Fig. 1 from SH and Fig. 6 from the present study reveal primarily the effects of the place of articulation of the initial consonant rather than the final consonant.

4. Effects of consonant voicing

Figure 9 shows the effects of consonant voicing for syllables that are symmetrical with respect to the voicing feature; for example, the data points identified as voiced (the

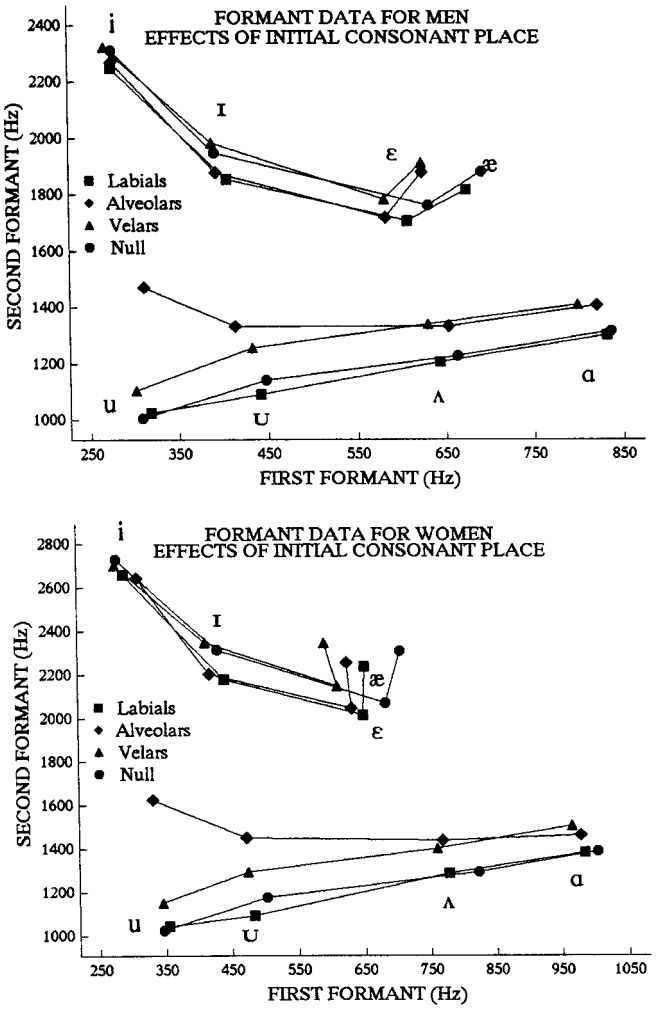


FIG. 7. Average formant frequencies at steady state as a function of the place of articulation of the initial consonant for men (top panel) and women (bottom panel).

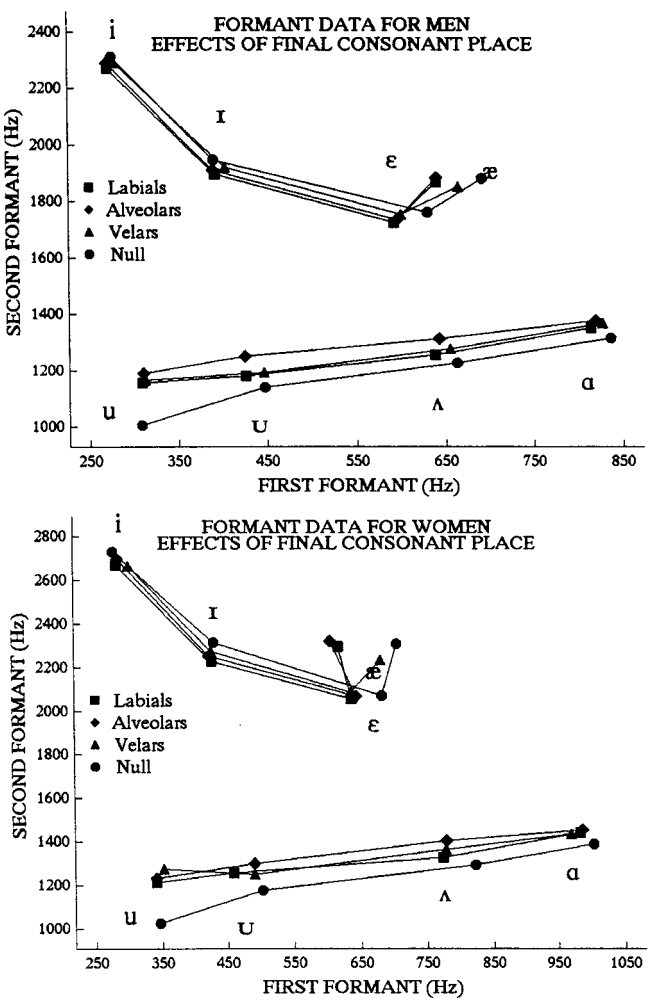


FIG. 8. Average formant frequencies at steady state as a function of the place of articulation of the final consonant for men (top panel) and women (bottom panel).

filled squares) represent averages pooled over all syllables with voiced initial and final consonants, regardless of place of production. For reference, formant values for the null environments are also shown. For the back/central vowels, the most consistent effect appears to be a tendency for the F_1 values of vowels flanked by voiced stops to be slightly lower than those in unvoiced-stop environments. For /ʌ/, the difference in F_1 between voiced and unvoiced environments is approximately 75 Hz for both men and women, but for the remaining back/central vowels the difference is quite small, typically averaging no more than 15–20 Hz. For the front vowels, the largest voicing-related differences are downward shifts in F_1 in voiced environments averaging about 90 Hz for /ɪ/, 90–120 Hz for /ɛ/, 70–100 Hz for /æ/, and negligible for /i/. In general, the tendency for F_1 values to be somewhat lower in the environment of voiced consonants is consistent with the findings of SH. We assume that these shifts in F_1 values are due at least in part to the slightly lower position of the larynx for voiced as compared to unvoiced consonants, with this difference carrying over into the vowel in the case of initial consonants and being anticipated in the case of final consonants.

Figures 10 and 11 show the effects of consonant voicing separately for initial and final consonant environments. For example, in Fig. 10, which shows initial environments, the values identified as unvoiced represent averages pooled over all non-null syllables with an unvoiced initial consonant, independent of either the place or voicing of the final consonant. Based strictly on visual inspection, Figs. 10 and 11 appear to show the same kinds of effects that were seen in symmetrical environments, but reduced in magnitude. The voicing effects, therefore, appear to derive approximately equally from initial and final consonants. The apparent attenuation of voicing effects in Figs. 10 and 11 is not surprising since half of the syllables whose formant values were pooled to calculate the means identified as voiced in Fig. 10, for example, were from syllables containing an unvoiced final consonant.

5. Spectral change patterns

Formant movement patterns, averaged across all phonetic environments, are shown in Fig. 12. The figure was created by connecting a line between the average formant values sampled at 20% of vowel duration and the average formant values sampled at 70% of vowel duration. The symbol for each vowel category is plotted at the location of the

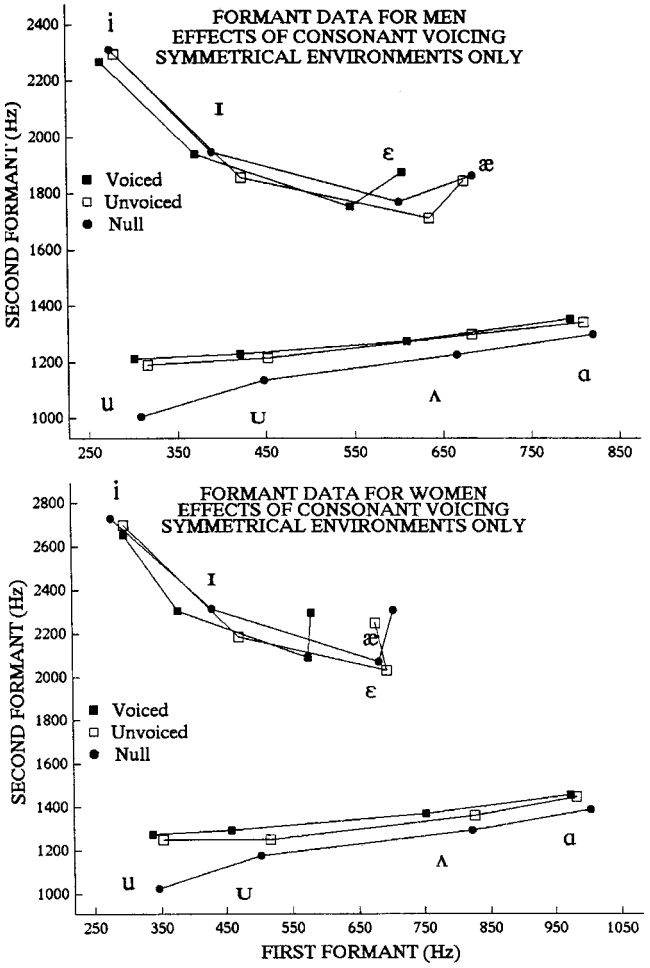


FIG. 9. Average formant frequencies at steady state as a function of the voicing of the surrounding consonants for men (top panel) and women (bottom panel). Data are shown for symmetrical environments only.

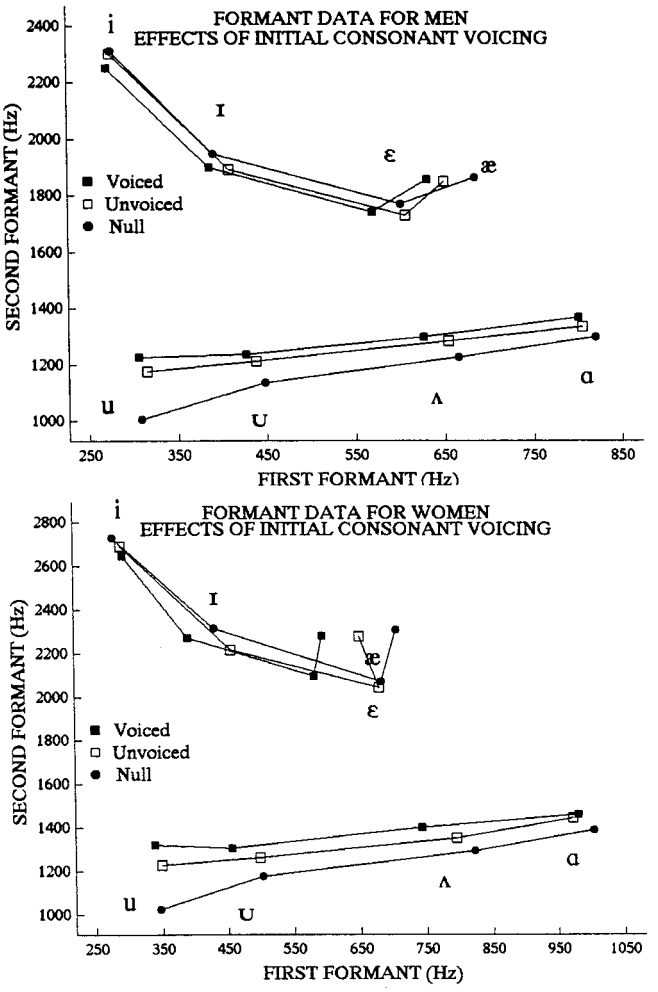


FIG. 10. Average formant frequencies at steady state as a function of the voicing of the initial consonant for men (top panel) and women (bottom panel).

second measurement; the larger symbols show formant values for the men. There are some similarities between these spectral change patterns and those observed in our earlier study of /hVd/ utterances (e.g., compare Fig. 12 with Fig. 1 of Hillenbrand and Nearey, 1999), but there are some important differences as well. Differences include: (a) the centralized offglide that was observed for /ʊ/ in the /hVd/ data is apparent in Fig. 12 as well, but the magnitude of the spectral movement is considerably attenuated; (b) the centralized offglide that was observed for /ʌ/ in the /hVd/ data is not evident in the present data; (c) the modest centralized offglide that was observed for /ɑ/ in the /hVd/ is not evident in the present data; in fact, a small movement toward the periphery is seen; and (d) the rather small centralized offglide that was observed for /ɛ/ in the /hVd/ is not evident in the present data; instead, a small movement toward the periphery was seen. Average spectral change patterns for /i/, /ɪ/, /æ/, and /u/ are grossly similar to those observed in the /hVd/ data.

6. Statistical treatment of acoustic data

A five-way factorial repeated measures analysis of variance was undertaken for the acoustic data for the stop-vowel-stop syllables.[5] The same treatment factors were considered here as in the previous analysis of the identification data, namely voicing and place of the initial consonant, vowel, and voicing and place of the final consonant. In all cases, formant frequencies were log transformed since there are clear indications that this improves the homogeneity of variance. Significance levels were determined using the Greenhouse–Geisser procedure for all F tests that involved more than one degree of freedom in the numerator. Many main effects and interactions turn out to be significant even by this conservative procedure. However, a substantial subset of the nominally significant effects accounts for a very small amount of the total phonetic variability (TPV). We have chosen to discuss only those significant interactions that account for at least 0.25% of the total variance due to all phonetic factors.

a. F_1. Significant main effects were found for all five phonetic factors: initial voicing, initial place, vowel, final voicing, and final place. (F values, F probabilities, and other numerical details concerning these and all other ANOVA results on the acoustic measurements can be found in the Appendix.) Not surprisingly, vowel effects were dominant in F_1, accounting for 97.6% of TPV. Initial and final voicing effects, though accounting for a very small proportion of TPV (0.2% and 0.3%, respectively), showed patterns consistent with previous observations. For initial consonants, voiced stops showed formant frequencies about 5.4% lower than voiceless. Initial and final place accounted for only

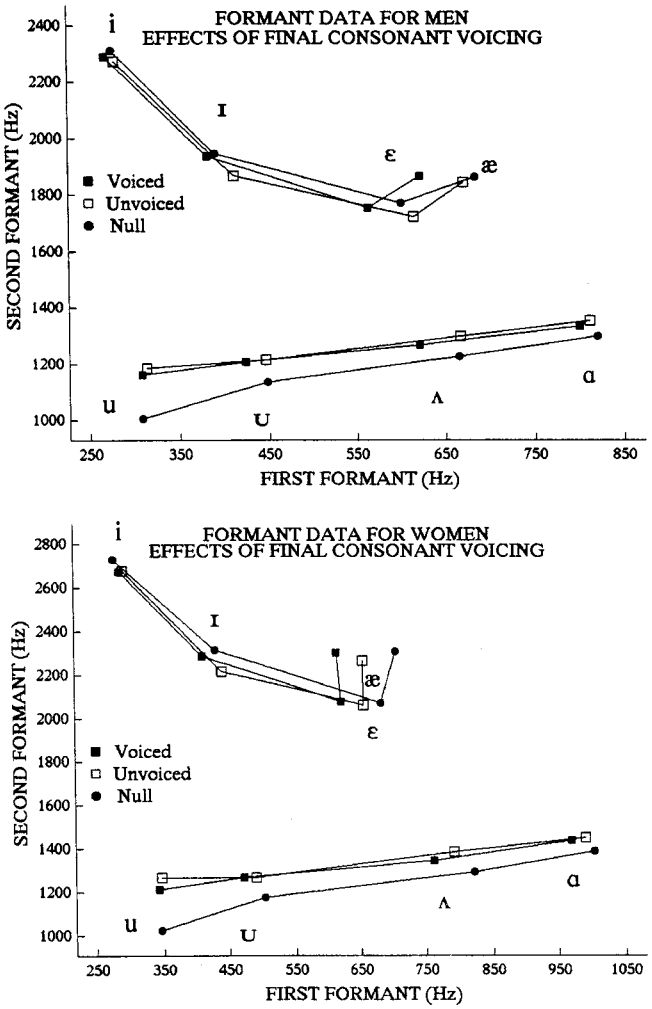


FIG. 11. Average formant frequencies at steady state as a function of the voicing of the final consonant for men (top panel) and women (bottom panel).

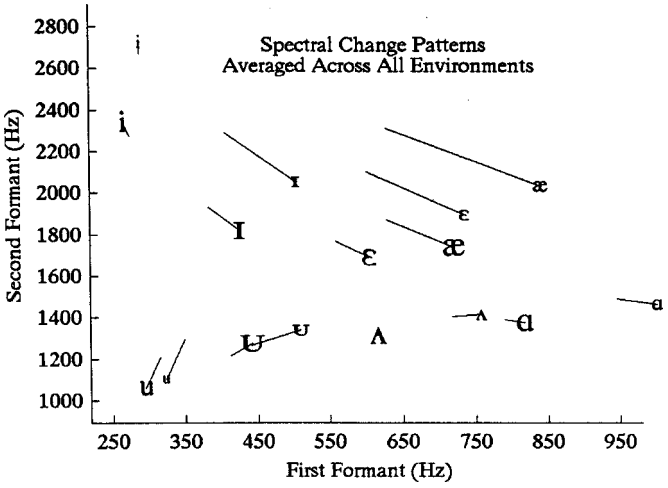


FIG. 12. Spectral change patterns for eight vowels averaged across all phonetic environments. The phonetic symbol identifying each vowel is plotted at the F_1-F_2 value for the second sample of the formant pattern (measured at 70% of vowel duration), and a line connects this point to the first sample (measured at 20% of vowel duration). The larger phonetic symbols designate formant values for men.

0.2% and 0.1% of phonetically induced variation in F_1. Sidak-corrected tests of contrasts showed that initial labials had F_1 values that were 2.4% above the mean, while velars were lower by about 2.1%. Final place effects were in the opposite direction, with final labials about 1.2% lower and final velars about 1.8% higher than average. There was only one significant interaction for F_1 that reached the variance-proportion criterion, namely initial voicing by vowel. This effect accounted for only about 0.3% of TPV. Sidak-corrected tests failed to yield any single contrast that was significantly different from zero.

b. F_2. Four of five main effects were significant for F_2. Vowels accounted for 91.2% of TPV (compared to over 97% for F_1). Initial place accounted for about 2.7% of TPV. Sidak-adjusted tests of contrasts revealed initial labial to be significantly (5.8%) lower, initial velars to be about 1.4% higher, and initial alveolars to be 4.7% higher than average. These place-induced deviations are consistent with the patterns noted by SH. Final place, while still significant, had substantially less effect on F_2 at steady state, accounting for only about 0.1% of TPV. Sidak tests showed that final labials were about 0.9% lower and final alveolars about 1.1% higher in frequency than average. Initial voicing accounted for about 0.1% of TPV, with initial voiceless stops showing F_2 values about 1.4% lower than voiced. Final voicing was not significant.[6]

By far the largest two-way interaction was initial place by vowel, which accounted for 4.8% of TPV. A large number of contrast coefficients were significant by the Sidak-adjusted test (see the Appendix). We will summarize the general findings here. The terms ''lower'' (or ''higher'') below can be interpreted as meaning that F_2 steady states for the initial place by vowel combination in question were lower (higher) than expected after adjusting for the main effects of the initial place and vowel in question. Significantly lower than expected were labial (by 9.7%) and velar (by 8.8%) before /u/; alveolars before front vowels (by 4.5% to 5.8%) and before /ɑ/ (by 2.8%). Significantly higher than expected were alveolars before /u/ (by an egregious 21.4%) and before /ʊ/ (by 5.8%); and labials before the front vowels /ɪ/, /ɛ/, and /æ/) (by 3.1% to 3.9%). Velars were also slightly higher than expected before /ʌ/ (by 1.6%) and before /ɪ/ (by 3.1%). Taken together with the main effects for place, the general trends can be viewed as being compatible with a degree of assimilation of the F_2 steady state towards roughly the expected F_2 locus for the initial consonant.

Although a number of other second- and higher-order interactions were significant, none accounted for more than about 0.2% of TPV, a criterion that corresponds to less than about 10% of the size of the main effect of initial place, or 5% of the initial place by vowel interaction.

c. F_3. The main effect of vowel accounted for 92.7% of TPV in F_3, with /u/ and /ʊ/ showing lower than average F_3 by about 8.5% and 5.6%, respectively. The vowels /i/ and /ɪ/ showed significantly higher than average F_3, by 15.8% and 2.1%, respectively. (We did not discuss the significance of contrast coefficients for vowel main effects in F_1 or F_2 because they are all significant and their pattern reflects the well-known expected locations of the vowel means in F_1-F_2 space.)

The main effect of initial place accounted for 0.9% of the TPV in F_3. Only initial velars showed a significant effect

size, being lower than average by 0.9%. The main effect of final voicing accounted for 0.8% of the TPV, with final voiceless consonants showing F_3 values about 1.2% lower than voiced. Final place of articulation accounted for about 0.3% of TPV. Sidak-adjusted tests of contrasts revealed final velars to be 0.4% lower and final alveolars to be 0.5% higher than average.

The initial place by vowel interaction was also significant, accounting for 2.7% of TPV (considerably more than the place main effect). Sidak comparisons showed initial alveolars and labials before /i/ and labials before /ɪ/ to be significantly lower than average (by 2.1% to 0.9%) while labials before /ʌ/ and /ɑ/ and velars before /i/ were higher than average (by about 1.1% to 3.2%). One additional two-way interaction, initial voicing by vowel, was also significant for F_3, accounting for 0.4% of TPV. However, Sidak tests of contrasts failed to identify any specific interaction coefficients that deviated significantly from average.

d. Summary. Statistical tests revealed a number of reliable effects of phonetic context on steady-state formant frequencies. By far the largest of these are associated with initial place of articulation in F_2. The general tendencies are in accord with preliminary observations of SH. Examination of the strong interaction effects of place with vowel also confirms that the effects of alveolars on the vowels /ʊ/ and especially /u/ are particularly strong. The large number of reliable effects in the production data contrasts with the paucity of context effects in perception. Furthermore, the largest effects in perception do not appear to correspond to those in the production data.

C. Discriminant analyses

Recall that in our earlier study of /hVd/ utterances (Hillenbrand *et al.*, 1995), discriminant analyses showed that vowels could be separated with substantially greater accuracy for pattern recognition models that incorporated spectral change as compared to otherwise comparable models that were trained on the formant pattern sampled at a single cross section of the vowel. The main purpose of the discriminant analyses reported here was to determine whether formant-frequency movements contribute to the separability of vowel categories for more complex CVC utterances in which formant movements are affected both by vowel identity and consonant environment. The pattern recognizer was a quadratic discriminant analysis technique (Johnson and Winchern, 1982) that was trained on various combinations of F_0, duration, and the three lowest formant frequencies. The formant values were sampled: (a) a single time at steady state, or (b) once at 20% of vowel duration and a second time at 70% of vowel duration.[7] For each parameter set, the pattern recognizer was run 12 separate times. On each run, the classifier was trained on 11 of the 12 talkers and tested on tokens from the single talker whose utterances had been omitted from the training. Excluded from both training and testing were: (a) tokens showing an unmeasurable formant in a formant slot that was included in the parameter list for a particular test, and (b) any token with a listener identification error rate of 15% or greater. In all cases linear frequencies in Hz were used.

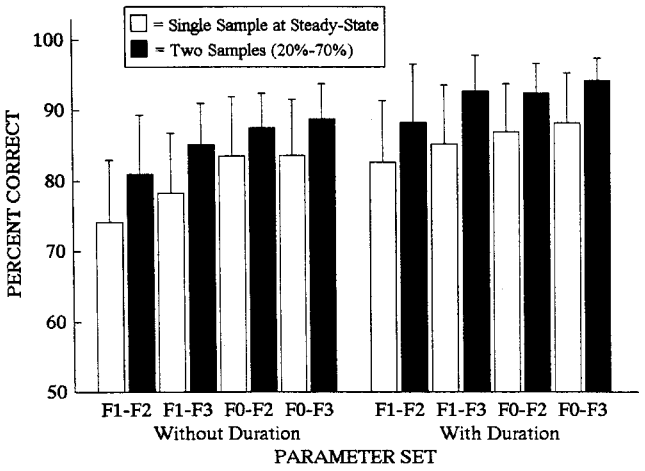


FIG. 13. Overall vowel classification accuracy for a quadratic discriminant classifier trained on various combinations of parameters.

Figure 13 shows recognition accuracy for the pattern classifier averaged across the 12 talkers for 16 different parameter sets, with the error bars showing the standard deviation calculated across the 12 talkers. It can be seen that the accuracy of the pattern classifier is higher when the model incorporates spectral change for all eight combinations of acoustic features. Averaged across the feature sets, classification accuracy was 6.1% higher for two samples of the formant pattern as compared with a single sample at steady state.[8] As shown in Table II, the improvement in classification accuracy for the two-sample models varies across vowels. Improvement with the addition of spectral change is the greatest for /ɪ/, /ɛ/, and /æ/, a cluster of vowels showing a good deal of overlap in static formant space (see Fig. 5).

As has been noted in other pattern recognition studies, there is also a substantial improvement in category separability with the addition of vowel duration. Averaged across the feature sets, classification accuracy was 6.3% higher with duration than without. As shown in Table II, very large improvements in classification accuracy averaging some 22%–25% were seen for /æ/ and /ɛ/. Substantial improvements of

TABLE II. Improvement in discriminant classification accuracy for each vowel with the addition of (a) spectral change (column 2) or (b) duration (column 3). Column 2 shows the average improvement in classification accuracy for two samples of the formant pattern as compared to a single sample. The averages were computed over the eight acoustic feature sets (i.e., F_1–F_2_NoDuration, F_1–F_3_NoDuration, F_0–F_2_NoDuration, F_0–F_3_NoDuration, F_1–F_2_Duration, etc.). Column 3 shows the average improvement in classification with the addition of duration as compared to otherwise identical parameter sets, averaged over the eight acoustic feature sets.

Vowel	Improvement with spectral change	Improvement with duration
/i/	2.5	0.4
/ɪ/	14.0	11.1
/ɛ/	15.2	24.5
/æ/	12.8	21.9
/ɑ/	9.8	12.9
/ʌ/	2.8	5.1
/ʊ/	3.5	0.6
/u/	4.2	0.2

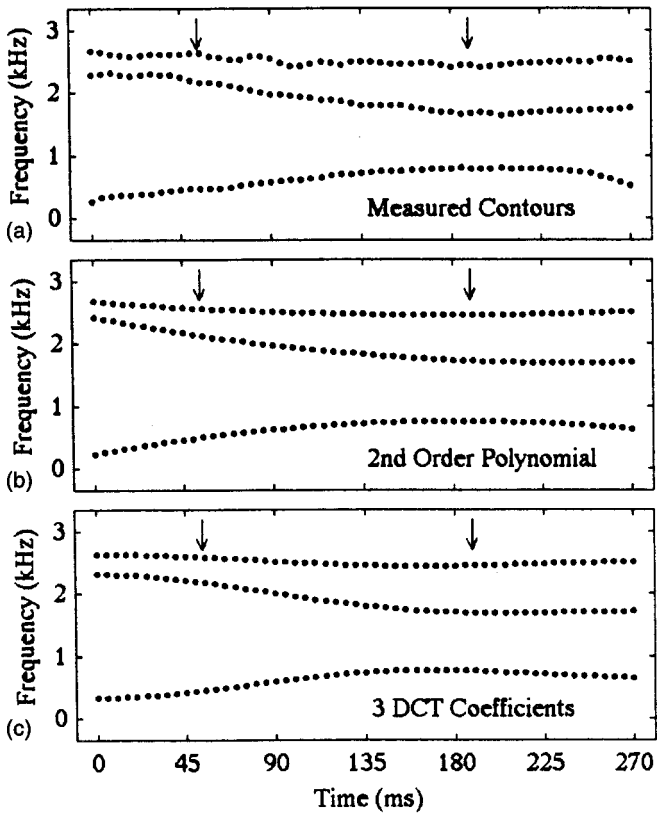


FIG. 14. From top to bottom: (a) the measured contours of F_1-F_3 for the vowel /æ/ in /gæd/; (b) a second-order polynomial fit to the contours; and (c) a three-coefficient discrete cosine transform fit to the contours. Arrows are drawn at 20% and 70% of vowel duration (see the text).

some 11%–13% were also seen for /ɪ/ and /ɑ/.

The two-sample method of capturing formant movements that was used in the present study, and in several previous studies, is not especially elegant and requires the more-or-less arbitrary selection of two discrete time points at which to sample the formant values. We experimented with two alternate schemes for capturing formant movements. One method involved fitting *n*th-order polynomials to the contours of F_1-F_3, followed by training and testing of the discriminant classifier on the coefficients of the polynomial. The curve fit was applied either to the full vowel or to the formant values from 20% to 70% of vowel duration. We experimented with several different orders of polynomial fits and different choices of sampling points. The second method involved the use of discrete cosine transform (DCT) coefficients to code the contours of the three lowest formants, as described by Zahorian and Jagharghi (1993). As with the polynomial method: (a) the coefficients were computed from the formant values of either the full vowel or the portion of the vowel from 20%–70% of vowel duration, and (b) we experimented with different numbers of DCT coefficients (see Fig. 14). The results from these two methods were not sufficiently promising to merit detailed description. Our main conclusion from this work is that both the polynomial and DCT method produced good classification results, but neither method was found to be superior to the simpler two-sample method that has been used in previous work.

For the listener data, we reported the results of statistical analyses showing the effects of consonant environment on recognition accuracy. This kind of statistical analysis is not possible with the pattern classification results for the simple reason that each token is given a single label by the pattern recognizer, providing no error term comparable to the variability in labeling responses across listeners. In the section below, we report the results of preliminary analyses that compare listener labeling responses with the output of the pattern classifier. However, there are two specific aspects of the pattern recognizer which warrant examination. As noted above, the largest context effect was a 500–600-Hz upward shift in F_2 for /u/ in the environment of initial-position alveolars. The obvious question is whether the pattern recognizer, trained on measurements from all phonetic contexts, would tend to misclassify /u/ in initial-position alveolar contexts. Using the best parameter set from those shown in Fig. 13 (duration, F_0, and two samples of F_1-F_3), the recognition rate for /u/ in the environment of initial-position alveolars was 98.3%, very similar to the 97.1% recognition rate for /u/ averaged across all contexts. The second-largest context effect was an upward shift of about 200 Hz in F_2 for /ʊ/ in the environment of initial-position alveolars. The recognition rate for /ʊ/ in initial-position alveolar environments was a respectable 85.1%, but lower than the recognition rate of 93.6% for /ʊ/ averaged across all contexts.

D. Correspondence between listener identification and discriminant analysis

As noted earlier, there is a disconnect of sorts between the listener data on the one hand and the formant measurements on the other. For example, the large number of statistically reliable effects of phonetic context on the formant frequencies contrasts sharply with the near uniformity in labeling accuracy across phonetic context. Additionally, the few reliable context effects that were observed in perception do not correspond with the largest of the effects found in the production data. At first glance, these findings would seem to discourage any consideration of a simple pattern recognition model that might account for the labeling behavior of listeners. However, as we have argued elsewhere (Hillenbrand and Nearey, 1999), it may not be adequate to compare variation in raw acoustic patterns to listeners' perception. The main reason for this is that even in the simplest model of categorization we can imagine, namely one based on minimum absolute distance of a token to a set of prototypes, more than the absolute location of a vowel token in pattern space must be considered. Specifically, the relative similarity to other category prototypes must also play a role. As noted in the example above, consider that tokens of /u/ following alveolars typically have much higher F_2 values than the population average of /u/. If those tokens have very low first formants, then they show second formants that are still substantially lower than /i/, the prototype of its nearest competitor category. Thus, it may still strike listeners as clearly more /u/-like than any other vowel, and hence the relatively large acoustic variation may produce little degradation of correct identification.

Discriminant analysis of the type reported above takes such factors of relative similarity directly into account and hence may provide perspective on the degree of correspon-

dence between classification by listeners and expected classification of the tokens based on the overall statistical properties of the distributions associated with each vowel. Here, we provide a brief analysis comparing aspects of discriminant analysis and listeners' categorization following methods described in detail in Hillenbrand and Nearey (1999). A quadratic discriminant analysis (QDA) was again performed using all 3390 tokens for which all of the following measurements were available: duration, F_0, and F_1-F_3 at 20% of the vowel duration and F_1-F_3 at 70% of the duration. When all the available tokens were used for both training and classification, 94.1% of the tokens were correctly classified. We define the modal response category of the panel of listeners as the category chosen by a plurality of listeners for each token. Under this hard-classification criterion, the panel showed a ''modally correct'' identification score of 98.1%, meaning that for all but about 2% of the tokens the label provided by a plurality of the panel agreed with the vowel intended by the talker. This is somewhat better than the QDA rate of 94.1%. There is substantial agreement between the QDA and the panel at the level of the individual token: 94.6% of tokens correctly identified by the panel were also correctly identified by QDA. There was also a reasonable level of agreement on the misidentified tokens: of the 64 tokens that were ''modally misidentified'' by the panel, the QDA chose the same incorrect category in 20 cases. Overall, the percent modal agreement between the QDA and the panel (i.e., where the panel of listeners agreed with the QDA on the best category, whether correct or incorrect) was 93.4%.

Following Hillenbrand and Nearey (1999), we also performed a *correct-response correlation analysis*, whereby the proportion of listeners' correct responses to each token was compared to the predicted *a posteriori* probability for the correct category from the QDA. The correct-response correlation r_c is defined as the Pearson correlation between the observed and predicted scores. The value of r_c will approach a maximum of 1.0 when variation in the relative probabilities of correct identification by listeners is matched by covariation in the predicted probabilities on a token-by-token basis. A modest but highly significant ($p<0.001$ by a randomization test) correlation of 0.28 was observed. This is generally similar to the value observed for the most similar model used in Hillenbrand and Nearey (1999, see Table VIII, model A). As in the case of the /hVd/ stimuli studied in Hillenbrand and Nearey, we have not yet matched listeners' performance in every respect. Nonetheless, the relatively simple discriminant analysis model adopted above seems to account well for the generally high identification rates by listeners. Similarly, it seems reasonable to suggest that lack of obvious correspondence between listeners' identification patterns and the magnitude of specific effects of context on acoustic properties is due in large measure to the relatively large degree of statistical separation among vowel classes. The reasonable success of the QDA demonstrates that the acoustic distinctiveness of the vowels is largely preserved despite variations in context.

IV. DISCUSSION

To summarize briefly, the primary purpose of this study was to evaluate the contribution of formant-frequency movements to the separability of vowel categories for CVC utterances involving variation in both the initial and final consonants; i.e., for utterances in which formant movements are influenced both by vowel identity and by coarticulatory phenomena. Recordings were made of six men and six women producing eight vowels (/i,ɪ,ɛ,æ,ɑ,ʌ,ʊ,u/) in isolation and in CVC syllables comprising all combinations of seven initial consonants (/h,b,d,g,p,t,k/) and six final consonants (/b,d,g,p,t,k/). A listening test showed high identification rates for the test utterances. More to the point, the effects of consonant environment on vowel intelligibility, while statistically significant in some cases, were quite small in magnitude, with only a few percent separating the most intelligible contexts from the least intelligible. In contrast to this perceptual stability, acoustic analysis showed a number of effects of consonant environment on vowel formant frequencies. The most important of these effects, which were generally consistent with SH, included: (1) a general tendency toward centralization for vowels produced in non-null environments; (2) large upward shifts in F_2 of 500–600 Hz for /u/ and 200–300 Hz for /ʊ/ in initial-position alveolar environments; (3) an upward shift of about 100 Hz in F_2 for /ɑ/ and /ʌ/ in initial-position alveolar environments; (4) an upward shift of about 100 Hz in F_2 for back vowels in initial-position velar environments; (5) a downward shift of about 85–100 Hz in F_2 for front vowels in initial-position labial environments; and (6) a tendency toward somewhat lower F_1 values for vowels in the environment of voiced consonants.

The central question that we sought to address was whether coarticulatory effects such as these would have the effect of obscuring the relationships between formant movement patterns and vowel identity that had been observed in several previous studies of isolated vowels or vowels in /hVd/ environments. Evidence from the pattern recognition studies reported in Sec. III C above is reasonably clear. Pattern recognition models that were trained on formant trajectories separated vowels with consistently greater accuracy than otherwise comparable models that were trained on static formant patterns. The improvement in overall classification accuracy with the addition of formant movement information was modest, averaging about 6 percentage points, but quite consistent across several combinations of parameters. Particularly large improvements averaging some 13–15 percentage points were seen for /ɪ/, /ɛ/, and /æ/.

As we have discussed elsewhere (e.g., Nearey, 1992; Hillenbrand and Nearey, 1999) pattern recognition evidence is not conclusive by itself for the simple reason that showing that a given feature improves category separability does not by itself prove that listeners make use of that feature in perception. There are, in fact, some clear examples in the literature of statistically based pattern classifiers greatly overestimating the perceptual importance of acoustic features (e.g., Hillenbrand and Gayvert, 1993b). As noted in the Introduction, in Hillenbrand and Nearey (1999) we were able to show that the pattern recognition evidence implicating an important role for spectral change in the classification of /hVd/

syllables was not, in fact, misleading. This was done by demonstrating that /hVd/ signals that had been resynthesized with flattened vowel formants were considerably less intelligible than otherwise comparable signals with formant movements matching the original utterances. In the absence of comparable perceptual evidence for the more complex CVC utterances studied here, we are forced to rely on a preliminary examination of the level of agreement between the listener data and the classification provided by the QDA. As discussed in Sec. III D, there are many—though by no means all—aspects of listener behavior that can be accounted for by a rather simple model incorporating $F0$, duration, and a very simple coding of spectral change consisting of two discrete samples of the formant pattern.

At first glance it might have been anticipated that the rather simple model of vowel classification that is embodied by the discriminant classifier would have been inherently incapable of accounting for the labeling behavior of the listener. As noted earlier, there are striking differences between the consonant-context effects that were observed in perception and those that were observed in the acoustic data. In our view, one of the most significant aspects of the listener data is the near uniformity in vowel intelligibility across different consonant environments. While there were a few statistically reliable effects of context on vowel intelligibility, these effects were small in absolute magnitude. This stands in contrast to the rather large number of reliable effects in production. Further, the largest of the effects that was observed in perception did not correspond in any obvious way to the largest context effects that were observed in perception. The clearest case of the apparently complex relationship between the production and perception data was the 500–600-Hz upward shift in F_2 values for /u/ following alveolar stops, which stands in contrast to the high intelligibility of postalveolar /u/, which was essentially indistinguishable from the intelligibility of /u/ in other environments. Taken together, these findings might well be seen as clear evidence favoring SH's conclusion that listeners internalize knowledge about the effects of context on vowel formants and invoke this knowledge in perception. While this remains a plausible account of the listener data, we do not believe that there is yet compelling evidence that a knowledge-based mechanism such as this is required. It should be kept in mind, for example, that the great majority of the statistically reliable context effects on the formant values are rather small in absolute terms. For example, the most consistent effect of context on F_1 values was a tendency toward lower frequencies in the context of voiced stops, but there was only one vowel (/ʌ/) for which the shift reached even a modest 75 Hz, and for the remaining vowels the shift averaged an auditorily undetectable 15–20 Hz. Further, of the rather large number of statistically reliable effects of context on F_2 values, the great majority averaged 100 Hz or less. There is, of course, the special case of the 500–600-Hz upward shift in F_2 for postalveolar /u/, but even for this very large context effect it is not obvious that a simple pattern recognition approach is incapable of accounting for the classification behavior of the listeners. As noted earlier, this shift has the effect of moving postalveolar /u/ upward into one of the few largely unoccupied regions of the crowded English vowel space where it remains unlikely to collide with other vowels. Our simple pattern classifier, in fact, recognized postalveolar /u/ with slightly greater accuracy than the recognition rate for /u/ averaged across all contexts.

Having made these arguments, we do not mean to imply that the variation in the acoustic properties of vowels induced by consonantal context is devoid of perceptual consequences. Specifically, we do not deny that such factors as (compensation for) consonant–vowel coarticulation (Strange, 1989; Nearey, 1989) or mechanisms of auditory contrast (e.g., Lotto and Kluender, 1998; Holt, Lotto, and Kluender, 2000) may play a role in listeners' identification of vowels in these stimuli. However, the preceding analyses indicate that even a simple two-target model, similar to that of Nearey and Assmann (1986), is adequate to account for a great deal of listeners' behavior. In future work, we plan to try to identify any systematic deviations from this baseline model and to study the relation of any such deviations to a wide range of hypotheses from the literature.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health (No. 2-R01-DC01661) to Western Michigan University and by a grant from the Canadian Social Sciences and Humanities Research Council (No. 410-93-0053) to the University of Alberta.

APPENDIX: SIDAK-CORRECTED MULTIPLE COMPARISON TESTS OF EFFECT COEFFICIENTS

The formula for Sidak correction is $p_{\text{adj}} = 1 - (1 - p_{\text{nom}})^k$, where p_{adj} is the Sidak-adjusted probability value, p_{nom} is the nominal alpha level of a single t test, and k is the number of tests in a family. The value is always less than (and hence, significance tests are more powerful than), but often very close to, the simpler Bonferroni adjusted level p_{nom}/k. Please see Tables AI–AVI.

TABLE AI. ANOVA table for F_1. TPV is percent of total variance due to the effect in question compared to that of all phonetic factors. Only significant effects accounting for at least 0.25% of TPV are shown.

Effect	TPV	MS	df_{effect}	df_{error}	F	eps_GG1	p_GG
$\text{Voicing}_{\text{init}}$	0.52%	2.5058	1	11	22.71	1.000 000	0.000 584
Vowel	97.6%	66.849	7	77	266.35	0.000 000	0.000 000
$\text{Voicing}_{\text{fin}}$	0.31%	1.4872	1	11	51.58	1.000 000	0.000 018
$\text{Voicing}_{\text{init}}$ ×vowel	0.29%	0.1957	7	77	6.28	0.000 240	0.001 510

[1]Some vowels, especially tokens of /æ/, sometimes showed a centralized offglide with a formant pattern that was as steady as that in the vowel nucleus. The 60% criterion was intended to reduce the likelihood of identifying these offglides as steady-state points.

[2]More specifically, the procedure was equivalent to the following: For the ith subject (whether talker or listener), a linear model of the following form was estimated: $Y_{iwpvzq}=M_i+W_{iw}+P_{ip}+V_{iv}+Z_{iz}+Q_{iq}+WP_{iwp}+WV_{iwv}\cdots+QW_{iwq}+WPV_{iwpv}+\cdots+ZQ_{izq}+WPVZ_{iwpvz}+\cdots PVZQ_{ipvzq}+WPVZQ_{iwpvzq}$. Here, w is the index for initial voicing, p for initial place, v for vowel, z for final voicing, and q for final place. The single corresponding capital letters indicate main-effect coefficient estimates for the corresponding terms. Pairs and n-tuples indicate two- and n-way interaction effects, respectively, while the ellipses indicate the presence of additional n-way interaction terms. The t values for specific effects are calculated by taking the mean of the specific term in the formula as the numerator and the standard error of estimate (standard deviation divided by the number of subjects). For example, tests of the vowel by initial place effect would be based on $t=X_{pv}/S_{pv}$ where $X_{pv}=n^{-1}{}_{\square i}(PV_{ipv})$ and $S_{pv}=\{n^{-1}{}_{\square i}[PV_{ipv}-X_{pv}]^2\}^{1/2}$. This approach corresponds to one in which the ANOVA was run via a regression model using effects-coding repeated measures regression (Myers and Well, 1991). As recommended by Myers and Well, a unique error term is used for each contrast rather than using a pooled estimate. In the case of perceptual data, with random listener and random talker effects, a $t'_{\min}$ statistic was calculated from the t_L (by listeners) and t_T (by talkers) analysis. The magnitude of effects is always reported as a percentage above or below expected average effects for the family (main effect or interaction) in question. This was done using an approximate inverse of the Studebaker (1985) arcsine transformation.

[3]In examining Fig. 5, Ken Stevens, who provided a critique of this manuscript, noted the relative scarcity of tokens with F_2 values of about 1500 Hz for men and about 1700 Hz for women. Also noted for back vowels was a paucity of tokens with F_1 values of about 530 Hz for men and 600 Hz for women. Since these values correspond roughly with the lowest resonances of the subglottal system (Stevens, 1999), Stevens speculated that speakers may stay clear of these regions of formant space to avoid the alignment of supraglottal and subglottal resonances which would have the effect of shifting formant values away from the values for the uncoupled upper airway.

[4]The greater upward shift in $F2$ in initial alveolar environments relative to SH is difficult to interpret unambiguously since the SH data included results pooled from dental and prepalatal consonants in addition to the alveolar consonants used in the present study.

[5]Despite careful efforts to obtain fully balanced data sets, a small proportion of the target recordings later proved unsatisfactory, leading to a small number of missing values (23 of the original 4128 utterances were omitted due to pronunciation errors that were not noticed during the recording session). The problem of missing values in experiments with random factors is a complex one, with no single widely accepted solution. While we investigated the possibility of using more sophisticated measures, we could find none that would work with problems of the size of the data at hand. We opted for an extension of a method recommended by Myers and Well (1991). This is an iterative procedure which, on the first pass, substitutes the grand mean of the complete case observations for each missing value. On subsequent passes, the estimated value of the missing cases from the

TABLE AII. Main effect and interaction contrasts for F_1 significantly different from zero (by Sidak adjusted t test, $p_$adj) for each effect in Table AI.

Effect	Level	Size (%)	t	df	p	$p_$adj
$\text{Voicing}_{\text{init}}$	$[+\text{voice}]_{\text{init}}$	−2.7	−4.8	11	0.000 584	0.001 169
	$[-\text{voice}]_{\text{init}}$	2.7	4.8	11	0.000 584	0.001 169
Vowel	/i/	−43.7	−28	11	0.000 000	0.000 000
	/u/	−34.8	−21	11	0.000 000	0.000 000
	/ɪ/	−18.6	−11	11	0.000 000	0.000 002
	/ʊ/	−9.7	−4.5	11	0.000 847	0.006 758
	/ɛ/	19.9	7.8	11	0.000 008	0.000 064
	/æ/	25.1	7.4	11	0.000 013	0.000 107
	/ʌ/	40	19	11	0.000 000	0.000 000
	/ɑ/	76.3	23	11	0.000 000	0.000 000
$\text{Voicing}_{\text{fin}}$	$[+\text{voice}]_{\text{fin}}$	−2.1	−7.2	11	0.000 018	0.000 036
	$[-\text{voice}]_{\text{fin}}$	2.1	7.2	11	0.000 018	0.000 036

TABLE AIII. ANOVA table for F_2. TPV is percent of total variance due to the effect in question compared to that of all phonetic factors. Only significant effects accounting for at least 0.25% of TPV are shown.

Effect	TPV	MS	df	dfe	F	eps_GG1	p_GG
$\text{Place}_{\text{init}}$	2.69	3.3331	2	22	147.95	0.000 000	0.000 000
Vowel	91.2	32.2500	7	77	174.83	0.000 000	0.000 000
$\text{Place}_{\text{init}}\times\text{vowel}$	4.84	0.8558	14	154	63.01	0.000 000	0.000 000

TABLE AIV. Main effect and interaction contrasts for F_2 significantly different from zero (by Sidak adjusted t test, $p_$adj) for each effect in Table AIII.

Effect	Level	Size (%)	t	df	p	$p_$adj
$\text{Place}_{\text{init}}$	$[\text{lab}]_{\text{init}}$	−5.8	−16	11	0.000 000	0.000 000
$\text{Place}_{\text{init}}$	$[\text{vel}]_{\text{init}}$	1.4	4.1	11	0.001 804	0.005 402
$\text{Place}_{\text{init}}$	$[\text{alv}]_{\text{init}}$	4.7	12	11	0.000 000	0.000 000
Vowel	/u/	−26.9	−9.5	11	0.000 001	0.000 010
Vowel	/ʊ/	−24.1	−14	11	0.000 000	0.000 000
Vowel	/ʌ/	−19.5	−15	11	0.000 000	0.000 000
Vowel	/ɑ/	−15.4	−11	11	0.000 000	0.000 003
Vowel	/ɛ/	14.6	9.6	11	0.000 001	0.000 009
Vowel	/æ/	24.2	11	11	0.000 000	0.000 003
Vowel	/ɪ/	24.7	17	11	0.000 000	0.000 000
Vowel	/i/	49.0	23	11	0.000 000	0.000 000
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{lab}]_{\text{init}}\times$/u/	−9.7	−8.3	11	0.000 004	0.000 107
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{vel}]_{\text{init}}\times$/u/	−8.8	−9	11	0.000 002	0.000 050
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times$/ɪ/	−5.8	−12	11	0.000 000	0.000 003
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times$/ɛ/	−5.4	−10	11	0.000 000	0.000 012
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times$/i/	−5	−12	11	0.000 000	0.000 002
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times$/æ/	−4.5	−8.1	11	0.000 006	0.000 147
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times$/ɑ/	−2.8	−6.6	11	0.000 040	0.000 967
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{vel}]_{\text{init}}\times$/ʌ/	1.6	4.7	11	0.000 687	0.016 366
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{vel}]_{\text{init}}\times$/ɪ/	3	9.4	11	0.000 001	0.000 031
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{lab}]_{\text{init}}\times$/ɪ/	3.1	8.3	11	0.000 005	0.000 112
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{lab}]_{\text{init}}\times$/æ/	3.7	7.5	11	0.000 011	0.000 274
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{lab}]_{\text{init}}\times$/ɛ/	3.9	13	11	0.000 000	0.000 001
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{lab}]_{\text{init}}\times$/i/	5.2	9.7	11	0.000 001	0.000 025
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times$/ʊ/	5.8	5.6	11	0.000 152	0.003 646
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times[\text{u}]$	21.4	12	11	0.000 000	0.000 003

TABLE AV. ANOVA table for F_3. TPV is percent of total variance due to the effect in question compared to that of all phonetic factors. Only significant effects accounting for at least 0.25% of TPV are shown.

Effect	TPV	MS	df	dfe	F	eps_GG1	p_GG
$\text{Place}_{\text{init}}$	0.89	0.070 345	2	22	5.80	0.009 465	0.010 973
Vowel	92.7	2.082 7	7	77	62.84	0.000 000	0.000 000
$\text{Voicing}_{\text{fin}}$	0.76	0.120 22	1	11	35.66	1.000 000	0.000 093
$\text{Place}_{\text{fin}}$	0.31	0.024 569	2	22	13.30	0.000 382	0.000 712
$\text{Voicing}_{\text{init}}\times\text{vowel}$	0.45	0.010 091	7	77	3.04	0.011 446	0.030 239
$\text{Place}_{\text{init}}\times\text{vowel}$	2.7	0.030 276	14	154	7.35	0.000 000	0.000 033

TABLE AVI. Main effect and interaction contrasts for F_3 significantly different from zero (by Sidak-adjusted t test, $p_$adj) for each effect in Table AV.

Effect	Level	Size (%)	t	df	p	$p_$adj
$\text{Place}_{\text{init}}$	$[\text{vel}]_{\text{init}}$	−0.9	−3.7	11	0.003 627	0.010 840
Vowel	/u/	−8.5	−9.3	11	0.000 001	0.000 012
Vowel	/ʊ/	−5.6	−7.9	11	0.000 008	0.000 061
Vowel	/ɪ/	2.1	4.8	11	0.000 578	0.004 612
Vowel	/i/	15.8	12	11	0.000 000	0.000 001
$\text{Voicing}_{\text{fin}}$	$[-\text{voice}]_{\text{fin}}$	−0.6	−6	11	0.000 093	0.000 186
$\text{Voicing}_{\text{fin}}$	$[+\text{voice}]_{\text{fin}}$	0.6	6	11	0.000 093	0.000 186
$\text{Place}_{\text{fin}}$	$[\text{vel}]_{\text{fin}}$	−0.4	−3.1	11	0.009 882	0.029 354
$\text{Place}_{\text{fin}}$	$[\text{alv}]_{\text{fin}}$	0.5	5.4	11	0.000 210	0.000 629
$\text{Place}_{\text{init}}\times$vowel	$[\text{alv}]_{\text{init}}\times$/i/	−2.1	−5.1	11	0.000 365	0.008 734
$\text{Place}_{\text{init}}\times$vowel	$[\text{lab}]_{\text{init}}\times$/ɪ/	−1.3	−4.4	11	0.001 019	0.024 169
$\text{Place}_{\text{init}}\times$vowel	$[\text{lab}]_{\text{init}}\times$/i/	−0.9	−4.3	11	0.001 184	0.028 022
$\text{Place}_{\text{init}}\times$vowel	$[\text{lab}]_{\text{init}}\times$/ʌ/	1.1	4.3	11	0.001 186	0.028 072
$\text{Place}_{\text{init}}\times$vowel	$[\text{lab}]_{\text{init}}\times$/ɑ/	1.4	4.9	11	0.000 494	0.011 783
$\text{Place}_{\text{init}}\times$vowel	$[\text{vel}]_{\text{init}}\times$/ɪ/	3.2	7.8	11	0.000 008	0.000 193

previous iteration is substituted. The estimates in question are based on all main effects and interactions except the highest-order interaction. This process is repeated until there is negligible change in any estimate on subsequent iteration. Statistics are based on standard methodology for fully balanced design except that degrees of freedom for error terms in ANOVA are based on the number of nonmissing cells associated with those terms. It is those adjusted degrees of freedom that are reported throughout the text and this Appendix.

[6]The absence of an effect for final voicing may have some relevance to SH's suggestion that the effects of consonant environment on vowel formants can be attributed primarily to articulatory undershoot. Since vowels preceding unvoiced stops are shorter on average than those preceding voiced stops (see Table I), one would presumably expect that the purely inertial effects associated with articulatory undershoot would be greater for the shorter-duration vowels preceding unvoiced stops. The absence of a final voicing effect would seem to call this interpretation into question.

[7]Pilot testing showed that the performance of the classifier was not strongly affected by different sampling points, making the choice of 20%–70% somewhat arbitrary. The 20%–70% sampling points performed slightly better than the 20%–80% points used in our earlier /hVd/ study.

[8]An anonymous reviewer suggested that the two-sample pattern classifier may have outperformed the one-sample classifier simply by reducing sampling error; i.e., by improving the odds that a representative sample was obtained of the effectively steady-state formant pattern. As a quick test of this possibility, we trained the pattern classifier on the same sets of parameters that were used in the tests that are summarized in Fig. 13, but tested the recognition model on parameter sets in which the ordering of the 20% and 70% samples was reversed. We reasoned that if the issue is simply sampling error, then the ordering of the samples will be unimportant. However, if it is the formant trajectory that is being captured by the two-sample method, as we argue, then the ordering of the samples will make a great deal of difference. The results showed that the ordering of the samples matters. Averaged across all parameter sets, classification accuracy was 10.6 percentage points higher for natural order than reverse order. Classification accuracy for the two-sample reverse-order features was also 4.3 percentage points lower on average than for a single sample of the formant pattern. See Nearey and Assmann (1986) for comparable tests with human listeners.

Andruski, J. E., and Nearey, T. M. (**1992**). ''On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables,'' J. Acoust. Soc. Am. **91**, 390–410.

Assmann, P., Nearey, T., and Hogan, J. (**1982**). ''Vowel identification: Orthographic, perceptual, and acoustic aspects,'' J. Acoust. Soc. Am. **71**, 975–989.

Black, J. W. (**1949**). ''Natural frequency, duration, and intensity of vowels in reading,'' J. Speech Hear. Disord. **14**, 216–221.

Clark, H. (**1973**). ''The language-as-fixed-effect fallacy,'' J. Verbal Learn. Verbal Behav. **12**, 335–339.

Crystal, T. H., and House, A. S. (**1988**). ''Segmental durations in connected-speech signals: Current results,'' J. Acoust. Soc. Am. **83**, 1553–1573.

Fairbanks, G., and Grubb, P. (**1961**). ''A psychophysical investigation of vowel formants,'' J. Speech Hear. Res. **4**, 203–219.

Fischer-Jorgensen, E. (**1964**). ''Sound duration and place of articulation,'' Zeitschrift Sprachwisse Phonetik **17**, 175–207.

Hillenbrand, J. (**1988**). ''MPITCH: An autocorrelation fundamental-frequency tracker,'' [Computer Program], Western Michigan University, Kalamazoo, MI.

Hillenbrand, J. M., and Gayvert, R. T. (**1993a**). ''Identification of steady-state vowels synthesized from the Peterson and Barney measurements,'' J. Acoust. Soc. Am. **94**, 668–674.

Hillenbrand, J. M., and Gayvert, R. T. (**1993b**). ''Vowel classification based on fundamental frequency and formant frequencies,'' J. Speech Hear. Res. **36**, 694–700.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Hillenbrand, J. M., and Nearey, T. N. (**1999**). ''Identification of resynthesized /hVd/ syllables: Effects of formant contour,'' J. Acoust. Soc. Am. **105**, 3509–3523.

Holt, L. L., Lotto, A. J., and Kluender, K. L. (**2000**). ''Neighboring spectral content influences vowel identification,'' J. Acoust. Soc. Am. **108**, 710–722.

Jenkins, J. J., and Strange, W. (**1999**). ''Perception of dynamic information for vowels in syllable onsets and offsets,'' Percept. Psychophys. **61**, 1200–1210.

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

Johnson, R. A., and Winchern, D. W. (**1982**). *Applied Multivariate Statistical Analysis* (Prentice-Hall, Englewood Cliffs, NJ).

Lotto, A. J., and Kluender, K. R. (**1998**). ''General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification,'' Percept. Psychophys. **60**, 602–619.

Miller, J. D. (**1989**). ''Auditory-perceptual interpretation of the vowel,'' J. Acoust. Soc. Am. **85**, 2114–2134.

Myers, J. L., and Well, A. D. (**1991**). *Research Design and Statistical Analysis* (HarperCollins, New York).

Nearey, T. M. (**1989**). ''Static, dynamic, and relational properties in vowel perception,'' J. Acoust. Soc. Am. **85**, 2088–2113.

Nearey, T. M. (**1992**). ''Applications of generalized linear modeling to vowel data,'' in *Proceedings of ICSLP 92*, edited by J. Ohala, T. Nearey, B. Derwing, M. Hodge, and G. Wiebe (University of Alberta, Edmonton, Alberta, Canada), pp. 583–586.

Nearey, T. M., and Assmann, P. (**1986**). ''Modeling the role of vowel inherent spectral change in vowel identification,'' J. Acoust. Soc. Am. **80**, 1297–1308.

Parker, E. M., and Diehl, R. L. (**1984**). ''Identifying vowels in CVC syllables: Effects of inserting silence and noise,'' Percept. Psychophys. **36**, 369–380.

Peterson, G. E., and Lehiste, I. (**1960**). ''Duration of syllable nuclei in English,'' J. Acoust. Soc. Am. **32**, 693–703.

Stevens, K. N. (**1999**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).

Stevens, K. N., and House, A. S. (**1963**). ''Perturbation of vowel articulations by consonantal context: An acoustical study,'' J. Speech Hear. Res. **6**, 111–128.

Strange, W. (**1989**). ''Dynamic specification of coarticulated vowels spoken in sentence context,'' J. Acoust. Soc. Am. **85**, 2135–2153.

Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). ''Dynamic specification of coarticulated vowels,'' J. Acoust. Soc. Am. **74**, 695–705.

Studebaker, G. A. (**1985**). ''A 'rationalized' arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.

Zahorian, S., and Jagharghi, A. (**1993**). ''Spectral shape features versus formants as acoustic correlates for vowels,'' J. Acoust. Soc. Am. **94**, 1966–1982.

Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?

Lori L. Holt[a)]
Department of Psychology and Center for the Neural Basis of Cognition, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

Andrew J. Lotto
Department of Psychology, Washington State University, P.O. Box 644820, Pullman, Washington 99164-4820

Keith R. Kluender
Department of Psychology, University of Wisconsin—Madison, 1200 West Johnson Street, Madison, Wisconsin 53706

(Received 11 August 2000; accepted for publication 16 November 2000)

For stimuli modeling stop consonants varying in the acoustic correlates of voice onset time (VOT), human listeners are more likely to perceive stimuli with lower $f0$'s as voiced consonants—a pattern of perception that follows regularities in English speech production. The present study examines the basis of this observation. One hypothesis is that lower $f0$'s enhance perception of voiced stops by virtue of perceptual interactions that arise from the operating characteristics of the auditory system. A second hypothesis is that this perceptual pattern develops as a result of experience with $f0$-voicing covariation. In a test of these hypotheses, Japanese quail learned to respond to stimuli drawn from a series varying in VOT through training with one of three patterns of $f0$-voicing covariation. Voicing and $f0$ varied in the natural pattern (shorter VOT, lower $f0$), in an inverse pattern (shorter VOT, higher $f0$), or in a random pattern (no $f0$-voicing covariation). Birds trained with stimuli that had no $f0$-voicing covariation exhibited no effect of $f0$ on response to novel stimuli varying in VOT. For the other groups, birds' responses followed the experienced pattern of covariation. These results suggest $f0$ does not exert an obligatory influence on categorization of consonants as [VOICE] and emphasize the learnability of covariation among acoustic characteristics of speech. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1339825]

PACS numbers: 43.71.An, 43.71.Es, 43.71.Pc [CWT]

I. INTRODUCTION

Among the world's languages, fundamental frequency ($f0$) and voicing tend to covary. Cross-linguistically, this observation is extremely reliable; so reliable, in fact, that this relationship has been said to arise as a result of physiological constraints on speech production. However, cross-linguistic analysis demonstrates that $f0$ and the acoustic correlates of voice onset time (VOT) covary only among consonants that are used distinctively by languages (Kohler, 1982, 1984, 1985; Kingston, 1986; Kingston and Diehl, 1994), thus suggesting that the influence is not a mandatory consequence of the speech-production system. Vowels immediately following voiced consonants (e.g., [b], [d], [g]) tend to have lower $f0$'s than those following voiceless consonants (e.g., [p], [t], [k]; House and Fairbanks, 1953; Lehiste and Peterson, 1961; Mohr, 1971; Hombert, 1978; Caisse, 1982; Peterson, 1983; Ohde, 1984).[1] For example, the fundamental frequency of the vowel [ʌ] (as in *bud*) tends to be lower in the utterance [dʌ] than in the syllable [tʌ] (Kingston and Diehl, 1994).

The covariation between $f0$ and voicing[2] in language production has a corresponding regularity in speech perception. When listeners categorize synthetic or digitally manipulated natural speech tokens of a phonetic series varying perceptually from voiced to voiceless (e.g., from [ba] to [pa]) listeners more often identify tokens as voiced (i.e., as [ba]) when $f0$ is low. For higher $f0$'s, listeners more often report hearing voiceless consonants (i.e., [pa]). This finding is extremely robust, and has been reported across multiple phonetic contexts, using a variety of measures (e.g., Chistovich, 1969; Haggard *et al.*, 1970; Fujimura, 1971; Massaro and Cohen, 1976, 1977; Derr and Massaro, 1980; Gruenenfelder and Pisoni, 1980; Haggard *et al.*, 1981; Kohler, 1985; Kohler and van Dommelen, 1986; Whalen *et al.*, 1993; Castleman and Diehl, 1996).

Perception of voiced versus voiceless consonants thus follows the regularities of speech production. Much has been made of this correspondence and a good deal of speculation has surrounded the question of why $f0$ and VOT covary in speech production (e.g., Kingston and Diehl, 1994). However, the mechanisms that govern the perceptual side of this correspondence remain largely unknown.

Diehl and Kluender (1989) offer an hypothesis that accounts for the regularities in speech perception and production. By their *auditory enhancement* account, constellations of articulations (such as those that lead to low $f0$ and other

a)Electronic mail: lholt@andrew.cmu.edu

characteristics of [+voice] consonants) tend to covary because these combinations confer a perceptual advantage to the listener. The covariance of $f0$ and voicing in speech production, they argue, is a result of language communities' tendency to coordinate components of speech that have mutually enhancing auditory effects. Thus, the perceptual interaction of $f0$ and VOT evident in listening studies is due to the stable operating characteristics of the auditory system. For this account to provide an explanation for the perceptual influence of $f0$ on voicing categorization, it must offer the additional hypothesis that the coupling of low $f0$'s with the acoustic characteristics of voiced consonants interacts, in some manner, to create an acoustic signal that is more robust than alternative combinations. In a proposal that builds upon the earlier work of Stevens and Blumstein (1981), Diehl *et al.* (1995) have argued that low $f0$ contributes to the presence of low-frequency energy during and near the consonant, thus enhancing the perception of voicing. This auditory enhancement account therefore implies that the natural covariation of $f0$ and voicing observed across languages confers an advantage in auditory processing by virtue of making low-frequency energy more salient for [+voice] consonants.

In a similar vein, Kingston and Diehl (1994) have argued that mutually enhancing characteristics of speech production are explicitly represented by a level of representation intermediate individual acoustic/phonetic correlates of voicing, like a low $f0$, and higher-level representations of distinctive features (such as [+voice]). The advantage of these *integrated perceptual properties* (IPPs) is in limiting energy expenditure in speech production and producing mutually enhancing acoustic effects, thus aiding communication for both the speaker and the listener. Kingston and Diehl propose that the auditory system literally treats a low $f0$ at vowel onset and a short VOT as perceptually equivalent because both act to increase the percept of low-frequency energy near the stop consonant. Like earlier auditory enhancement accounts, this hypothesis implies that the influence of $f0$ on categorization of consonants as voiced or voiceless is a result of demands upon language to provide a robust, readily intelligible, speech signal. The distinction of this account from earlier treatments of auditory enhancement is that it posits an additional level of representation (Diehl *et al.*, 1995).

These auditory enhancement accounts are in agreement that the reason $f0$ acts to shift listeners' perception of voicing is that the acoustic cues provided by $f0$ and voicing interact to enhance some perceptual property (e.g., the perceived presence of low-frequency energy). By these accounts, stable characteristics of the auditory system are responsible for producing perceptual interactions among the acoustic characteristics of $f0$ and voicing. However, preexisting auditory characteristics may not be the only feasible explanation for the interaction of $f0$ and voicing in speech perception. After all, $f0$ and voicing covary in speech production and, as a result, the language environment is rich with structured covariance. Should listeners be sensitive to this covariance, experience with it could influence categorization of voiced and voiceless consonants that vary in $f0$. That is, perceptual interactions between $f0$ and voicing might arise from learning. Recent results have demonstrated the considerable influence covariation within the language environment has upon speech perception. For example, Pitt and McQueen (1998) have shown that listeners are sensitive to the natural covariation of phonetic segments by demonstrating that experimental manipulation of transitional probabilities between speech sounds can elicit predictable context effects. Saffran *et al.* (1996) likewise have argued that adult as well as infant listeners use natural covariation among syllables in word segmentation.

It is possible that listeners' experience with the natural covariation between $f0$ and voicing shapes speech perception. By this proposal, $f0$-voicing covariation in speech *production* arises from speakers' tendency to produce voiced consonants with lower $f0$'s than their voiceless counterparts. Listeners may learn and use this covariation such that $f0$ and voicing interact in speech *perception*. This learning account makes fewer predictions than the auditory enhancement account about why speech production should be so patterned. One possibility is that articulations of $f0$ and VOT are not fully independent. However, as Kingston and Diehl (1994) have argued, no one has yet explained how the articulations of $f0$ and VOT depend on one another. Whatever the nature of the linguistic habits of speakers that promote $f0$/VOT covariation, they remain to be fully uncovered.

The following experiment was designed to tease apart the relative roles that stable preexisting auditory characteristics and effects of experience with $f0$-voicing covariation have in explaining the influence of $f0$ on categorization of consonants as voiced or voiceless. Among human listeners, especially native English speakers who have had extensive experience with $f0$-voicing covariation, this is an extraordinarily difficult task. It would be most desirable to have a population of listeners who are inexperienced with covariation between $f0$ and voicing. Among these individuals, it would be possible to exercise complete experimental control over experience and thus assess relative contributions of audition versus learning.

Nonhuman animals are just such a population. An extensive literature now exists to demonstrate the feasibility of using nonhuman animals in experiments aimed at understanding human speech perception. For the most part, nonhuman animals have provided two distinct services in developing our knowledge of speech perception. In one way, they have served as ''pristine'' auditory systems, unblemished by the experience with speech that human listeners bring to the laboratory. In experiments designed to exploit this characteristic, it is possible to examine the contributions of audition to speech perception while factoring out potential effects of experience. From these experiments, we have learned that nonhuman animals respond to speech categorically (Morse and Snowdon, 1975; Kuhl and Miller, 1975, 1978; Waters and Wilson, 1976), exhibit phonetic context effects (Dent *et al.*, 1997; Lotto *et al.*, 1997), and are sensitive to acoustic trading relations (Kluender, 1991; Kluender and Lotto, 1994). Nonhuman animals have also provided a means of directly manipulating experience with speech to test its effect (Kluender *et al.*, 1987, 1998; Lotto *et al.*, 1999). In experiments of this sort, animals have served as a population in which

there is the possibility of exquisite experimental control over speech experience. These methods have allowed rather precise characterization of effects of experience that can be difficult to garner with human adult or infant listeners (see Holt *et al.*, 1998 for a discussion of these issues) and have led to demonstrations that nonhuman animals exhibit learning-dependent hallmarks of speech perception such as phonetic categorization (Kluender *et al.*, 1987) and internal phonetic category structure (Kluender *et al.*, 1998; Lotto *et al.*, 1999).

Under both experimental paradigms, nonhuman animals have served well, demonstrating that they often respond to speech in much the same manner as human listeners. The present experiment is a fusion of these two experimental approaches. Here, the aim is to delineate the relative contributions of perceptual interactions arising from stable operating characteristics of the auditory system and those arising from experience with covariation in the environment. The present design investigates the influence of $f0$ upon nonhuman animals' responses to stimuli that vary in voicing via manipulation of VOT across three conditions; two conditions provide experience with $f0$-voicing covariation and a third strictly eliminates such experience. Nonhuman animal listeners are essential for this endeavor because they allow rigorous experimental control over the characteristics of experience with $f0$-voicing covariation.

II. EXPERIMENT

Japanese quail (*Coturnix coturnix japonica*), an avian species that has been used extensively in auditory and speech perception research, served as listeners. [See Dooling and Okanoya (1995) for behaviorally derived quail audiograms.] Quail were chosen because they have proven to be very capable subjects in auditory learning tasks (Kluender *et al.*, 1987; Lotto *et al.*, 1999) and avian subjects, in general, are known to exhibit phonetic categorization that mimics essential aspects of human phonetic categorization (Kluender *et al.*, 1987, 1998; Lotto *et al.*, 1999). In addition, quail have demonstrated the ability to respond behaviorally to voiced versus voiceless stimuli (Kluender, 1991; Kluender and Lotto, 1994).

The experiment was designed to assess potential influences of auditory constraints and experience with $f0$-voicing covariation on quails' responses to voicing as $f0$ varies. To achieve this aim, quail were assigned to one of three conditions. Birds in each condition first were trained to respond to either voiced or voiceless syllable-initial stop consonants by pecking a key. Conditions were distinguished by the manner that $f0$ and voicing covaried in the set of stimuli used to train the quail. For birds in the first condition, $f0$ and voicing covaried in the natural manner; voiced consonants had lower $f0$'s than voiceless consonants. A second group of quail was trained with stimuli that varied in the reverse manner; voiced consonants had *higher* $f0$'s and voiceless consonants had *lower* $f0$'s.

A final subset of the quail experienced training stimuli that had no orderly covariation between $f0$ and voicing; the two dimensions were uncorrelated among these training stimuli. As a result, the quail in this last group received no orderly experience with $f0$-voicing covariation. Thus, this condition provides a control group with which to compare the other two conditions. If auditory interactions between $f0$ and voicing influence birds' response to novel test stimuli in the control condition, then birds should respond more robustly to voiced consonants with low $f0$ and voiceless consonants with high $f0$ despite that they have no experience with $f0$-voicing covariation. If experience with covariation between $f0$ and voicing is responsible for the perceptual trading relation, then this group of quail should exhibit no effect of $f0$ upon voicing response.

The remaining conditions are critical to the issue of whether experience with $f0$-voicing covariation influences patterns of perception. If experience with covariation is responsible for the effect of $f0$ on voicing identification, quail that experience natural covariation during training should exhibit an influence of $f0$ on response to novel stimuli in the direction of covariation. However, because the pattern of experience for this condition follows that observed naturally in languages, effects observed are difficult to disentangle from putative auditory influences. However, if quail that experience *reversed* covariation exhibit a ''reversed'' influence of $f0$ on response to voicing such that consonants with *higher* $f0$'s are more often responded to as voiced consonants, then the pattern of behavior should mirror that of the input covariance and diverge from the pattern of results predicted by auditory interactions.

A. Method

1. Subjects

Twenty-one adult Japanese quail (*Coturnix coturnix japonica*) served as listeners in the experiment. Some of the quail with smaller body weights failed to reach criterion performance (pecking ten times more often to positive stimuli than to negative stimuli, see Sec. II B) leaving 16 quail to enter into testing. Free-feed weights ranged from 104 to 160 g.

2. Stimuli

a. Stimulus synthesis. A bilabial stop-consonant series of 19 stimuli varying perceptually from [ba]-[pa] was synthesized using the parallel branch of the Klatt (1980) speech synthesizer. Endpoint stimuli were based upon the productions of a single male talker who uttered ''ba'' and ''pa'' in isolation. For all stimuli in the series, nominal formant frequencies were equivalent. First through third formants ($F1-F3$) were 150, 800, and 2100 Hz, respectively, at stimulus onset. Frequencies for all three formants changed linearly over the next 40 ms to 750, 1220, and 2600 Hz for $F1-F3$. Formant frequencies remained at these values for the remainder of the 250-ms total duration.

Voice onset time was modeled acoustically by varying amplitude of voicing (Klatt parameter AV, set to zero during VOT duration), noise in the signal (AH=55 during VOT duration) and amplitude of the first formant (A1=0 during VOT duration, to model $F1$ cutback). These parameters were varied in 5-ms steps to mimic acoustic changes from 5 to 95 ms VOT. From this base set of stimuli, fundamental frequency ($f0$) was manipulated to create a full stimulus corpus that varied in VOT (5–95 ms) and $f0$. To accomplish

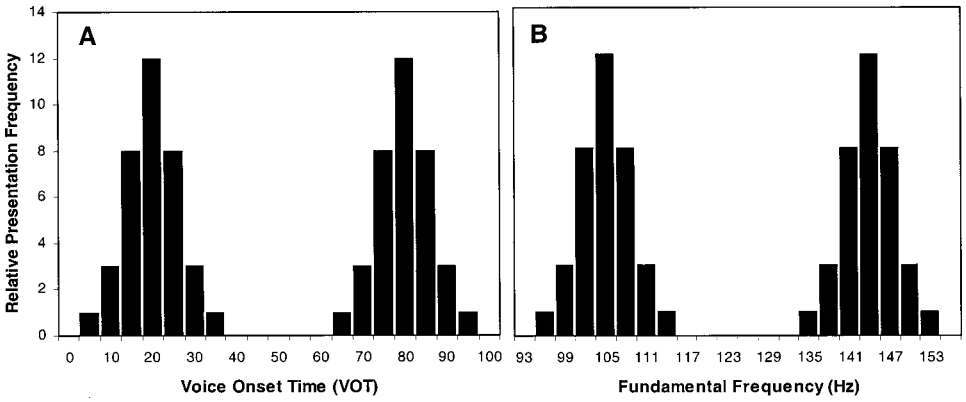


FIG. 1. Sampling distributions for voice onset time [VOT, panel (a)] and fundamental frequency [$f0$, panel (b)] from which stimuli were drawn for presentation to quail during training.

this, the 19-member series varying perceptually from [ba]-[pa] was synthesized at 14 different $f0$'s, 96–114 Hz in 3-Hz steps and 135–153 Hz in 3-Hz steps. This created an $f0 \times$VOT stimulus space consisting of a corpus of 266 stimuli (19 VOT values $\times 14$ $f0$'s). Two distinct ranges for $f0$ were created to provide a ''low'' versus ''high'' distinction, with non-overlapping values roughly corresponding perceptually to male and female voice. For each stimulus, $f0$ was constant across the entire stimulus duration.[3]

b. Stimulus sampling. Typically, speech perception experiments that examine phonetic labeling do so using one or several phonetic series that vary perceptually from one clearly identifiable phonetic endpoint to another via an acoustic manipulation. This is generally true of both human and nonhuman studies of speech perception. Here, one of the primary goals is to examine the role of experience in shaping perception. As a result, the traditional approach is adapted to better model some of the statistical characteristics of speech sound distributions that human listeners encounter.

Although there have been few large-scale efforts to measure the acoustic characteristics of multiple phonetic segments across multiple speakers, those that exist (e.g., Peterson and Barney, 1952; Lisker and Abramson, 1964) suggest that there is a good deal of variability in the acoustic characteristics of speech sounds across speakers. Fortunately for listeners, there is also a good deal of regularity. In an early inventory of cross-language stop-consonant voicing, for example, Lisker and Abramson (1964) observed between- and within-speaker variation in VOT. However, they also reported very regular underlying acoustic patterns for the phonemes of a particular language. Across speakers and productions, estimated VOT values tended to cluster around a particular mean value that occurred most frequently across productions of a particular phoneme (see Newman, 1997). In addition, there was variance such that values adjacent to the mean were also observed, but less frequently. To put this observation in more concrete terms, the measured values roughly approximated a normal (Gaussian) distribution. Therefore, there is reason to believe that normal distributions are a reasonable approximation of the distributions underlying variability in speech production.

In line with these observations, stimuli from the VOT $\times f0$ stimulus space were not presented equally often during the experiment. Rather, there was a statistical structure to the manner in which stimuli were sampled from the space. Independent distributions for sampling $f0$ and VOT were created by modeling Gaussian (normal) distributions with variance of 1.25 stimulus steps (based upon series of 3-Hz steps for $f0$ and 5-ms steps for VOT). For each dimension, two distributions were created. These distributions corresponded to high versus low $f0$ (distribution means were 105 and 144 Hz, respectively) and voiced versus voiceless consonants (with distribution means of 20 and 80 ms).[4] Stimuli were presented with relative frequencies that were discrete approximations of these continuous Gaussian distributions. Figure 1 illustrates relative frequencies for $f0$ and VOT values. Using distributions of $f0$ and VOT values rather than individual stimuli (e.g., one stimulus with a low $f0$ versus one with a high $f0$) allowed for a more sensitive test of interactions between $f0$ and VOT because it encouraged quail to generalize to novel stimuli. Likewise, it provided a more realistic model of $f0$/VOT covariation.

c. Stimulus presentation. Stimuli were synthesized with 12-bit resolution at a 10-kHz sampling rate, matched in rms energy and stored on a computer disk. Stimulus presentation was under the control of an 80386 computer. After D/A conversion (Ariel DSP-16), stimuli were low-pass filtered (4.8-kHz cutoff frequency, Frequency Devices #677), amplified, and presented to quail via a single 13-cm speaker (Peerless 11592) in a tuned enclosure providing flat frequency response from 40 to 5000 Hz. Sound level was calibrated by placing a small sound-level meter (Bruel & Kjaer 2232) in the chamber with the microphone positioned at approximately the same height and distance from the speaker as a bird's head.

B. Procedure

1. Training stimuli

Reinforcement contingencies were structured to train quail to respond differentially to training stimuli drawn from the VOT distributions shown in Fig. 1(a) (i.e., 5–35 ms VOT versus 65–90 ms VOT). Half of the birds in each condition were rewarded for pecking in response to voiced stimuli (5–35 ms VOT, designated +voice birds). Longer VOT (65–95 ms) signaled reinforcement for the remaining quail (−voice birds). The exact stimuli that were presented to quail in a given training session were determined by randomly sampling from the Gaussian distributions described earlier (see Fig. 1). Stimuli were constrained in the number

of positive versus negative stimuli that could occur in a session, but otherwise $f0$ and VOT were randomly selected from the appropriate distribution.

During training, stimuli varied by condition. One group of quail was trained with stimuli exhibiting ''natural'' covariance. That is, voiced stimuli had low $f0$ and voiceless stimuli had high $f0$. Quail in this group heard stimuli created to mimic 5–35-ms VOT that were synthesized with an $f0$ varying between 96 and 114 Hz and 65–95-ms VOT stimuli with an $f0$ of 135–153 Hz. The matching between $f0$ and VOT adhered to these constraints, although the precise $f0$/VOT stimuli varied randomly from within the sampling distributions (see Fig. 1). Another group of quail was trained on the ''reverse'' of this covariation. These birds heard 5–35-ms VOT stimuli synthesized with an $f0$ of 135–153 Hz and 65–95-ms VOT stimuli synthesized with an $f0$ of 96–114 Hz. For a final group of ''control'' quail, $f0$ and VOT did not covary. Stimuli presented to these quail had random assignment of $f0$ to VOT. For this group of quail, $f0$ and VOT values were chosen from the sampling distributions in the same manner as for the other quail. However, no constraints upon $f0$ to VOT mapping were enforced; $f0$ was assigned randomly to VOT values. For any presentation, $f0$ could be chosen from either the high or the low distribution, independent of VOT value.

2. Training procedure

Following 18 to 22 h of food deprivation (adjusted to each bird individually for optimal performance)[5] birds were weighed and placed in a small sound-attenuated chamber within a larger single-wall sound-attenuated booth (Suttle Acoustics Corp.). In a go/no-go identification task, quail pecked a lighted key (1.2 cm square) located 15 cm above the floor and centered below a speaker from which stimuli were presented. A computer recorded responses and controlled reinforcement.

Following magazine training and autoshaping procedures, reinforcement contingencies were gradually introduced over 8 days in sessions of 60 to 72 trials. During this time, average amplitude of the stimuli was increased from 50 to 70 dB SPL to introduce sound without startling the birds. Average trial duration increased from 5 to 30 s, intertrial interval decreased from 40 to 15 s, average time to reinforcement increased from 5 to 30 s, and the ratio of positive to negative trials decreased from 4:1 to 1:1. After the gradual introduction of reinforcement contingencies over eight days, daily training sessions consisted of 72 stimuli (36 positive and 36 negative).

On each trial, a stimulus was presented repeatedly once per 1550 ms at an average peak level of 70 dB SPL. On a trial-by-trial basis, the intensity of stimuli varied randomly from the mean of 70 dB by ± 0–5 dB via a computer-controlled digital attenuator (Analog Devices 7111). The average duration of each trial was 30 s, varying geometrically from 10 to 65 s. The intertrial interval was 15 s. Responses to positive stimuli were reinforced on a variable-interval schedule by 1.5–2.5 s of access to food from a hopper beneath the peck key. Duration of reinforcement was also adjusted for each bird for consistent performance. Average interval to reinforcement was 30 s (10–65 s) so that positive stimuli were reinforced on an average of once per trial. Note that when a trial was long (e.g., 65 s) and times to reinforcement were short (e.g., 10 s), reinforcement was available more than once. Likewise, on shorter positive trials, reinforcement did not become available if time to reinforcement was longer than the trial. Any reinforcement interval that did not expire during one positive trial carried over to the next positive trial. Such intermittent reinforcement encouraged consistent peck rates during later non-reinforced testing trials. During negative trials, birds were required to refrain from pecking for 5 s for the trial to be terminated. This procedure has been used successfully to train Japanese quail in other speech perception tasks (Kluender, 1991; Kluender and Lotto, 1994; Lotto *et al.*, 1997, 1999).

3. Testing

All birds learned quickly to respond differentially to VOT. Birds continued to train with the distributions until they responded with 10:1 performance for positive versus negative stimuli. Among the 21 quail that began magazine and autoshape training, 16 quail made it to criterion performance and were entered in the testing procedure. Of these birds, five were in the ''natural'' condition, seven were in the ''reverse'' condition, and four were in the ''control'' condition.

Following training, quail were tested on novel, intermediate members of the [ba]-[pa] series (VOT from 40 to 60 ms in 5-ms steps) synthesized with $f0$ of 105 and 141 Hz, the modes of the $f0$ sampling distributions of training stimuli. In all, ten stimuli were tested (2 $f0$'s$\times$5 novel intermediate series members). In each daily test session, the ten test trials (each with a fixed trial duration of 30 s) were randomly interspersed among normal training trials. Each testing session began with 15 trials of training stimuli. Novel test trials could not occur until after these 15 trials as an assurance that birds had ''settled into'' the task before responding to test stimuli. After these initial trials, 10 test trials were randomly interspersed with 60 training trials for a total of 70 trials per test session. Contingencies remained the same for training stimuli, but during test trials no contingencies were in effect. Birds neither received food reinforcement nor needed to refrain from pecking for presentation to terminate after 30 s. Training and testing stimuli were randomly ordered for each bird. Testing continued for 20 daily sessions, providing a data set consisting of birds' peck responses to 20 repetitions of each of the ten test stimuli.

C. Results

The data set was submitted to an analysis of peck responses across high and low $f0$ for all test stimuli. For each bird, raw pecks were collected for each test trial. There is inherent variance in peck rates across individual birds. Therefore, total pecks to each test stimulus were summated across the 20 repetitions of the novel stimuli. Mean peck rates (i.e., pecks per 30-s trial) were calculated for both high- and low-$f0$ test stimuli. These means were then transformed into normalized peck rates by dividing peck rates to test

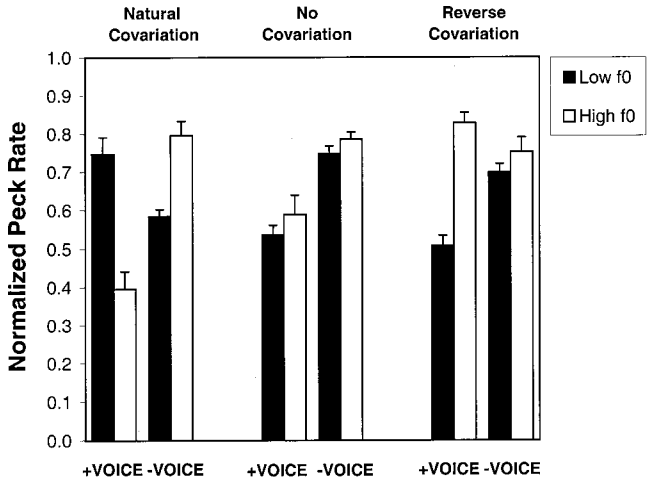


FIG. 2. Average normalized peck rates to low (105 Hz, black bars) versus high (141 Hz, white bars) $f0$. For each VOT$\times f0$ covariation condition (natural, reverse, control), data is presented for birds reinforced to peck to voiced stimuli (+voice) and for those reinforced to peck to voiceless stimuli (−voice).

stimuli by the individual quail's highest peck rate to the ten test stimuli. This transformation adjusted peck rates to a scale between zero and one for each bird, thus minimizing the variance that arises from the fact that some birds are ''heavier'' peckers than others (Bush *et al.*, 1993). This normalization method has been used previously to mitigate natural variance across individual animals (e.g., Lotto *et al.*, 1997).

Normalized mean peck rates and corresponding standard errors are presented in Fig. 2. Data are displayed by $f0$ (black bars correspond to $f0=105$ Hz, white bars show $f0=141$ Hz), sorted by condition (natural, control, reverse), and presented for +voice and −voice birds. Matched-pairs *t*-tests were computed for the difference between normalized peck rates to low- and high-$f0$ test stimuli for +/−voice birds in each condition.

1. Influence of ''natural'' f0/VOT covariation

Birds that were trained to peck in response to *voiced* consonants and heard natural covariation of $f0$ and VOT during training demonstrated a difference in their response to novel stimuli as a function of $f0$. These quail pecked significantly more ($t=4.80$, $p<0.01$) to novel, intermediate VOT series members synthesized with a *low* $f0$ (0.75 average normalized rate) than to the same stimuli synthesized with a *higher* $f0$ (0.39 average normalized rate). Quail trained to peck to *voiceless* consonants in the natural condition also exhibited a significant shift in behavior as a function of $f0$ ($t=3.02$, $p<0.01$), pecking more robustly to novel stimuli with a *higher* $f0$ (0.79 average normalized rate) than to those with a *lower* $f0$ (0.58 average normalized rate).

Thus, these birds' responses to novel stimuli mirrored natural covariation of $f0$ and VOT. Quail trained to peck in response to voiced stimuli pecked most vigorously to novel stimuli with a *low* $f0$, whereas quail trained to peck to voiceless stimuli responded most to stimuli with a *high* $f0$. This avian pattern of results mirrors data that have been observed in human perception (Chistovich, 1969; Haggard *et al.*, 1970; Fujimura, 1971; Massaro and Cohen, 1976, 1977; Haggard *et al.*, 1981; Whalen *et al.*, 1993; Castleman and Diehl, 1996). Stimuli with high $f0$ tend to be labeled as voiceless whereas otherwise similar stimuli with a low $f0$ are more often labeled as voiced.

These data demonstrate that the influence of $f0$ upon voicing need not arise from species-specific mechanisms. However, this single condition does not allow determination of whether the general mechanisms that may govern the interaction of $f0$ and voicing arise from auditory perceptual interactions or from experience with $f0$/VOT covariation. Quail in this condition experienced $f0$ and VOT covariation that modeled the covariation found among many of the world's languages. Although it is possible that experience with this pattern of covariation may have influenced their perception of novel stimuli, it is also possible that auditory interactions led them to respond to stimuli with a higher $f0$ as better exemplars of voiceless consonants than those with lower $f0$. Thus, it is necessary to turn to the behavior of quail in the remaining conditions to evaluate the relative influence of auditory constraints and learning.

2. Outcome of the control condition: No f0/VOT covariation

First, consider the control condition in which $f0$ and VOT did not covary during training. For stimuli presented to quail in this condition, $f0$ and voicing characteristics varied independently. If experience with covariation rather than auditory enhancement explains the effect of $f0$ upon VOT labeling for ''natural'' quail, there should be no effect of $f0$ upon responses of control quail in this condition. However, if the auditory system conspires to bias low-$f0$ stimuli to be responded to as short VOT stimuli, control quail should peck in a manner that mirrors effects typically observed for English listeners.

In fact, quail in the control condition did *not* demonstrate an influence of $f0$ upon their pecking behavior. With no covariation between $f0$ and VOT in the training stimuli, quail exhibited no effect of $f0$ on response to novel stimuli. Neither the birds trained to peck to *voiced* consonants ($t=0.56$, $p=0.30$, average normalized peck rates of 0.53 and 0.58 for low and high $f0$, respectively) nor those trained to respond to *voiceless* consonants ($t=1.05$, $p=0.18$, average normalized peck rates of 0.75 and 0.79 for low and high $f0$, respectively) were influenced by $f0$.

3. Influence of ''reverse'' f0/VOT covariation

Consider, now, the final condition. If experience with $f0$/VOT covariation is responsible for effects observed in the ''natural'' quail, then covariation in the opposite direction (i.e., low $f0$ paired with voiceless consonants) should influence birds' behavior in the opposite manner. In fact, birds trained to peck to *voiced* consonants for which covariation with $f0$ was in the direction opposite natural covariation exhibited a significant difference in their pecking behavior contingent on $f0$ ($t=4.26$, $p<0.01$), pecking most vigorously to stimuli with high $f0$ (0.83 average normalized rate) and less to stimuli with low $f0$ (0.51 average normalized rate). Birds' behavior mirrored the covariation inherent in

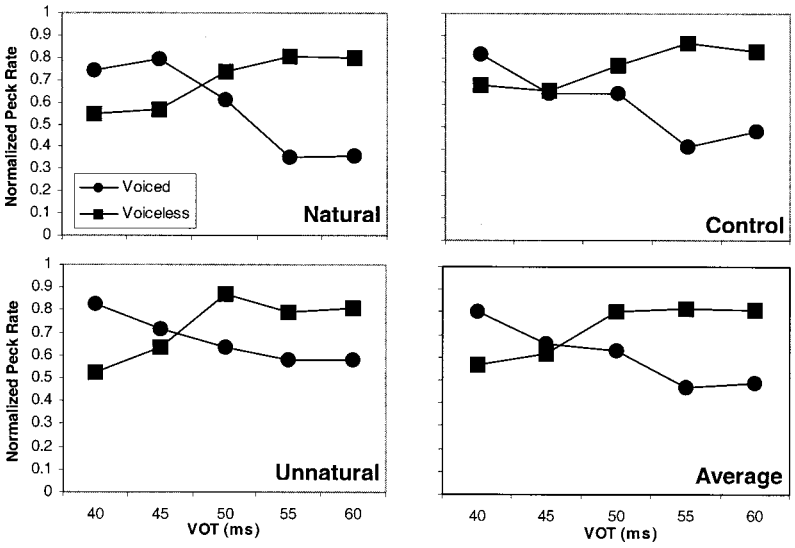


FIG. 3. Average normalized peck rates collapsed across $f0$ for each condition and averaged across conditions. Square symbols indicate −voice birds responses to novel test stimuli that vary across the VOT series. Circles correspond to +voice birds responses. These data indicate that quail did not perform the task simply by responding to changes in $f0$.

their training stimuli and was opposite the direction predicted by auditory enhancement hypothesis. Thus, these data suggest that the influence of $f0$ on VOT labeling is not bound to correspondence with the pattern typically observed in the world's languages, but rather is influenced by the pattern of $f0$/VOT covariation present in the speech input.

The results for birds trained to peck to *voiceless* consonants are less clear in that they did not exhibit a significant effect of $f0$ upon their response to novel stimuli ($t=.85$, $p=0.22$). Low $f0$ (0.70 average normalized rate) and high $f0$ (0.75 average normalized rate) did not differentially influence quails' pecking behavior to novel stimuli. It appears likely that the failure to find an influence of $f0$ upon birds' response to novel stimuli in this condition is related to rather high variability. Two of the four birds in this condition did exhibit a modest influence of $f0$ upon response to novel stimuli in the direction predicted by their experience. However, the remaining quail did not differentially respond as a function of $f0$.

To summarize these results across conditions and counterbalancing, birds in five of the six conditions tested (3 types of experience×2 mappings to voicing) exhibited behavior that supports the hypothesis that experience with covariation is fundamental to effects of $f0$ upon voicing perception.

4. Influence of VOT on birds' responses

The results are therefore in line with the hypothesis that the influence of $f0$ is related to experience with $f0$/VOT covariance. However, the data presented thus far have reported only the influence of $f0$. It is possible that these results reflect a tendency by birds to respond solely on the basis of $f0$, to the exclusion of VOT. If this is the case, then the results do not bear upon the hypotheses under investigation, but rather merely reflect the ability of quail to map $f0$ variation to a pecking response. To examine this possibility, it is necessary to inspect birds' responses across novel VOT stimuli, independent of $f0$. Figure 3, which illustrates these data, suggests that the birds' responses were sensitive to VOT. Data points in Fig. 3 correspond to average normalized peck rates to novel stimuli collapsed across the two test $f0$'s for +voice and −voice quail. Quail differentially responded to test stimuli varying in VOT, demonstrating that their behavior was not solely the result of sensitivity to $f0$. Note that the data points illustrated are drawn from the middle of the series and can thus be expected to be ''boundary'' stimuli. Broader identification functions are common for animal subjects. Typically, this characteristic is taken as indicative of attentional differences between animals and humans (e.g., Kuhl and Miller, 1978), but it also may be due to the animals' more limited experience with speech sounds. Quail trained to respond to +voice exhibited a very different pattern of response to novel test stimuli than did their −voice counterparts. This indicates that although $f0$ influenced quails' responses, they used both $f0$ and VOT to perform the task.

III. DISCUSSION

The objective of the present experiment was to examine the relative contributions of stable auditory perceptual interactions and experience with covariation in understanding the influence of $f0$ upon [VOICE] labeling. Many studies have demonstrated that listeners' categorization of synthetic or digitally manipulated natural speech varying perceptually from voiced to voiceless across a phonetic series can be shifted by changes in $f0$. Stimuli with lower $f0$'s are more often categorized as voiced consonants whereas stimuli with otherwise identical acoustic characteristics, but with higher $f0$'s, tend to be labeled as voiceless. This pattern of perception mimics patterns of speech production commonly observed across languages. Voiceless consonants tend to be produced with higher $f0$'s than their voiced counterparts.

The mechanisms behind this phenomenon are largely unknown, but there are at least two prominent hypotheses. The first hypothesis, in line with the tenets of auditory enhancement (Diehl and Kluender, 1989; Kingston and Diehl, 1994; Diehl *et al.*, 1995), suggests that general auditory interactions among mutually enhancing acoustic characteristics of $f0$ and VOT couple to improve the intelligibility of voiced consonants with low $f0$ and voiceless consonants

with high $f0$. Another possibility is that experience with covariation of $f0$ and voicing within the speech signal is responsible for effects of $f0$ upon [VOICE] labeling.

Using a nonhuman animal model, these two hypotheses were teased apart in the present experiment. Results from quail in the ''natural'' and ''reverse'' covariation conditions demonstrate that experience with $f0$/VOT covariance influences quails' response to novel stimuli. Quail that responded to +voice and −voice in the ''natural'' and those who pecked to +voice in the ''reverse'' condition pecked more often to $f0$/VOT combinations that matched their pattern of experience. One subset of these quail (those who were trained to peck to −voice stimuli in the ''reverse'' condition) did not adhere to this pattern of results. However, $f0$ had a null effect on responses to novel stimuli for these quail, so it is difficult to interpret these data. Overall, three of the four groups of quail in conditions where $f0$ covaried with VOT demonstrated an effect of $f0$ upon response to novel stimuli that mirrored the covariance experienced during training. The data of the +voice quail in the reverse condition, especially, are difficult to explain from an account that relies upon auditory interactions because they suggest $f0$ does not exert an obligatory influence on perception of [VOICE] consonants in the absence of covariation with VOT.

The results of the ''control'' condition complement these findings. Quail that did not experience regularity in $f0$/VOT covariation during training showed no shift in response to test stimuli contingent on $f0$. These results suggest that the influence of $f0$ is not strictly related to stable mutually enhancing interactions that are auditory in nature.

A. Ties to previous experiments

The present results may relate well to findings of some previous experiments that examined human listeners' perception of voicing as a function of $f0$. For example, Bernstein (1983) found that adult listeners make use of $f0$ in identifying words that vary in word-initial voicing (*gate* versus *Kate*), but 4- and 6-year-old children do not. These results are consistent with the hypothesis that experience may play a role in determining the influence of $f0$ upon perception of consonantal voicing.

Likewise, Haggard *et al.* (1981) reported an intriguing cross-linguistic difference that arose serendipitously from a study of the influence of $f0$ upon phoneme boundaries between voiced and voiceless consonants. One of their 35 listeners identified members of a series of stimuli varying in VOT quite differently from the rest of the listeners tested. This listener exhibited uncharacteristically flat identification functions across VOT, suggesting that she relied almost entirely upon the $f0$ variation and almost completely disregarded variation in VOT. Upon recalling the listener to determine whether she suffered from a hearing deficit, Haggard *et al.* found that the listener had been born in Italy and had emigrated to the United States at age five, though she had no command of Italian as an adult. Haggard *et al.* suggested that these odd data might underscore the importance of early learning in determining phoneme boundaries and acoustic cue weighting in speech perception. Massaro and Cohen (1976, 1977) have also reported marked individual differences in the influence of $f0$ on listeners' categorization of fricatives as /s/ versus /z/.

Although neither of these results (one a null finding for children, the other based on a single listener) is exceptionally strong evidence that experience plays a role in determining the influence of $f0$ upon perception of voicing, coupled with the findings of the present experiment, they hint at a role for experience and offer intriguing possibilities for further research.

Sinnott and Saporita (2000) recently have presented data from American English and Spanish adults as well as monkeys (*Macaca fuscata*) on the perceptual influence of first formant ($F1$) transition onset frequency covariation with gap duration in a speech series that varied from *say* to *stay*. Unlike English, Spanish does not have native consonantal cluster contrasts like *say* versus *stay*. Incremental increases in gap duration in Sinnott and Saporita's stimuli caused perception to change from *say* to *stay* for all subjects. However, subjects varied in their use of the $F1$ onset cue, which covaried with gap duration. American English listeners exhibited a strong influence of $F1$-onset frequency on stimulus identification. Spanish listeners differed in the degree to which they used $F1$ onset as a cue. Monkeys did not appear to use $F1$ onset at all. Sinnott and Saporita (2000) suggest that the important factor delineating these subject populations is degree of exposure to English and thus to $F1$ onset and gap duration covariation. These findings thus appear to suggest a perceptual learning component for another example of cue covariation in speech perception.

B. The generality of auditory mechanisms

There is at least one important criticism that could be leveled against the current experiment. The present results are much less clear if the avian auditory system of the Japanese quail is sufficiently different from the human auditory system so as not to capture putatively important characteristics that may contribute to auditory interactions between $f0$ and voicing. Though there are significant anatomical and physiological distinctions between avian and human auditory systems (see, e.g., Popper and Fay, 1980), there are several reasons to believe that the data presented here reliably represent audition quite generally. First, avian species have been shown to exhibit a number of effects in speech perception that rely upon audition, with no influence of learning (e.g., Kluender and Lotto, 1994; Dooling *et al.*, 1995; Dent *et al.*, 1997; Lotto *et al.*, 1997). These effects quite closely mirror human perceptual results, so despite distinctions between avian and human auditory systems, it appears that there is a great deal of functional correspondence. Furthermore, the present stimuli differed critically in VOT and $f0$—two acoustic cues that rely on rather low-frequency hearing. Previously, Kluender and Lotto (1994; Kluender, 1991) have shown that quail are quite capable of this sort of task, exhibiting effects of F1 on their response to voiced and voiceless stimuli. Thus, the conclusions drawn from the observation that quail in the ''control'' condition failed to exhibit an effect of $f0$ upon response to novel stimuli, although depen-

dent upon the functional correspondence of human and avian audition, are supported by these previous positive results.

Holt *et al.* (1999) have presented data from a mammalian model (the chinchilla, *Chinchilla villidera*) performing a similar task that provide further support for the present conclusions. Chinchillas' audiograms closely model those of humans (Henderson *et al.*, 1969). Furthermore, the properties of their auditory system are well mapped and, psychoacoustically, their behavior corresponds quite well with that of humans. For these reasons, chinchillas are one of the most common species used in systems auditory neurophysiology research. As a part of a larger project examining auditory cues to voicing, Holt *et al.* (1999) tested chinchilla perception of VOT as a function of changes in $f0$. Unlike quail in the ''natural'' and ''reverse'' conditions of the current experiment, chinchillas were not trained with covariance between $f0$ and VOT. Thus, the chinchillas' experience closely modeled that of quail in the ''control'' condition. Their pattern of response was also similar. Like control quail, chinchillas did not exhibit an effect of $f0$ upon response to voicing differences. Thus, the generalizability of the current results is supported by the fact that a mammalian species (in possession of an auditory system that more closely models that of humans) also fails to exhibit an effect of $f0$ upon VOT response when there is no history of experience with covariation between $f0$ and VOT.

C. Role of learning in speech perception

The present results should not be taken as evidence against an auditory enhancement account of speech perception. Though its emphasis is very clearly upon general auditory perceptual mechanisms, the auditory enhancement account does not fail to posit a role for learning. Diehl *et al.* (1990), for example, explicitly point out that a successful theory of speech perception must provide both an account of ''the transfer function of the auditory system'' as well as ''the listener's tacit knowledge of speech and language-specific facts that are relevant to phonetic categorization'' (p. 244). They go on to argue that it is presently possible to be much more unequivocal about influences of the auditory system than it is to be explicit about speech-relevant knowledge of listeners. As a consequence, they begin by seeking explanation ''in terms of general auditory mechanisms before appealing to speech-specific tacit knowledge'' (p. 245). We agree that the determination of the role of experience and learning ought to be a fundamental pursuit in understanding speech perception. Data from nonhuman species and those from nonspeech studies (e.g., Kluender *et al.*, 1998; Saffran *et al.*, 1999) suggest that very general learning processes may play an important role. Furthermore, phonetic categorization is unlikely to be the only domain where experience is important to speech perception.

An important lesson from the present work is that structured experience influences perception. The speech signal, in general, is richly structured with regularities imposed both by physical constraints on articulatory processes and by linguistic constraints that shape the habits of talkers. Experience with this structure shapes perception, and nonhuman animal models can contribute to our understanding of these processes.

ACKNOWLEDGMENTS

This work was supported in part by a National Science Foundation Predoctoral Fellowship to the first author. Additional support was provided by NSF Young investigator Award DBS-9258482 to the third author. Some of the data were presented at the 138th Meeting of the Acoustical Society of America in Columbus, OH. The authors thank Eric P. Lotto for his assistance in conducting the experiment and serving as the male voice upon which the experiment's stimuli were based. The authors also gratefully acknowledge the helpful comments of Randy L. Diehl, John Kingston, and Joan Sinnott. Correspondence and requests for reprints should be addressed to Lori L. Holt, Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213 (e-mail: lholt@andrew.cmu.edu).

[1]Although it is accurate to refer to the syllable-initial English voiceless stops that have mainly been examined in these studies as ''voiceless aspirated'' stops, they will be referred to here simply as ''voiceless'' stops for ease of reading.

[2]There are many acoustic cues that correlate with voicing. For example, presence of voicing during consonant constriction, low first formant ($F1$) near consonant constriction, absence of significant aspiration after consonant release, relatively short closure interval, and relatively long preceding vowel are all effective perceptual cues to the [VOICE] contrast. Lisker and Abramson (1964) demonstrated that the primary acoustic correlate of voicing is variation in voice onset time (VOT) in utterance-initial position. In its most precise usage, VOT refers to an articulatory characteristic—namely, the interval of time between the release of a stop consonant and the onset of voicing of the following vowel. Hereafter, we abandon the cumbersome phrase ''acoustic correlates of voice onset time'' and extend the usage of VOT to include the acoustic effects of the VOT as well, sacrificing precision but preserving readability. The *Stimulus* section (11.A.2) describes the precise acoustic cues synthesized to model voicing in the present experiment.

[3]There has been a good deal of debate about whether overall frequency of $f0$ or direction of $f0$ contour contributes more reliable cues to voicing (e.g., Lea, 1973; Umeda, 1981; Ohde, 1982; Silverman, 1986; Castleman and Diehl, 1996). Rather than modeling these more complex aspects of $f0$, the present stimuli had a flat $f0$. As a result, these stimuli might be thought to model initial or peak $f0$, each of which has been found to systematically vary as a function of voicing (Umeda, 1981). There is substantial evidence that stimuli synthesized with flat $f0$ contours influence listeners' perception of voicing.

[4]These distributions differ in some ways from what is typical of English voicing categories. For example, the modal VOT values for the voiced and voiceless distributions were 20 and 80 ms, respectively. The mid-point ''boundary'' for these stimuli was thus approximately 50 ms VOT whereas, in English, the labial VOT boundary is approximately 25 ms (Lisker and Abramson, 1964). This difference was tolerated in an effort to avoid inclusion of stimuli with VOT values less than or equal to 0 ms for fear that the 0-ms boundary might introduce unwanted discontinuities in the stimulus set. Another difference is that the standard deviations of our stimuli were equivalent across voiced and voiceless modes. Lisker and Abramson (1964) have shown that the standard deviations of VOT distributions depend on modal VOT; shorter VOT categories tend to have smaller standard deviations than longer VOT categories. This detail of distributions was not modeled here. Manipulations of both distribution modes and standard deviations may prove to be interesting variables for further research of how complex phonetic categories are learned. However, as a starting point, we chose to test the hypothesis in the most straightforward manner.

[5]Optimal performance was operationally defined as the highest ratio of pecks to positive versus negative stimuli. Birds are idiosyncratic in the amount of food deprivation necessary to achieve stable optimal perfor-

mance. Weights ranged from 85% to 100% of free-feed weight during training and testing.

Bernstein, L. (**1983**). ''Perceptual development for labeling words varying in voice onset time and fundamental frequency,'' J. Phonetics **11**, 383–393.

Bush, L. L., Hess, U., and Wolford, G. (**1993**). ''Transformations for within-subjects designs: A Monte Carlo investigation,'' Psychol. Bull. **113**, 566–579.

Caisse, M. (**1982**). ''Cross-linguistic differences in fundamental frequency perturbation induced by voiceless unaspirated stops,'' M. A. thesis, Univ. of California—Berkeley.

Castleman, W. A., and Diehl, R. L. (**1996**). ''Effects of fundamental frequency on medial and final [voice] judgments,'' J. Phonetics **24**, 383–398.

Chistovich, L. A. (**1969**). ''Variations of the fundamental voice pitch as a discriminatory cue for consonants,'' Sov. Phys. Acoust. **14**, 372–378.

Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., and Pierce, A. (**1997**). ''Perception of synthetic /ba/-/wa/ speech continuum by budgerigars (*Melopsittacus undulates*),'' J. Acoust. Soc. Am. **102**, 1891–1897.

Derr, M. A., and Massaro, D. W. (**1980**). ''The contribution of vowel duration, f_0 contour, and frication duration as cues to the /juz/-/jus/ distinction,'' Percept. Psychophys. **27**, 51–59.

Diehl, R. L., and Kluender, K. R. (**1989**). ''On the objects of speech perception,'' Ecological Psychol. **1**, 121–144.

Diehl, R. L., Castleman, W. A., and Kingston, J. (**1995**). ''On the internal perceptual structure of phonological features: The [voice] distinction,'' J. Acoust. Soc. Am. **97**, 3333–3334.

Diehl, R. L., Kluender, K. R., and Walsh, M. A. (**1990**). ''Some auditory bases of speech perception and production,'' in *Advances in Speech, Hearing, and Language Processing, Volume 1*, edited by W. A. Ainsworth (JAI, London), pp. 243–267.

Dooling, R. J., and Okanoya, K. (**1995**). ''The method of constant stimuli in testing auditory sensitivity in small birds,'' in *Methods in Comparative Psychoacoustics*, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhäuser Verlag, Basel, Switzerland), pp. 161–169.

Dooling, R. J., Best, C. T., and Brown, S. D. (**1995**). ''Discrimination of synthetic full-formant and sinewave/ra-la/continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*),'' J. Acoust. Soc. Am. **97**, 1839–1846.

Fujimura, O. (**1971**). ''Remarks on stop consonants: Synthesis experiments and acoustic cues,'' in *Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen*, edited by L. L. Hammerich, R. Jakobson, and E. Zwirner (Akademisk Forlag, Copenhagen), pp. 221–232.

Gruenenfelder, T. M., and Pisoni, D. B. (**1980**). ''Fundamental frequency as a cue to postvocalic consonantal voicing: Some data from speech perception and production,'' Percept. Psychophys. **28**, 514–520.

Haggard, M., Ambler, S., and Callow, M. (**1970**). ''Pitch as a voicing cue,'' J. Acoust. Soc. Am. **47**, 613–617.

Haggard, M., Summerfield, Q., and Roberts, M. (**1981**). ''Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading F_0 cues in the voiced-voiceless distinction,'' J. Phonetics **9**, 49–62.

Henderson, D., Onishi, S., Eldredge, D. H., and Davis, H. (**1969**). ''A comparison of chinchilla auditory evoked response and behavioral response thresholds,'' Percept. Psychophys. **5**, 41–45.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**1998**). ''Incorporating principles of general learning in theories of language acquisition,'' in *Chicago Linguistic Society, Volume 34: The Panels*, edited by M. Gruber, C. Derrick Higgins, K. S. Olson, and T. Wysocki (Chicago Linguistics Society, Chicago), pp. 253–268.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**1999**). ''Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?'' J. Acoust. Soc. Am. **106**, 2247(A).

Hombert, J. M. (**1978**). ''Consonant types, vowel quality, and tone,'' in *Tone: A Linguistic Survey*, edited by V. Fromkin (Academic, New York), pp. 77–111.

House, A. S., and Fairbanks, G. (**1953**). ''The influence of consonant environment on the secondary acoustical characteristics of vowels,'' J. Acoust. Soc. Am. **25**, 105–135.

Kingston, J. (**1986**). ''Are F_0 differences after stops deliberate or accidental?'' J. Acoust. Soc. Am. **79**, S27(A).

Kingston, J., and Diehl, R. L. (**1994**). ''Phonetic knowledge,'' Lang. **70**, 419–454.

Klatt, D. K. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.

Kluender, K. R. (**1991**). ''Effects of first formant onset properties on voicing judgments result from processes not specific to humans,'' J. Acoust. Soc. Am. **90**, 83–96.

Kluender, K. R., and Lotto, A. J. (**1994**). ''Effects of first formant onset frequency on [−voice] judgments result from general auditory processes not specific to humans,'' J. Acoust. Soc. Am. **95**, 1044–1052.

Kluender, K. R., Diehl, R. L., and Killeen, P. R. (**1987**). ''Japanese quail can learn phonetic categories,'' Science **237**, 1195–1197.

Kluender, K. R., Lotto, A. J., Holt, L. L., and Bloedel, S. L. (**1998**). ''Role of experience in language-specific functional mappings for vowel sounds as inferred from human, nonhuman and computational models,'' J. Acoust. Soc. Am. **104**, 3568–3582.

Kohler, K. J. (**1982**). ''F_0 in the production of lenis and fortis plosives,'' Phonetica **39**, 199–218.

Kohler, K. J. (**1984**). ''Phonetic explanation in phonology. The feature fortis/lenis,'' Phonetica **41**, 150–174.

Kohler, K. J. (**1985**). ''F_0 in the perception of lenis and fortis plosives,'' J. Acoust. Soc. Am. **78**, 21–32.

Kohler, K. J., and van Dommelen, W. A. (**1986**). ''Prosodic effects on lenis/fortis perception: Preplosive F_0 and LPSC synthesis,'' Phonetica **43**, 70–75.

Kuhl, P. K., and Miller, J. D. (**1975**). ''Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants,'' Science **90**, 69–72.

Kuhl, P. K., and Miller, J. D. (**1978**). ''Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli,'' J. Acoust. Soc. Am. **63**, 905–917.

Lea, W. (**1973**). ''Segment and suprasegmental influences of fundamental frequency contour,'' Conson. Types and Tone, South. Cal. Occasional Papers in Ling. **1**, 17–69.

Lehiste, I., and Peterson, G. E. (**1961**). ''Some basic considerations in the analysis of intonation,'' J. Acoust. Soc. Am. **33**, 419–425.

Lisker, L., and Abramson, A. S. (**1964**). ''A cross-linguistic study of voicing in initial stops: Acoustical measurements,'' Word **20**, 384–422.

Lotto, A. J., Holt, L. L., and Kluender, K. R. (**1999**). ''Structure of phonetic categories produced by general learning mechanisms,'' J. Acoust. Soc. Am. **106**, 2247.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1997**). ''Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*),'' J. Acoust. Soc. Am. **102**, 1134–1140.

Massaro, D. W., and Cohen, M. M. (**1976**). ''The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction,'' J. Acoust. Soc. Am. **60**, 704–717.

Massaro, D. W., and Cohen, M. M. (**1977**). ''Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction,'' Percept. Psychophys. **22**, 373–383.

Mohr, B. (**1971**). ''Intrinsic variations in the speech signal,'' Phonetica **23**, 65–93.

Morse, P. A., and Snowdon, C. T. (**1975**). ''An investigation of categorical speech discrimination by rhesus monkeys,'' Percept. Psychophys. **17**, 9–16.

Newman, R. S. (**1997**). ''Individual differences and the link between speech perception and speech production,'' unpublished doctoral dissertation, SUNY Buffalo.

Ohde, R. N. (**1982**). ''The effects of linguistic context on temporal and F0 properties of speech,'' J. Acoust. Soc. Am. **72**, LL5(A).

Ohde, R. N. (**1984**). ''Fundamental frequency as an acoustic correlate of stop consonant voicing,'' J. Acoust. Soc. Am. **75**, 224–230.

Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.

Peterson, N. R. (**1983**). ''The effect of consonant type on fundamental frequency and larynx height in Danish,'' Annu. Report Inst. Phon., Univ. Copenhagen **17**, 55–86.

Pitt, M. A., and McQueen, J. M. (**1998**). ''Is compensation for coarticulation mediated by the lexicon?'' J. Mem. Lang. **39**, 347–370.

Popper, A. N., and Fay, R. R. (**1980**). *Comparative Studies of Hearing in Vertebrates* (Springer-Verlag, New York).

Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996**). ''Statistical learning by 8-month-olds,'' Science **274**, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (**1999**). ''Statistical learning of tone sequences by human infants and adults,'' Cognition **70**, 27–52.
Silverman, K. E. A. (**1986**). ''F_0 segmental cues depend on intonation: The case of the rise after voiced stops,'' Phonetica **43**, 76–91.
Sinnott, J. M., and Saporita, T. A. (**2000**). ''Differences in American English, Spanish, and monkey perception of the *say-stay* trading relation,'' Percept. Psychophys. **62**, 1312–1319.
Stevens, K. N., and Blumstein, S. E. (**1981**). ''The search for invariant acoustic correlates of phonetic features,'' in *Perspectives on the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Erlbaum, Hillsdale, NJ), pp. 1–38.
Umeda, N. (**1981**). ''Influence of segmental factors on fundamental frequency in fluent speech,'' J. Acoust. Soc. Am. **70**, 350–355.
Waters, R. A., and Wilson, Jr., W. A. (**1976**). ''Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants,'' Percept. Psychophys. **19**, 285–289.
Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (**1993**). ''$F0$ gives voicing information even with unambiguous voice onset times,'' J. Acoust. Soc. Am. **93**, 2152–2159.

Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system

Catherine T. Best[a)]
Department of Psychology, Wesleyan University, Middletown, Connecticut 06459 and Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

Gerald W. McRoberts
Department of Psychology, Lehigh University, Bethlehem, Pennsylvania 18015 and Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

Elizabeth Goodell[b)]
Department of Psychology, University of Connecticut, Storrs, Connecticut 06269 and Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

(Received 18 October 1999; revised 4 October 2000; accepted 18 October 2000)

Classic non-native speech perception findings suggested that adults have difficulty discriminating segmental distinctions that are not employed contrastively in their own language. However, recent reports indicate a gradient of performance across non-native contrasts, ranging from near-chance to near-ceiling. Current theoretical models argue that such variations reflect systematic effects of experience with phonetic properties of native speech. The present research addressed predictions from Best's perceptual assimilation model (PAM), which incorporates both contrastive phonological and noncontrastive phonetic influences from the native language in its predictions about discrimination levels for diverse types of non-native contrasts. We evaluated the PAM hypotheses that discrimination of a non-native contrast should be near-ceiling if perceived as phonologically equivalent to a native contrast, lower though still quite good if perceived as a phonetic distinction between good versus poor exemplars of a single native consonant, and much lower if both non-native segments are phonetically equivalent in goodness of fit to a single native consonant. Two experiments assessed native English speakers' perception of Zulu and Tigrinya contrasts expected to fit those criteria. Findings supported the PAM predictions, and provided evidence for some perceptual differentiation of phonological, phonetic, and nonlinguistic information in perception of non-native speech. Theoretical implications for non-native speech perception are discussed, and suggestions are made for further research. [DOI: 10.1121/1.1332378]

PACS numbers: 43.71.Hw, 43.71.Es, 43.71.Ft [KRK]

I. INTRODUCTION

Adults' perception of speech contrasts is strongly influenced by experience with the phonological system of their native language (e.g., Abramson and Lisker, 1970). A traditional account for this phenomenon has been a perceptual version of the concept that a native-language ''phonological filter'' operates in production of non-native segments (Polivanov, 1931; Trubetskoy, 1939/1969). That is, it has been assumed that mature listeners have difficulty discriminating phonetic distinctions that do not occur as a native phonological contrast. Perhaps the most widely cited example of such perceptual difficulty is the poor discrimination of English /r/-/l/ by speakers of languages that lack this contrast, such as Japanese (e.g., Goto, 1971; Miyawaki *et al.*, 1975; Mochizuki, 1981; Best and Strange, 1992; MacKain *et al.*, 1981). Similarly poor non-native speech perception performance has been documented for speakers of other languages. For example, English speakers have difficulty discriminating contrasts such as Hindi retroflex versus dental stops and Nthlakampx velar versus uvular ejectives (Werker *et al.*, 1981; Werker and Tees, 1984). Discrimination in such cases is near chance, in striking contrast to the ceiling-level performance typically found with native language distinctions.

What is it about native language experience that results in such difficulties with non-native speech discrimination? One type of explanation emphasized exposure in early development as being critical to the ''tuning'' of relevant sensorineural mechanisms. For example, some argued that innate, linguistically specialized neural mechanisms, initially tuned to universal settings of phonetic categories and/or boundaries, are modified by early exposure to specific phonetic features (e.g., Eimas, 1975, 1991). Others posited the nonlinguistic view that early exposure to specific acoustic properties maintains or enhances the tuning of prewired psychophysical mechanisms that respond selectively to those properties (e.g., Aslin and Pisoni, 1980). Such prewired mechanisms are generally assumed, by the latter view, to be components of general auditory processing skills that are

[a)]Electronic mail: cbest@wesleyan.edu
[b)]Present address: Landmark School, 429 Hale St., Prides Crossing, MA 01965.

part of our mammalian (or vertebrate) evolutionary endowment (e.g., Kuhl, 1988; Dooling, 1989).

It has since become apparent, however, that neither account of critical early tuning can adequately explain all aspects of adults' non-native speech perception. Numerous studies have shown that discrimination of unfamiliar phonetic contrasts can be improved even in adults through extensive natural experience, intensive laboratory training, or experimental manipulations that reduce task memory demands (e.g., Logan *et al.*, 1991; Lively *et al.*, 1993; Pisoni *et al.*, 1982; MacKain *et al.*, 1981; Strange and Dittman, 1984; Werker and Logan, 1985; Werker and Tees, 1984). That is, exposure need not occur early in development—even limited exposure in adulthood can improve performance to some extent.[1] In response to those findings, some have proposed that language experience affects higher-level processes that remain malleable, such as phonological encoding or memory retention, rather than lower-level sensorineural responsivities that are relatively permanently changed by early experience (Tees and Werker, 1984; Werker and Tees, 1984).

Further damaging to the sensorineural tuning hypotheses is the fact that early exposure to specific phonetic or acoustic properties, or lack thereof, does not guarantee good versus poor discrimination, respectively. For example, discrimination of non-native stop voicing contrasts is poor in American English listeners (Abramson and Lisker, 1970), despite the fact that the range of voice onset times (VOTs) involved is amply manifested in English stop allophones (cf. MacKain, 1982), and adults' discrimination levels for non-native contrasts are not systematically related to whether or not the associated phonetic features occur within the listeners' native language (Polka, 1992). Native English speakers' discrimination of four Hindi dental-retroflex stop consonant contrasts differing in voicing (prevoiced, voiceless unaspirated, voiceless aspirated, and breathy voiced) varied between poor and excellent, irrespective of whether the specific voicing type occurs in English (Polka, 1991). Conversely, adults discriminate certain non-native contrasts quite well, even with virtually no prior exposure to their distinctive phonetic-acoustic features in speech. American English listeners show very good to excellent discrimination of Zulu click consonants, despite their lack of experience or training with such clicks (Best *et al.*, 1988).

Two important conclusions can be drawn from these findings: Adult discrimination of non-native speech contrasts is *not* uniformly poor, as the classic view would have it. Instead, we see wide variation in performance level, from poor to excellent, that does not depend on the presence or absence of the critical phonetic/acoustic features in native speech.

If early sensorineural tuning fails to explain the variation in non-native speech discrimination, then what does account for it? Several recent theoretical models posit that native speech experience provides an organizing perceptual framework that shapes discrimination of unfamiliar speech contrasts. Best's perceptual assimilation model (PAM: Best, 1994a, b, 1995; Best *et al.*, 1988), Flege's speech learning model (SLM: Flege, 1986, 1990, 1995; cf. Guion *et al.*, 2000), and Kuhl's Native Language Magnet model (NLM: Grieser and Kuhl, 1989; Iverson and Kuhl, 1996; Kuhl, 1991, 1992; Kuhl *et al.*, 1992) all presume that adults' discrimination of non-native speech contrasts is systematically related to their having acquired a native speech system. However, the models differ in how they conceive of the native perceptual framework, as in ongoing theoretical debates about speech perception. One basic view is that speech perception depends on the same general-purpose auditory processes employed for perception of nonspeech sounds (e.g., Diehl and Kluender, 1989; Kluender, 1994; Kuhl, 1988; cf. theorical overview in Best, 1995). An opposing view is that a specialized linguistic-phonetic module is involved in perception of speech alone (Liberman *et al.*, 1967; Liberman and Mattingly, 1989). Also under debate is whether the perceptual mechanisms, whether general or specialized, operate on acoustic (e.g., Diehl and Kluender, 1989; Kuhl, 1988, 1991, 1992) or articulatory information (e.g., Best, 1995; Fowler, 1986, 1989; Liberman and Mattingly, 1989).

SLM addresses primarily how adult speakers acquire phonological segments for a second language (L2), particularly in production and particularly by relatively experienced L2 speakers. It proposes that non-native phones are ''equivalence-classified'' relative to native language (L1) phonemes on the basis of phonetic similarity. New L2 phonological categories are more likely to be developed, hence produced (and perceived) fairly accurately, the more dissimilar they are from the closest native phonemes. SLM remains neutral regarding general versus specialized mechanisms and extraction of acoustic versus linguistic-phonetic information from speech.

NLM instead proposes that early in life, listeners develop acoustic prototypes for native phonemic categories. NLM assumes that speech perception involves general auditory mechanisms that process acoustic rather than specifically phonetic information. In NLM, native prototypes have magnetlike effects, in which the nearby perceptual space is ''shrunk,'' making it more difficult to discriminate phonetic variation around prototypes than around non-prototypes, or poor exemplars, of the same category. So NLM, unlike SLM, predicts an asymmetry for discriminating prototypical versus non-prototypical stimuli. Listeners fail to develop prototypes for non-native categories, due to lack of relevant acoustic experience. Hence within-category discrimination for non-native phones is expected to be uniform rather than asymmetrical.

Both SLM and NLM have contributed importantly to our understanding of the perception of non-native speech segments, and have generated substantial research. In a quest to account for discrimination of nonnative *contrasts*, however, they both have limitations. SLM's primary shortfall is its focus on individual phonemes—it makes no explicit predictions about discrimination of non-native contrasts (Flege, personal communication). As for NLM, recent concerns have been raised that the perceptual magnet effect may not be robust across listener groups (Lively, 1993; cf. Frieda *et al.*, 1999). Moreover, given listeners' frequent identification of the ''non-prototype'' stimuli as exemplars of a different category altogether than the prototype, the asymmetry in dis-

crimination may actually reflect better between-category than within-category discrimination (Lotto *et al.*, 1996; Sussman and Lauckner-Morano, 1995), that is, the classic phenomenon of categorical perception (Lotto *et al.*, 1998). Additionally, NLM espouses basic principles of the early tuning accounts which, as we summarized earlier, are problematic. Neither SLM nor NLM comprehensively explains the variations in non-native discrimination that cannot be traced to the presence versus absence of features in the listeners' language, as discussed earlier.

Those theoretical gaps are addressed by the perceptual assimilation model (PAM). That model was originally developed to account for the previously unexpected finding that American English listeners discriminate Zulu clicks quite well; the authors hypothesized that this was due to the fact that they had perceived the clicks as nonspeech sounds (Best *et al.*, 1988). Of the three non-native speech perception models, only PAM makes explicit predictions about assimilation and discrimination differences for diverse types of non-native contrasts. And PAM alone incorporates principles of phonological theory, the branch of linguistics that concerns the linguistic function and structure of the native system of phonological contrasts. The specific phonological theory that PAM draws from is articulatory phonology (Browman and Goldstein, 1986, 1989, 1990a, b, 1992), which is compatible with PAM's direct realist (ecological) position that what listeners detect in speech is information regarding the articulatory gestures that generated the signal (e.g., Best, 1995; Fowler, 1986, 1989; Fowler *et al.*, 1990). Gestures are defined by the articulatory organs (active articulator, including laryngeal gestures), constriction locations (place of articulation), and constriction degree (manner of articulation) employed. We will use the term ''native phoneme'' to refer to a functional equivalence class of articulatory variants that serve a common phonological function, as evidenced by their contribution to distinguishing lexical items, identifying morpho-syntactic units, and participating in other phonological alternations such as context-conditioned allophony. PAM posits that non-native speech perception is strongly affected by listeners' knowledge (whether implicit or explicit) of native phonological equivalence classes, and that listeners perceptually assimilate non-native phones to native phonemes whenever possible, based on detection of commonalities in the articulators, constriction locations and/or constriction degrees used (Best, 1993, 1994a, b, 1995).

According to PAM (see Best, 1995), a given non-native phone may be perceptually assimilated to the native system of phonemes in one of three ways: (1) as a **C**ategorized exemplar of some native phoneme, for which its goodness of fit may range from excellent to poor; (2) as an **U**ncategorized consonant or vowel that falls somewhere in between native phonemes (i.e., is roughly similar to two or more phonemes); or (3) as a **N**onassimilable nonspeech sound that bears no detectable similarity to any native phonemes. Adults' discrimination of a non-native contrast is predicted to depend on how each of the contrasting phones is assimilated. Several pairwise assimilation types are possible. The non-native phones may be phonetically similar to two different native phonemes and assimilate separately to them, which was termed **T**wo **C**ategory assimilation[2] (**TC**). Both may, instead, assimilate equally well or poorly to a single native phoneme, termed **S**ingle **C**ategory assimilation (**SC**). Or both might assimilate to a single native phoneme, but one may fit better than the other, termed a **C**ategory **G**oodness difference (**CG**). Alternatively, one non-native phone may be **U**ncategorized, as defined above, while the other is **C**ategorized, forming an **U**ncategorized–**C**ategorized pair (**UC**). Or both non-native phones might be **U**ncategorized speech segments (**UU**). Finally, the two phones' articulatory properties may both be quite discrepant from any native phonemes, and be perceived as **N**on-**A**ssimilable (**NA**) nonspeech sounds.

As for discrimination of non-native contrasts, it can be hindered, aided, or unaffected by native phonology, depending on how the non-native phones relate to native phonemes and contrasts (Best, 1994a, 1995). Native phonology should aid discrimination when the two phones are separated by native phonological boundaries, but should hinder it when both phones assimilate to the same native phoneme. However, discrimination of non-native elements that are heard as nonspeech sounds is neither helped nor hindered by native phonology. NA contrasts, therefore, are predicted to show good to excellent discrimination, depending on their perceived differences as nonspeech sounds. However, TC and UC contrasts should be discriminated quite well because in both cases the contrasting phones fall on opposite sides of a native phonological boundary. On the other hand, with CG and SC types, both phones assimilate to the same native phoneme, so discriminability is hindered by native phonology. If one phone is a good fit and the other is poor, discrimination will be very good (CG difference), but not as good as in TC contrasts because it is hindered by assimilating to a single native phoneme. In SC cases, both non-native phones are equivalent in phonetic goodness, hence discrimination is poor, hindered both by lack of phonological contrast and by lack of difference in fit. For example, Japanese speakers are likely to assimilate English /r/ and /l/ as poor examples of a single Japanese phoneme (/r/ or perhaps /w/: Best and Strange, 1992; Takagi and Mann, 1995; Yamada and Tohkura, 1992), and discriminate the /r/-/l/ contrast poorly. For uncategorized-uncategorized (UU) assimilations, discrimination is less strongly affected by native phonological equivalence classes, and should range between fair and good, dependent on perceived similarity of the non-native phones to each other and to the set of nearby native phonemes.

The current research focuses on those contrasts involving only non-native phones that are perceptually categorized to native phonemes, as defined above, that is, the TC, CG, and SC assimilation types. PAM predicts the following gradient of discrimination levels for these: TC>CG>SC (Best, 1994a, 1995). PAM's predictions about each of these assimilation types have been supported by a number of cross-language perception studies (see Best, 1994a, b, 1995). As noted, English speaking adults fail to assimilate Zulu click consonants to English consonants, instead perceiving them as nonspeech sounds, consistent with the NA pattern. In keeping with PAM's predictions about non-native NA contrasts, discrimination of the clicks is good to very good (Best

et al., 1988). English listeners' perception of clicks as non-speech is supported by recent evidence that whereas Zulu listeners show right ear superiority for click discrimination in a dichotic listening task, presumed to reflect left hemisphere language specialization, American English listeners do not (Best and Avery, 1999). Moreover, English-learning infants fail to show a developmental decline in discrimination of the clicks by 10–12 months (Best *et al.*, 1988) comparable to that found for other non-native consonant contrasts (e.g., Werker, 1989; Werker and Pegg, 1992; Werker *et al.*, 1981; Werker and Lalonde, 1988). In particular, 10–12 month olds discriminated a click contrast but failed to do so with a non-native contrast from Werker *et al.* (1981) on which adults' perception had been consistent with SC assimilation (Best *et al.*, 1995). In cross-language studies of adults' non-native speech perception, Japanese listeners displayed SC assimilation of American English /r/-/l/ and CG assimilation of English /w/-/r/, with better discrimination of the latter, as predicted by PAM (Best and Strange, 1992). French listeners categorized and discriminated English /w/-/r/ in a CG pattern, consistent with French and English /r/ articulatory differences (Hallé *et al.*, 1999).

Studies from other research groups also are consistent with certain PAM predictions. In her study of English listeners' perception of four Hindi dental-retroflex stop contrasts differing in voicing type, Polka (1991) reported that, based on listeners' descriptions of the contrasts, SC-type assimilations were associated with lower discrimination performance than TC-type assimilations, as PAM predicts. She also found that English listeners tended to assimilate Farsi voiced velar versus uvular stops (/g/-/G/) as a CG contrast and Salish velar versus uvular ejectives (/k′/-/q′/) as a SC (or NA[3]) contrast, with a tendency toward better discrimination of the former distinction, as would be expected according to PAM (Polka, 1992).[4] In another recent study, Japanese listeners' discrimination of UU and UC assimilations of English consonant contrasts fit PAM predictions in all but one UC case (Guion *et al.*, 2000). Interestingly, two studies of early bilinguals revealed poor discrimination of contrasts that fit a SC pattern with respect to the L1, but a TC pattern with respect to the L2, suggesting long-term effects of L1 phonological organization even in listeners who have been fluent in the L2 from a young age (Calderón and Best, 1996; Pallier *et al.*, 1997).

However, no findings have yet been published on non-native contrasts that clearly fit the TC pattern, i.e., in which neither non-native phone is a good match for a native phoneme yet the both are perceptually assimilated to two different native phonological classes. Evidence on the TC pattern is important, given that the predicted excellent, nativelike (or nearly so) levels of discrimination and categorization performance would be quite unexpected according to the more traditional assumption that adults should have difficulty discriminating any contrasts that do not occur in the native language (Polivanov, 1931). Relatedly, reports are still lacking on systematic comparison of TC, CG, and SC assimilation types needed to evaluate PAM's strong prediction for significantly better discrimination of TC than CG assimilation types, which in turn should show better discrimination of SC types. Alternative outcome patterns remain possible for those three types of non-native contrast. One is that there could be equally poor discrimination for the three types of contrast, as suggested by traditional claims about non-native speech perception. As summarized earlier, this outcome is highly unlikely in light of previous findings that discrimination levels can differ substantially among non-native contrasts. Another possibility might be that discrimination differences could be determined by some other factor, such as acoustic differences among the contrasts, and not by their phonological assimilations. To evaluate these possibilities, Experiment 1 systematically compared discrimination levels among non-native contrasts that were expected to yield TC, CG, and SC assimilation patterns.

II. EXPERIMENT 1

To optimize comparisons of performance among SC, CG, and TC assimilations, all three stimulus contrasts were taken from a single language, Zulu. None were phonological contrasts in English. All three were differentiated by laryngeal gestures. The goal was to include one non-native contrast that American English (AE) listeners were likely to assimilate to two contrasting English phonemes (TC), another that they should assimilate as a noticeable category goodness difference within a single English phoneme (CG), and a third that they should assimilate with nearly equal fit to a single phoneme in English (SC). The following contrasts were selected, based on their articulatory-phonetic characteristics relative to English (Ladefoged and Maddieson, 1996; Maddieson, 1984; Ruhlen, 1975):

(i) voiceless versus voiced lateral fricatives (/ɬ/-/ɮ/);
(ii) voiceless aspirated versus ejective (glottalized) velar stops (/k^h/-/k′/);
(iii) plosive versus implosive voiced bilabial stops (/b/-/ɓ/).

The Zulu lateral fricative contrast uses a place of articulation that is non-native for AE fricatives, though the articulatory organs and constriction locations involved are similar to AE /l/. Voiceless–voiced fricative distinctions do occur in AE at other constriction locations. In both languages, fricative voicing contrasts are signaled by a laryngeal gesture of glottal abduction (voiceless), versus a glottal setting that results in vocal fold vibration (voiced), during frication at the supralaryngeal constriction location. In the other two Zulu contrasts, the laryngeal distinction itself rather than the constriction location was non-native. Location for the Zulu velar stop constriction corresponds to that for AE /k/. The glottal abduction for Zulu /k/ makes it essentially identical to AE /k/; both are narrowly transcribed as [k^h], i.e., long-lag voice onset with positive airflow through the open glottis during release of velar closure, resulting in aspiration. The distinctive laryngeal gesture for the contrasting ejective /k′/ is a glottal adduction, with a resulting (near-)cessation of glottal airflow during release of the velar stop closure. The latter laryngeal gesture is not used in utterance-initial AE stops (although some speakers produce ejectives in forceful releases of utterance-final voiceless stops). As for the third

Zulu contrast, the glottal setting is similar in Zulu and AE /b/, in that Zulu /b/ displays a short unaspirated voicing lag (i.e., [p]), as is the case for the common [p] allophone of AE /b/ (which can also be realized as fully voiced [b]). The implosive Zulu /ɓ/ involves voicing during bilabial closure and release (as in the AE [b] allophone), but adds a simultaneous rapid lowering of the larynx, which causes a brief negative airflow during release. Larynx lowering is not used distinctively in AE; but the fact that voicing continues during Zulu /ɓ/ release makes the implosive gesturally quite similar to voiced AE /b/ in both location of supralarygneal constriction and basic glottal setting.

Based on articulatory similarities and differences between the Zulu consonants and the most closely corresponding AE consonants, the following assimilation predictions were made: The lateral fricatives were expected to show two category (TC) assimilation by most AE listeners, as some phonological contrast in English, such as a voiceless apical fricative (e.g., /θ s ʃ/, perhaps clustered with /l/) versus /l/ (voiced lateral approximant) or some voiced apical fricative (e.g., /ð z ʒ/, perhaps clustered with /l/), which involve the same articulators (tongue tip and dorsum, glottis), constriction locations (dental/alveolar and posterior constrictions), and constriction degree (fricative) as these Zulu consonants. The velar stops were expected to show a notable category goodness difference (CG) in assimilation to good versus poor AE /k/, that is, to a native consonant involving the same articulatory organs (tongue dorsum, glottis), constriction location (velar), and degree (stop). The bilabial stops were expected to show single category (SC) assimilation as nearly equivalent exemplars of AE /b/ (same organs, constriction location and degree), at least for most listeners. The associated discrimination predictions were that performance would be excellent for the lateral fricatives, quite good but significantly lower for the velar stops, and substantially poorer for the bilabial stops, i.e., TC>CG>SC.

Discrimination of each contrast was tested before assimilation was assessed, to minimize the potential influence that labeling or describing the Zulu consonants may have had on discrimination. Given Werker and Logan's finding (1985) that short-term memory constraints affected English listeners' discrimination of a Hindi contrast that fits the SC definition, but not Hindi listeners' discrimination of the same contrast (native ~ TC), we also assessed whether this factor might influence discrimination more for SC assimilations than for TC assimilations. A difference in memorial influences would further support the differentiation of SC and TC contrasts. It must be noted, however, that neither PAM nor the other two non-native speech perception models (SLM and NLM) make explicit predictions about memorial effects on non-native speech discrimination.

A. Method

1. Participants

The listeners were 22 native speakers of American English (15 female, 7 male) with a mean age of 18.4 years (range=18–20 yr), recruited from an Introductory Psychology subject pool. Participants received course points for their participation. None of the participants had experience with Zulu or any other languages employing the consonant contrasts used in this study. None had a personal or family history of developmental speech, language, or reading disorders. Five other participants were tested but their data were later removed due to ear infection on the test day ($n=1$), delayed language development ($n=1$), or familial speech problems ($n=1$) or reading impairments ($n=2$).[5]

2. Stimulus materials

An adult female native Zulu speaker from Durban, South Africa, was recorded producing multiple tokens of each of the six target consonants in CV nonsense syllable pairs. The syllables used were [ɮɛ]-[ɬɛ] (lateral fricatives), [k^ha]-[k′a] (velar stops), and [bu]-[ɓu] (bilabial stops).[6] All syllables had high tone on the vowel.[7] The syllables were read aloud individually from a randomly ordered list containing 20 repetitions of each.

The recording was digitized on a VAX 11-780 computer using the Haskins Laboratories' Pulse Code Modulation (PCM) system (Whalen *et al.*, 1990). Individual syllables were extracted and acoustically analyzed by the third author, using a signal analysis program called HADES, which was developed at Haskins Laboratories. The measures included the durations of consonant noise, vowel, and full syllable; rms amplitude of the consonantal noise; VOT (for stops only); spectral centroid values at 15%, 50%, and 85% into the consonant noise; and frequency peaks for $F0$ and each of the first three formants at 15%, 50%, and 85% into the vowel. ''Centroid'' refers to the spectral center of gravity, or amplitude-weighted mean frequency, calculated as the first moment of a DFT. Centroid values primarily reflect front cavity size and configuration (see Nittrouer *et al.*, 1989). Only those tokens that our Zulu speaker identified in a listening task as unequivocal productions of each category were further considered. Also, any tokens displaying list intonation effects or other odd voice qualities were ruled out as potential stimuli for the perceptual tests. Six tokens were then selected per category, matched as closely as possible between the contrasting syllables of each pair for overall duration, fundamental frequency and contour, and vowel formant frequencies at the 50% point. The first author then independently remeasured the acoustic properties of the selected stimuli (see final values, Table I), using the Signalyze program (Keller, 1994) on a Macintosh computer. Note, however, that centroid values (HADES) could not be computed in Signalyze, and $F0$ and formant measures were added for the first pitch pulse of each stimulus. Note also that final formant values (Signalyze) were based on fast Fourier transforms (FFTs) rather than linear predictive coding (LPC) estimation (as in HADES), and that final $F0$ values (Signalyze) were calculated as the inverse of the period of the glottal pulse nearest to the designated time slice.

In summarizing the acoustic analyses, we use the term ''discrete difference'' to describe measures that showed no overlap in range of values between the contrasting consonants, and the term ''overlapping difference'' for measures whose values showed partially overlapping ranges, thus inconsistently differentiating the contrast. For the remaining measures, the range of values for one category was com-

TABLE I. Acoustic attributes of the stimulus tokens selected for each target category in experiment 1.

Stimulus syllable	Syllable duration[a]	Consonant duration[a]	Vowel duration[a]	Consonant VOT[a]	Consonant amplitude[b]	Consonant centroid[c]	Vowel $F1$[d]	Vowel $F2$[d]	Vowel $F3$[d]	Vowel $F0$[d]
Lateral fricatives										
/ɮɛ/ (voiced)	**309.9** (271–336)[e]	**108.3** (86–131)	**197.6** (185–217)		**25.3** (22–28)					
first pulse[f]							**475.2** (436–499)	**2318.9** (2259–2384)	**2813.4** (2623–2951)	**204.8** (153–233)
15%[g]						**3512.3** (3239–3722)	**528.0** (499–592)	**2443.6** (2384–2586)	**2812.1** (2638–2943)	**221.5** (214–232)
50%[g]						**3616.5** (3549–3767)	**563.0** (554–608)	**2560.4** (2493–2586)	**2875.5** (2676–2996)	**224.0** (215–233)
85%[g]						**3683.5** (3572–3805)	**550.5** (514–592)	**2539.7** (2477–2586)	**2873.0** (2780–2973)	**226.0** (220–231)
/ɬɛ/ (voiceless)	**345.2** (299–400)	**151.7** (134–179)	**187.9** (161–216)		**31.8** (24–36)					
first pulse							**542.7** (530–577)	**2279.5** (2228–2322)	**2850.6** (2810–2899)	**246.2** (212–267)
15%						**3654.9** (3621–3685)	**579.1** (561–592)	**2374.1** (2322–2430)	**2800.9** (2705–2907)	**240.4** (236–244)
50%						**3704.2** (3636–3810)	**592.1** (577–608)	**2495.5** (2337–2571)	**2818.3** (2690–2936)	**232.5** (227–239)
85%						**3690.3** (3550–3810)	**555.7** (514–577)	**2557.8** (2509–2633)	**2900.3** (2847–2959)	**229.8** (221–238)
Velar stops										
/**k^{h}a**/ (aspirated)	**284.8** (246–314)	**8.5** (7–11)	**202.8** (158–234)	**82.5** (76–88)	**30.2** (28–33)					
first pitch pulse							**1075.7** (1006–1162)	**1546.4** (1468–1655)	**2439.5** (2325–2534)	**225.2** (201–262)
15%						**2059.1** (1841–2423)	**1076.9** (991–1177)	**1519.1** (1431–1610)	**2505.3** (2392–2646)	**209.7** (201–219)
50%						**1972.1** (1876–2268)	**1063.2** (1014–1125)	**1442.1** (1379–1535)	**2523.9** (2467–2673)	**197.6** (185–204)
85%						**1913.7** (1814–2164)	**1049.6** (1006–1095)	**1525.3** (1438–1714)	**2632.0** (2556–2750)	**195.5** (188–203)
/**k′a**/ (ejective)	**263.5** (231–278)	**37.3** (23–55)	**175.4** (154–194)	**88.3** (70–113)	**42.3** (38–47)					
first pitch pulse							**1063.3** (1029–1117)	**1455.8** (1319–1528)	**2514.0** (2400–2661)	**197.1** (134–233)
15%						**2580.3** (2391–2885)	**1104.2** (1029–1178)	**1604.8** (1528–1729)	**2552.5** (2467–2623)	**203.0** (185–213)
50%						**2616.5** (2408–2788)	**1099.3** (1050–1170)	**1584.9** (1431–1729)	**2624.6** (2430–2772)	**198.3** (189–204)
85%						**2709.3** (2477–2937)	**1112.9** (1038–1200)	**1578.7** (1498–1692)	**2613.4** (2452–2743)	**194.9** (189–200)
Bilabial stops										
/**bu**/ (plosive)	**261.4** (232–278)	**9.7** (6–13)	**248.3** (211–267)	**13.5** (10–23)	**15.8** (13–17)					
first pitch pulse							**463.4** (443–499)	**895.0** (836–919)	**2461.7** (2337–2586)	**227.5** (211–242)
15%						**2806.8** (2284–3096)	**475.1** (443–494)	**935.1** (886–967)	**2365.5** (2307–2448)	**230.5** (213–239)
50%						**2707.0** (2400–3220)	**476.8** (453–494)	**935.1** (866–987)	**2387.4** (2317–2569)	**228.5** (213–244)
85%						**2758.1** (2378–3216)	**478.5** (463–494)	**955.3** (917–987)	**2424.3** (2226–2851)	**232.0** (225–239)

TABLE I. (*Continued*.)

Stimulus syllable	Syllable duration[a]	Consonant duration[a]	Vowel duration[a]	Consonant VOT[a]	Consonant amplitude[b]	Consonant centroid[c]	Vowel $F1$[d]	Vowel $F2$[d]	Vowel $F3$[d]	Vowel $F0$[d]
/ɓ**u**/ (implosive)	**293.8** (260–341)	**12.1** (7–33)	**221.5** (204–252)	**−58.7** (−105/−67)	**26.7** (23–31)					
first pitch pulse							**450.2** (405–483)	**907.1** (873–927)	**2472.2** (2307–2680)	**230.1** (219–255)
15%						**2637.8** (2389–2907)	**529.8** (499–577)	**877.7** (810–982)	**2383.9** (2290–2493)	**254.2** (238–270)
50%						**2591.5** (2250–3022)	**496.1** (483–514)	**981.9** (935–1044)	**2391.7** (2275–2462)	**246.4** (227–256)
85%						**2287.5** (1982–2782)	**493.4** (483–499)	**937.5** (826–997)	**2370.9** (2275–2462)	**237.8** (219–254)

[a]Duration in milliseconds (ms); consonant duration refers to burst for velar and bilabial stops, to frication for lateral fricatives; VOT=voice onset time for stops only.
[b]Root mean square amplitude across full duration of consonant noise (frication or burst).
[c]Centroid frequencies of consonant noise at onset, middle, and offset of noise (frication or burst).
[d]Frequency in Hz at first pitch pulse and at 15%, 50%, and 85% into vowel for lowest three formants and $F0$ (LPC estimates).
[e]Minimum and maximum values, to nearest integer.
[f]First pitch pulse of vocalic portion.
[g]Percent into consonant noise (frication or burst) for centroid measures; percent into vocalic portion for formant and $F0$ measures.

pletely subsumed within the range of the other (or nearly so). These latter measures are designated as ''no difference'' even if the means appear to diverge between the contrasting categories, because of the complete overlap in range.

The acoustic differences for each stimulus contrast are consistent with their productions (e.g., aerodynamic differences). For the lateral fricatives, the discrete differences were that the voiceless syllables showed a higher $F1$ at the first vocalic pitch pulse (M_{diff}=67.5 Hz), a higher $F0$ at 15% into the vowel (M_{diff}=18.9 Hz), and a longer duration of frication (M_{diff}=43.4 ms, or 40% longer than voiced fricatives). The overlapping differences were that voiceless stimuli displayed inconsistently higher mean amplitude (M_{diff}=6.5 rms), higher mean centroid frequency at 50% into the frication (M_{diff}=87.7 Hz), higher $F2$ at 15% into the vowel (M_{diff}=69.5 Hz), higher $F0$ at the first pitch pulse and at 50% into the vowel (M_{diff}= 41.4 and 8.5 Hz, respectively), and longer mean duration for the full syllable (M_{diff}=35.3 ms or 11% longer than voiced fricative syllables). The remaining acoustic measures showed no difference. Thus, 3 of 23 measures showed discrete differences; another 6 showed overlapping differences.

For the velar stops, the discrete differences were that ejective release bursts were higher in amplitude (M_{diff}=12.1 rms), longer in duration (M_{diff}=28.8 ms, or 339% longer), and higher in centroid frequencies at all three measured points in the consonant noise (M_{diff}=653.8 Hz) than the bursts of the aspirated stops. These properties are consistent with phonetic descriptions of Zulu ejective velar stops (and our own perceptual observations of them) as somewhat affricated. The overlapping differences were that ejectives also had inconsistently higher mean $F0$ at the first pitch pulse (M_{diff}=28.1 Hz) and lower $F2$ at the first pitch pulse (M_{diff}=90.6 Hz), but higher $F2$ at 15% and 50% into the vowel (M_{diff}=114.3 Hz). All other measures, including VOT, showed no difference. Thus, 5 of 24 measures showed discrete differences, and another 4 showed overlapping differences.

For the bilabial stops, the discrete differences were that the implosives had higher $F0$ and $F1$ frequencies at 15% into the vowel (M_{diff}= 54.7 and 23.7 Hz, respectively), and had higher-amplitude bursts (M_{diff}=10.9 rms) and substantial prevoicing as compared to the small, unaspirated voicing lag of the plosives (VOT M_{diff}=72.2 ms). The overlapping differences were that the implosives had inconsistently higher-frequency centroids at 50% and 85% into the consonant noise (M_{diff}=293.1 Hz), slightly shorter vowels (M_{diff}=26.5 ms) but longer full-syllable durations (M_{diff}=32.4 ms), higher $F0$ at 50% into the vowel (M_{diff}=17.5 Hz), lower $F1$ at the first pitch pulse (M_{diff}=13.2 Hz) but higher $F1$ at 50% and 85% into the vowel (M_{diff}=17.1 Hz), and higher $F2$ at 50% (M_{diff}=46.8 Hz) but lower $F2$ at 85% into the vowel (M_{diff}=17.8 Hz). The remaining measures displayed no differences. Thus, 4 of 24 measures showed discrete differences, and another 10 showed overlapping differences.

To summarize, the consonantal portions of all three stimulus sets showed several discrete or overlapping differences between the contrasting phone sets. The velar and bilabial contrasts showed discrete differences in amplitude of consonantal noise; the velar and lateral fricative contrasts showed discrete differences in consonant noise duration which were proportionally much larger for the velars, and the bilabials differed in VOT. The velar contrast also displayed systematic and pervasive centroid frequency differences in the consonant noise bursts, whereas the lateral fricative and bilabial contrasts showed systematic $F0$ and $F1$ frequency differences early in their vocalic sections. The other acoustic measures showed inconsistent or no differences between the contrasting consonants. It is noteworthy that English listeners show low levels of perceptual confusion, across a range of signal-to-noise ratios, for native stop and fricative voicing distinctions, as well as for native affrication differences (i.e., fricative versus stop manner). These observations are most relevant to the Zulu contrasts tested here. By comparison, confusion levels for native place of

articulation distinctions are fairly high (Miller and Nicely, 1955). The similarity in number of discrete acoustic differences for our three nonnative contrasts, together with the classic findings of Miller and Nicely, offer little *a priori* acoustic basis for predicting discrimination differences.

3. Procedure

Listeners first completed a categorial AXB discrimination test for each of the three Zulu contrasts. In this procedure, A and B are tokens of contrasting non-native phonemes. Listeners are told to circle on their answer sheets for each trial whether the middle item (X, or target) is the same syllable as the first or third item. The X is a different physical token than that of the categorially matched A or B item, so that listeners cannot make a simple acoustic identity judgment (e.g., Best *et al.*, 1988; Polka, 1991, 1992). This procedure was used for several reasons: (1) Because the categorial approach asks listeners to determine whether physically different tokens have the same identity or not, it better approximates natural listening conditions than do tasks that present physically identical tokens for judgment (as in ''same'' trials in AX tasks). (2) Observers display much lower, and easily estimated, response bias in 2AFC (two alternative forced choice) tasks such as AXB than in single-interval decision tasks (e.g., AX), and 2AFC tasks allow measurement of sensitivity to smaller stimulus differences than may be easily assessed with single-interval yes/no tasks such as AX (MacMillan and Creelman, 1991, p. 134). (3) The AXB task was used in the previous investigations of PAM predictions (Best *et al.*, 1988; Best and Strange, 1992).

Each AXB test contained 96 trials in 12-trial blocks (interstimulus interval [ISI]=1 s; intertrial interval=3.5 s; interblock interval=5 s), presented to listeners via audio tape. This was the ISI used in previous PAM reports (Best *et al.*, 1988; Best and Strange, 1992);[8] it should also minimize backward and forward masking between adjacent stimuli. Though this length of ISI might place a load on memory, we assessed this via analysis of short term memory effects (see the next section). The four trial types (AAB, ABB, BBA, BAA) were equally represented for each contrast, and within each test the trial order was randomized. Each of the six tokens per stimulus set occurred four times in each trial type, twice in A and twice in B position, but never paired with the same opposing token more than once.

Following the discrimination tests, a second questionnaire task was conducted involving transliteration of the syllables using English orthography, followed by eliciting additional descriptions, for each set of syllables, in order to evaluate perceptual assimilations. On each trial, the six tokens for a given target syllable were presented. Participants were then directed to write down what the syllable sounded like to them, using English orthography (i.e., ''spell as you would in English''), *if and only if* the consonants sounded to them like anything resembling English consonants. The questionnaire then asked them to write any further description they could give regarding the way the stimuli sounded to them [e.g., ''it sounded like the speaker was doing _____ when she pronounced the consonant'' or ''it sounded like _____ (some nonspeech sound)''']. Participants could not see each others' responses nor discuss their perceptions of the stimuli during the test session.

Listeners were tested in groups of four to six, along one side of a large table in a sound-attenuated room. Stimuli were presented via an Otari MX5050 BQ-II reel-to-reel tape deck connected to a Kenwood amplifier, which fed to a Jamo compact loudspeaker. The speaker was centered on the opposite side of the table, facing the participants (approximately 3 ft from them). Output from the loudspeaker was set to 70±3 dB, as measured from the participants' location.

B. Results

1. Discrimination analyses

Percent correct performance was analyzed in a three-way within-subjects analysis of variance (ANOVA) on the factors stimulus contrast (laterals, velars, bilabials)×trial type (whether X matched the first item of the trial [AAB and BBA trials] versus the third [ABB, BAA])×native similarity (whether X matched the more English-like comparison item [AAB, BAA] versus the less English-like one [ABB, BBA]). Trial type provides an index of memory influences. Better performance on trials where X matches the third item of the trial would suggest a recency-type effect, posited to reflect auditory short-term memory constraints (see Crowder, 1971, 1973). Native similarity was determined by considering the phonetic properties of each Zulu phone relative to the closest English phoneme(s), in addition to considering the listeners' assimilations (below). In the case of the Zulu velar stops, the voiceless /k/ is virtually identical to English /k/, whereas ejective /k′/ is obviously more deviant from English /k/. Listeners' assimilations were consistent with this phonetic analysis: they produced more orthographically regular English spellings for Zulu /k/ than for /k′/. The Zulu plosive bilabial has essentially the same pronunciation as English /b/, whereas the Zulu implosive /ɓ/ is less English-like in that it employs a non-English larynx-lowering gesture which results in negative airflow. However, determining native similarity was more difficult for the Zulu lateral fricatives. Both are deviant from English fricatives in terms of place of articulation, both using the same tongue constriction locations (which are similar to AE /l/), and the voicing distinction is virtually identical to that for AE fricatives. However, one apparent basis of difference in English likeness is evident in the listener assimilations. The phonotactically permissible English spellings they wrote for the voiceless lateral fricative (/s ʃ tʃ sl ʃl/) have a higher mean frequency of occurrence in word-initial position ($M_{\text{frequency}}=0.0166$), according to the Francis and Kuçera database (1982), than did their permissible spellings (/l z/) for the voiced lateral fricative ($M_{\text{frequency}}=0.0037$). Given the lack of other bases for deciding, the voiceless lateral fricative was designated the more English-like item for the native similarity factor.

The main effect of contrast was significant, $F(2,42)=178.91$, $p<0.0001$. Tukey tests revealed that discrimination for the lateral fricatives was significantly better ($M=95\%$ correct, s.e.=0.49) than for the velar stops ($M=89.4\%$, s.e.=1.4), which was significantly better than for

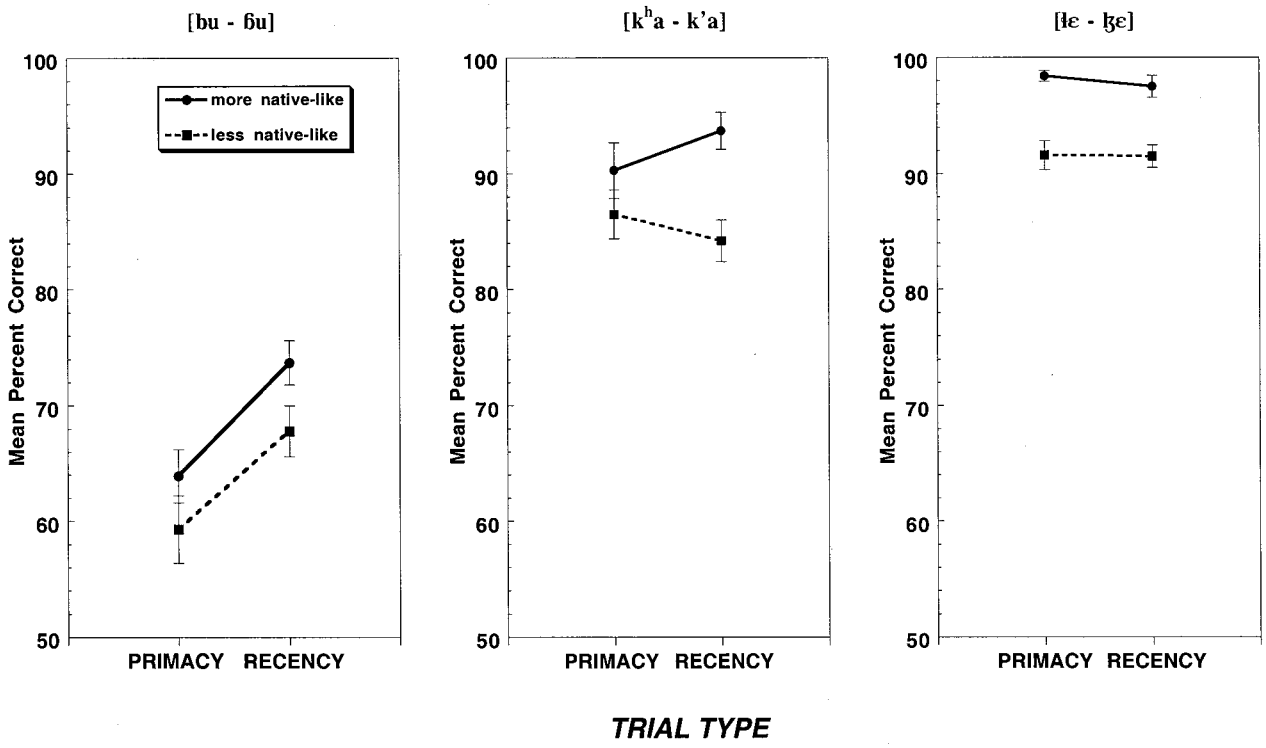


FIG. 1. The AXB discrimination performance in experiment 1 for the factors of contrast×trial type×native likeness. The three panels display results for (a) Zulu plosive versus implosive bilabial stops (SC), (b) voiceless aspirated versus ejective velar stops (CG), and (c) voiceless versus voiced lateral fricatives (TC).

the bilabial stops ($M=65.9\%$, s.e.$=1.5$) (all p's<0.01). Nevertheless, even for the bilabials discrimination was significantly above chance (50% correct), $t(21)=11.59$, $p<0.001$.

The main effect of trial type was only marginally significant ($p=0.08$). However, the trial type×contrast interaction was significant, $F(2,42)=5.495$, $p<0.008$. Simple effects tests revealed that trial type was significant only for the bilabial test, $F(2,21)=7.29$, $p<0.01$, with performance higher on recency-type trials ($M=69.9\%$, s.e.$=1.9$) than on primacy-type trials ($M=61.9\%$, s.e.$=2.1$). This suggests that auditory memory influenced discrimination of the bilabials, but not of the velars nor of the lateral fricatives, which failed to show recency effects. Nonetheless, discrimination of the bilabials was significantly above chance for both recency-type trials, $t(21)=10.57$, $p<0.0001$, and primacy-type trials, $t(21)=5.57$, $p<0.001$ (see Fig. 1).

The native similarity main effect was also significant, $F(1,27)=34.41$, $p<0.0001$. Discrimination was significantly better when the target (X) was more English-like ($M=86.79\%$, s.e.$=1.78$) than when it was less English-like ($M=80.13\%$, s.e.$=1.81$) (see Fig. 1). Although the contrast ×native similarity interaction was nonsignificant, we ran a simple effects test on it in order to determine whether the similarity effect was significant for each contrast individually. The effect was indeed significant for each contrast: SC, $F(1,42)=7.296$, $p<0.013$; CG, $F(1,42)=47.518$, $p<0.0001$; TC, $F(1,42)=29.2$, $p<0.0001$. To determine whether the effect differed in magnitude among the three contrasts, we then calculated difference scores (more-English-like minus less-English-like) and conducted a contrast×trial type ANOVA. No main effects or interactions were significant in this analysis, and Tukey tests among the contrasts were all *ns*, indicating a lack of variation in magnitude of the native similarity effect.

Although response bias is low for AXB and other 2AFC discrimination procedures, we applied MacMillan and Creelman's recommended bias-correction procedure to the percent correct data (1991, p. 127). The formula, $q_{2\mathrm{AFC}}=[p(c)_{2\mathrm{AFC}}-0.5]/(1-0.5)$, yields the proportion of guessing-corrected performance above chance, which we multiplied by 100 to obtain corrected percent above-chance performance. Since this is a linear transformation of the raw percentage data, the ANOVA results were identical to those for the uncorrected scores. Both the uncorrected and corrected cell means are listed in Table II.

2. Assimilation patterns

English spellings and descriptions of the Zulu consonants by each participant, on each contrast, were categorized according to whether the participant used the same or different consonant spellings for the contrasting syllable onsets, as well as whether their additional written descriptions identified any differences they noticed in the productions or sound of the consonants. If both consonant onsets were spelled identically, or were phonologically equivalent in English orthography (e.g., CA and KA), and the participant's additional descriptions failed to note any other consonantal differences,[9] the participant's assimilation pattern for that contrast was categorized as SC. If instead the contrasting consonants were spelled with a common letter, yet one member of the pair was further modified by punctuation marks or

TABLE II. Mean percent correct discrimination for trial type×native likeness×contrast.

	Trial type			
	Primacy (AAB, BBA)		Recency (BAA, ABB)	
Contrast	More nativelike	Less nativelike	More nativelike	Less nativelike
Zulu [bu–ɓu]	63.98 (27.97)[a]	59.32 (18.66)	73.67 (47.34)	67.85 (35.69)
Zulu [k^ha–k′a]	90.35 (80.71)	86.45 (72.90)	93.68 (87.36)	84.17 (68.34)
Zulu [ɮɛ–ɬɛ]	98.42 (96.86)	91.57 (83.14)	97.52 (95.04)	91.46 (83.29)

[a]Values in ( ) are corrected for guessing/bias (Macmillan and Creelman, 1991; see text), i.e., corrected percent above chance.

by additional letters to emphasize some phonetic feature (e.g., K followed by H to indicate aspiration), and/or the participant's written description noted some phonetic (or acoustic) discrepancy between the two consonants, then the assimilation of that contrast was considered a CG difference within a single English consonant. That is, a perceived goodness difference was inferred from the discrepant notation/description. But if the two consonant onsets were spelled with different letters or combinations of letters that indicate phonologically different English pronunciations, the assimilation pattern was categorized as a TC type. If the spelling and description had referred to a stimulus set as falling somewhere in between two or more English consonants (e.g., ''between 'sh' and 'th''' or ''sometimes sounds like 's' sometimes like 'sh' or 'zh''') for one or both Zulu consonants, then the pattern would have been categorized as UC (uncategorized-categorized) or UU assimilation, respectively. Alternatively, if the listener gave a name or description only of some nonspeech sound (e.g., ''snapping'' or ''popping sound'' or ''whooshing''), it would have been designated as a NA type. No participants indicated that any of the Zulu consonants were heard as uncategorized speech sounds or as NA nonspeech sounds. That is, all were described as English consonants or consonant sequences.

All 22 participants showed the expected TC assimilation of the lateral fricatives to some phonological distinction in English. Each labeled the voiceless lateral fricative as some AE voiceless fricative or affricate involving the same articulators (tongue tip/body); ten combined this with /l/, /h/, /t/ or /z/. For the voiced lateral fricative, ten gave the label ''l,'' five gave ''z,'' three wrote a voiced fricative combined with other fricatives involving tongue tip/body, and the remaining four provided clusters of voiced fricative +''l'' (thus combining the same articulators, constriction locations, degree, and/or laryngeal setting) (see Table III). A few participants provided additional articulatory descriptions, most involving tongue tip/body constrictions (silent ''n,'' unpronounced ''l,'' soft ''c;'' stronger ''s'' or ''l''). Only one participant offered a more acoustic-oriented description, indicating a ''slight click on the 'l''' he had heard for the voiced fricative.

For the velar stops, all participants were again consistent in their assimilations, this time reporting the expected CG difference in goodness of fit to AE /k/. All listed /k/ as their primary response (same supralaryngeal articulator, constriction location, and degree, and same laryngeal gesture), but all notated the ejective with a q, c, g, ch, or h or a mark such as an apostrophe or a dash following the /k/. All but 4 participants also wrote further descriptions of the ejective, with 16 indicating that the ejective included some unusual articulation in the throat (pharynx) (choke; gagging; gurgle; throaty; clearing throat; in back of throat), and/or involving the tongue tip/body (clucking; clicking [painful; nasalized; at roof of mouth; Bushman-like]), and two giving more acoustic-oriented descriptions (slight clacking noise; broken up).

The bilabial stops yielded a somewhat less consistent assimilation pattern, although about $\frac{2}{3}$ of the participants ($n=15$) showed the expected SC assimilation to a single AE consonant lacking any notated differences in goodness of fit. These SC listeners reported both bilabial consonants as /b/ (same articulatory organ, supralaryngeal constriction location, and constriction degree) without additional spelling, marking, or descriptive differences; however, a subset of these did report vowel or intonation differences ($n=9$). Discrimination performance for SC listeners overall was poor ($M=64.9\%$, s.e.$=1.95$); it was no higher for the subset who noted vowel or intonation differences ($M=63.93\%$, s.e. $=2.55$) than for those who did not ($n=6$: $M=65.87\%$, s.e. $=2.52$). Two other participants showed a CG assimilation pattern to /b/, providing added articulatory descriptions for the implosive (harder; ''mb'' described as softer /b/). Their discrimination performance ($M=72.23\%$, s.e.$=2.33$) was better than that of the SC participants. Four others showed TC assimilation as /b/ vs /v/ (different constriction location and degree); however, their discrimination ($M=65.91\%$, s.e.$=2.25$) was no better than that of the SC participants. The remaining participant failed to describe one Zulu bilabial; his assimilation was not classifiable.

3. Discrimination reevaluated

Given the individual variations in assimilation of the bilabial contrast, we tested whether the discrimination results would be upheld for just the 15 participants who had shown the predicted assimilation types on all three contrasts: TC assimilation of the lateral fricatives, CG assimilation of the velar stops, and SC assimilation of the bilabial stops. The results remained essentially the same as for the full group.

TABLE III. American English-speaking adults' assimilations of Zulu consonants. Values in parentheses indicate number of participants providing each transcription type.

Lateral fricatives				Velar stops				Bilabial stops			
Voiceless: /ɬɛ/		Voiced: /ɮɛ/		Voiceless: /ka/ ([k^ha])		Ejective: /k′a/		Plosive: /bu/		Implosive: /ɓu/	
s[a]	(8)	l[b]	(10)	k[c]	(19)	k[c,d]	(6)	b	(20)	b[e]	(16)
sh	(3)	z	(5)	kh	(2)	k_[c,f]	(8)	bh	(1)	v	(4)
sl[g]	(3)	zl[h]	(2)	chz	(1)	ck[i]	(4)	missed	(1)	vb	(1)
ch	(1)	thl	(2)			kh[j]	(2)			mb	(1)
ts[k]	(6)	th with z/v	(2)			kch[l]	(2)				
cth	(1)	szt	(1)								

[a]Includes ''s'' or ''hs'' or '''soft' c'' or '''soft' sc.''
[b]Includes ''hl.''
[c]Includes ''k'' or '''hard' c.''
[d]Described by all listeners with additional features: choking, throat-clearing, clacking, clicking, gagging.
[e]Described by four listeners as ''harder''' or ''with pursed lips'' or ''with tensed speech muscles.''
[f]Includes transcriptions of ''k'' with apostrophe or with epenthetic vowel after ''k.''
[g]Includes ''shl'' or ''chl.''
[h]Includes ''zhl.''
[i]Includes ''qk'' or ''gk'' or ''tk.''
[j]Includes ''gh.''
[k]Includes ''tz'' or ''tsc'' or ''zs.''
[l]Includes ''chg.''

The main effect of contrast was significant, $F(2,28) = 114.01$, $p < 0.0001$, supporting the predicted performance pattern of TC>CG>SC. The trial type main effect remained nonsignificant, while the contrast×trial type interaction became marginal, $F(2,28) = 2.805$, $p < 0.08$. However, simple effects tests again indicated an advantage on recency-type trials for the bilabial contrast, $F(1,14) = 4.41$, $p = 0.05$, but not for the fricatives or velars. Bilabial discrimination remained significantly above chance both for recency-type trials ($M = 68.67\%$, s.e.$= 2.36$), $t(14) = 7.92$, $p < 0.0001$, and primacy-type trials ($M = 60.74\%$, s.e.$= 2.83$), $t(14) = 3.8$, $p < 0.002$. The main effect of native similarity also remained significant, $F(1,14) = 18.99$, $p < 0.0007$, and did not interact significantly with trial type or contrast. Thus, listeners who showed the predicted assimilation for all three contrasts performed better on all of them when X was more English-like.

C. Discussion

The assimilation results are largely consistent with the PAM predictions made on the basis of articulatory-phonetic similarities between Zulu and AE consonants. As expected, the lateral fricatives were assimilated as a TC contrast, and the velar stops as a CG difference within a single English consonant, by all listeners. Over $\frac{2}{3}$ of the participants also showed the predicted SC assimilation of the bilabial stops, reporting no differences in the consonants' goodness of fit to AE /b/. The remainder, who showed either CG assimilation of the bilabials to /b/ or TC assimilation to /b/ vs /v/ or /w/, displayed clear responsiveness to articulatory properties, in that they always reported hearing consonantal constrictions involving lips as the articulator, distinguished either by a noncontrastive difference in degree of constriction (e.g., ''more pursed or tense'') or by a phonotactically permissible AE contrast in constriction degree and/or location (/v/). Most listeners' assimilations referred to articulatory properties of the stimuli; that is, the listeners seem to have approached the task as ''naive phoneticians,'' with a focus on articulators, constriction location, and degree.

The variations in discrimination across the contrasts, considered in light of the assimilation patterns, also supported the PAM prediction of the performance pattern TC >CG>SC. That is, the lateral fricatives were discriminated better than velar stops, which were discriminated better than bilabial stops. By comparison, the Zulu click contrasts tested by Best and colleagues (1988), which had yielded clear NA assimilation, were discriminated between 80.6% and 99.1% correct. Thus, click discrimination ranged between the TC and CG levels we found here, and was substantially better than the SC level, again in keeping with PAM predictions.

The native similarity effects with AE listeners are of particular theoretical interest. That discrimination performance was influenced by the native likeness of the target item in the AXB trials is consistent with the PAM claim that perceivers are sensitive to variations of a native consonant. For the CG assimilation case, this effect may appear to be consistent with NLM predictions (Grieser and Kuhl, 1989; Kuhl, 1992; Kuhl *et al.*, 1992) about goodness-related discrimination effects (cf. Miller, 1994; Volaitis and Miller, 1992). Specifically, NLM predicts that discrimination should be worst among tokens that are acoustically similar to the prototype, and best among tokens that are non-prototypical, of the same native category. However, our finding seems to show the opposite pattern—*better* discrimination performance for more nativelike targets (i.e., more prototypical) and *poorer* performance for less nativelike (non-prototypical) ones. Possibly, methodological differences contribute to this apparent reversal of NLM findings. We employed a categorial AXB task whereas Kuhl and colleagues tested detection of stimulus changes against a repeating background. Perhaps the perceived equivalence between target and matching items in our categorial AXB task are greater when the target is more English-like (i.e., corre-

sponds to an NLM prototype) than when it is less English-like (i.e., corresponds to a non-prototype). This might provide an NLM-compatible interpertation of native similarity effects (see also Polka and Werker, 1994). However, the native similarity effects for SC and TC assimilations are inconsistent with NLM expectations, given that these types involve a notable difference in goodness of fit to the associated native phoneme. That is, NLM should expect a native similarity effect in discrimination only for the CG contrast, with significant differences in magnitude of the effect between the CG contrast and the other two types. This expectation was not supported; all contrasts showed the effect, and its magnitude did not differ significantly among them.

Two aspects of the Zulu bilabial findings must also be addressed: above-chance discrimination by listeners who showed SC assimilation, and poor discrimination by those who showed TC assimilation. SC listeners' discrimination was poor, as predicted, but was nevertheless significantly above chance, even according to the bias-corrected scores. This may perhaps seem unsurprising, in light of ample evidence from studies of categorical perception that within-category discrimination is usually significantly better than chance. But the important question is, why isn't it at chance, specifically in the present SC case? These listeners had failed to detect any sort of phonological contrast, or even any differences in phonetic goodness of fit to AE /b/. Obviously, whatever remaining properties they detected did not support very good discrimination. But what actual stimulus differences might they have heard? There was a reliable difference in voicing between the unaspirated /bu/ and prevoiced /ɓu/; however, both voicing values are found in allophones of AE /b/ and are difficult for English listeners to discriminate (Lisker and Abramson, 1967), and few participants reported such differences. On the other hand, nine listeners reported differences in vowel quality or intonation, perhaps associated with slight differences in $F1$ and $F2$ onsets, and mid-vowel $F0$ differences, respectively, for /bu/ vs /ɓu/. Still, those who reported such differences discriminated the Zulu bilabials no better than those who did not.

The recency-type effect for the bilabial contrast alone may suggest another clue, although as noted earlier, neither PAM nor NLM and SLM make *a priori* predictions about auditory memory effects on discrimination. Discrimination in cases where the listener fails to detect either a phonological contrast or a phonetic goodness difference would be expected to involve detection of nonlinguistic auditory differences, and thus to show an influence of auditory memory reflected in a recency effect (see Crowder, 1971, 1973). Recency effects in discrimination would not be expected for CG or TC assimilation, which involve detecting phonetic or phonological differences, respectively, rather than nonlinguistic differences. That is, we speculate that detection of contrastive phonological distinctions versus non-contrastive phonetic details versus nonlinguistic auditory properties is somehow differentiated in non-native speech perception. Although this three-way division is superficially consistent with Werker and Logan (1985)'s proposal for separate phonological, phonetic, and auditory processing levels, our own view more closely follows the direct realist position that listeners detect information in signals about the nature of the event that produced the signal. In the case of speech signals, listeners could detect several types of event information: articulatory patterns that signal phonological distinctions in a language, articulatory patterns that are noncontrastive phonetic variants of phonemes in the language, or nonlinguistic aspects of vocal (or other) sound-producing events such as breathiness, emotional intonation, murmuring, clacking noise, choppiness, etc.

Consistent with the preceding reasoning, recency effects were found for SC but not for CG or TC assimilations (see Fig. 1). However, the difference in mean discrimination levels for the three contrasts raises the possibility that the recency effect for bilabials is due simply to the generally poor performance level rather than to the detection of nonlinguistic, as opposed to phonetic or phonological, information *per se*. However, we can assess this possibility by testing for recency effects in discrimination of the nine NA (nonassimilable) Zulu click contrasts examined by Best and colleagues (1988). Performance on those clicks was higher than the current SC discrimination, and comparable to CG and TC discrimination levels in the present study (80%–99% correct). Yet discrimination of the clicks apparently involved detection of nonlinguistic rather than phonetic and phonological differences, as with the SC contrast in the current study. Therefore, we reexamined the click discrimination data in a new ANOVA on trial type×feature type (voicing contrasts versus place of articulation contrasts)×phonetic contrast (voicing contrasts: prevoiced/short-lag unaspirated, unaspirated/long-lag aspirated, prevoiced/aspirated; place contrasts: dental-lateral, lateral/palatal, dental/palatal). Only trial type was significant, $F_{(1,8)}=22.71$, $p<0.002$, indicating a recency effect: better discrimination for trials in which the target matched the third item of the trial ($M=92.95\%$, s.e. $=0.77$) rather than first item ($M=89.92\%$, s.e.$=1.03$). So, recency effects appear to be associated specifically with detection of nonlinguistic as opposed to phonological or phonetic differences, rather than being associated with poor discrimination.

There is another puzzle, however. The small number of listeners who reported TC assimilation of the bilabials showed *poor* discrimination, no better than the SC listeners. Why? We suspect the reason that these unexpected and infrequent cases of TC assimilation for the bilabials, unlike the expected and unanimous TC cases for the lateral fricatives, showed an assimilation-discrimination discrepancy may be attributable to task order. We had listeners complete the discrimination task prior to the spelling/description task in order to minimize influences of categorization on discrimination performance; this was a necessary experimental control for evaluating PAM hypotheses about the influence of perceptual assimilation patterns on discrimination. However, this minority of listeners may have felt compelled to generate some AE phonological distinction when presented with the two bilabial categories in the second task, even though their poor AXB performance strongly suggests that they had not detected any such differences during the preceding discrimination task.

The more pervasive TC effect, however, was the pre-

dicted one of excellent discrimination in the case of unanimous TC assimilations of the lateral fricatives. Here, the order of experimental tasks was crucial for ruling out any possibility that categorization experience with the stimuli could have directly affected discrimination performance within the experimental context. This TC assimilation pattern with near-ceiling discrimination is the most surprising of the PAM predictions, from the perspective of classic reports that adults have serious difficulties in labeling and discriminating nonnative phonetic contrasts. Moreover, this TC pattern has received the least prior research attention, having been reported only by Best and Strange (1992), and there only for categorical perception of a synthetic continuum rather than of multiple natural utterances. For these reasons, we conducted a second experiment to extend our investigation of TC assimilation to another non-native contrast from a different language.

III. EXPERIMENT 2

For this study, we chose a stop consonant contrast from a second African language, Ethiopian Tigrinya, which is from a different language family (Afro-Asiatic: Semitic: Ethiopic) than Zulu (Niger-Kordofanian: Niger-Congo: Bantu) (Ruhlen, 1975). The contrast was between the ejective bilabial versus alveolar stops /p′/ and /t′/. The consonants contrasted in constriction locations that occur in AE rather than in laryngeal gestures, as in experiment 1; the laryngeal gesture of both was non-native to English. We also tested discrimination of two native AE fricative voicing contrasts involving phonemes that had appeared in the experiment 1 participants' spellings of the Zulu lateral fricatives: /s/-/z/ and /ʃ/-/ʒ/ and involved tongue tip/body as the active articulators, for comparison to the results with that contrast. Because the lateral fricatives had been presented with the lax vowel /ɛ/ in open CV syllables, which is phonotactically impermissible in English, we used the same vowel and CV context in all contrasts tested in experiment 2. To directly compare the results to those for the lateral fricatives (TC) of experiment 1, we used the same testing procedures.

A. Method

1. Participants

The listeners were 19 native speakers of American English (10 female, 9 male) with a mean age of 18.7 years (range=18–20 yr). None had experience with Tigrinya or any other languages employing ejective consonants. None had a personal or family history of developmental speech, language, or reading disorders. Eight other participants were tested but their data were removed from the final data set due to developmental and/or familial speech impairments ($n=5$), familial language disorders ($n=1$), or reading impairments ($n=1$) (see footnote 5), or chance-level discrimination of the English control contrasts ($n=1$).

2. Stimulus materials

A male native Tigrinya speaker from Ethiopia (Eritrea) was recorded producing multiple tokens of each of the two non-native CV nonsense syllables /p′ɛ/-/t′ɛ/. The supralarngeal articulators, constriction locations, and constriction degrees for these two stops correspond to those of the AE voiceless stops /p/-/t/; however, the ejective laryngeal gesture of both is not used in English. The syllables were read aloud individually from a randomly ordered list containing 20 repetitions of each. The AE contrasts /sɛ/-/zɛ/ and /ʃɛ/-/ʒɛ/ were recorded according to the same procedure by a female native AE speaker (author CTB).[10] The recordings were digitized and analyzed as in experiment 1. Six tokens were selected per category, matched as closely as possible between the contrasting syllables of each pair for overall duration, fundamental frequency and contour, and vowel formant frequencies (see Table IV).

The discrete differences between the final stimulus sets for the contrasting Tigrinya ejectives are in the spectrum, duration, and amplitude of the release bursts. The /t′/ bursts were longer in duration ($M_{\text{diff}}=8.6$ ms, or 72% longer), higher in amplitude ($M_{\text{diff}}=9.9$ rms), higher in centroid frequencies throughout ($M_{\text{diff}}=986.3$ Hz), and had higher $F2$ and $F3$ values at the first pitch pulse of the vowel ($M_{\text{diff}}=516.0$ and 1182.2 Hz, respectively) than the /p′/ bursts. All other acoustic measures on the syllables, including VOT, showed no difference between the two categories. Acoustic measures for the AE fricative contrasts were quite similar between paired stimulus sets, except for the obvious voicing difference.

3. Procedure

As in Experiment 1, listeners first completed categorial AXB discrimination tests for the Tigrinya and AE contrasts. Following the discrimination tasks, listeners completed the assimilation questionnaire for each set of Tigrinya syllables, as in experiment 1. Listeners were tested in groups of four to six in the same experimental setup as before.

B. Results

1. Discrimination analyses

The AXB discrimination data were submitted to ANOVA for the within-subject effects of contrast (the two AE contrasts versus the Tigrinya contrast)×trial type (primacy versus recency).[11] This time, the only significant effect was contrast, $F_{(2,36)}=6.792$, $p<0.003$. Discrimination performance was essentially at ceiling for the two AE contrasts (for /s/-/z/, $M=98.8\%$ correct, s.e.=0.33; for /ʃ/-/ʒ/, $M=98.8\%$ correct, s.e.=0.37), but was somewhat lower and more variable, though still excellent, for Tigrinya /p′/-/t′/ ($M=91.4\%$ correct, s.e.=2.02). The trial type effect and the interaction were nonsignificant. Cell means for percent correct, as well as for bias-corrected percent performance above chance, are shown in Table V.

Performance on the lateral fricatives of experiment 1 was compared to performance for each of the AE fricative voicing contrasts in between-subject contrast×trial types ANOVAs. Discrimination was significantly lower, though still excellent, for Zulu /ɮe/-/ɬe/ ($M=95\%$ correct, s.e. =0.49) as compared to both AE /s/-/z/, $F(1,39)=26.243$, $p<0.0001$, and AE /ʃ/-/ʒ/, $F(1,39)=32.14$, $p<0.0001$ (see preceding paragraph for AE means).

TABLE IV. Acoustic attributes of the stimulus tokens selected for each target category in experiment 2.

Stimulus syllable	Syllable duration[a]	Consonant duration[a]	Vowel duration[a]	Consonant VOT[a]	Consonant amplitude[b]	Consonant centroid[c]	Vowel $F1$[d]	Vowel $F2$[d]	Vowel $F3$[d]	Vowel $F0$[d]
Tigrinya ejectives:										
/**pʼɛ**/	**235.0** (223–248)[e]	**12.0** (9–16)	**157.9** (144–169)	**74.8** (66–92)	**40.8** (37–43)					
first pulse[f]							**456.7** (318–507)	**2001.5** (1871–2190)	**2632.0** (2517–2780)	**127.1** (104–143)
15%[g]						**3602.3** (3282–3993)	**618.7** (593–627)	**2052.3** (2027–2086)	**2868.0** (2643–2998)	**107.3** (105–111)
50%[g]						**3365.8** (2977–3780)	**626.3** (612–637)	**2085.8** (2023–2205)	**2792.8** (2742–2815)	**97.3** (94–101)
85%[g]						**3223.5** (2991–3595)	**624.0** (612–630)	**2068.5** (2028–2209)	**2872.7** (2815–2998)	**90.7** (86–97)
/**tʼɛ**/	**232.8** (213–248)	**20.6** (18–27)	**157.1** (135–176)	**75.1** (57–105)	**48.7** (47–52)					
first pulse							**478.0** (472–490)	**2517.5** (2448–2570)	**3814.2** (3724–3986)	**126.7** (111–145)
15%						**4614.0** (4391–4848)	**622.2** (612–631)	**2078.0** (1885–2195)	**2863.8** (2805–2937)	**114.9** (106–123)
50%						**4494.2** (3958–5112)	**621.3** (612–631)	**2088.0** (2033–2189)	**2882.5** (2804–3015)	**102.2** (93–106)
85%						**4042.4** (3372–4874)	**611.2** (551–630)	**2142.3** (2033–2199)	**2818.5** (2805–2936)	**91.6** (86–101)
Eng. alveolar fricatives										
/**sɛ**/	**487.8** (471–509)	**239.3** (221–254)	**319.3** (313–329)		**50.3** (49–52)					
first pitch pulse							**319.3** (313–329)	**1677.2** (1557–1741)	**2785.2** (2645–2817)	**218.3** (186–237)
15%						**6822.2** (6511–6971)	**477.2** (468–487)	**1731.3** (1713–1756)	**2693.5** (2505–2821)	**188.6** (188–192)
50%						**7177.2** (6992–7473)	**486.7** (475–496)	**1722.5** (1575–1878)	**2711.5** (2639–2822)	**172.8** (165–174)
85%						**7348.0** (6965–7750)	**476.2** (461–496)	**1648.2** (1617–1692)	**2617.0** (2508–2791)	**166.0** (159–178)
/**zɛ**/	**465.0** (440–497)	**211.8** (197–244)	**238.6** (222–253)		**48.8** (47–51)					
first pitch pulse							**308.5** (299–319)	**1720.3** (1554–1881)	**2725.8** (2517–2804)	**185.4** (179–189)
15%						**6345.8** (5533–7182)	**477.0** (468–483)	**1781.3** (1713–1881)	**2810.2** (2801–2821)	**184.8** (180–190)
50%						**7177.6** (6809–7563)	**482.3** (466–498)	**1731.3** (1554–1876)	**2765.3** (2672–2831)	**172.6** (168–177)
85%						**7402.2** (7175–7768)	**466.5** (455–481)	**1698.5** (1564–1881)	**2737.5** (2500–2819)	**158.1** (153–161)
Eng. palatal fricatives										
/**ʃɛ**/	**490.7** (476–500)	**256.7** (243–268)	**229.1** (223–237)		**46.5** (45–48)					
first pitch pulse							**325.0** (313–348)	**1618.5** (1541–1866)	**2474.0** (2203–2657)	**227.5** (215–239)
15%						**5477.5** (5250–5644)	**478.2** (458–498)	**1853.5** (1741–1866)	**2740.2** (2517–2821)	**186.4** (180–195)
50%						**5629.3** (5473–5818)	**483.0** (473–495)	**1801.3** (1716–1886)	**2630.7** (2495–2786)	**171.4** (165–180)
85%						**5539.3** (5451–5613)	**482.5** (472–490)	**1732.2** (1724–1744)	**2714.3** (2508–3129)	**168.0** (164–174)

TABLE IV. (*Continued*.)

Stimulus syllable	Syllable duration[a]	Consonant duration[a]	Vowel duration[a]	Consonant VOT[a]	Consonant amplitude[b]	Consonant centroid[c]	Vowel $F1$[d]	Vowel $F2$[d]	Vowel $F3$[d]	Vowel $F0$[d]
/ʒɛ/	**485.6** (467–509)	**229.5** (213–256)	**242.9** (227–264)		**45.3** (43–47)					
first pitch pulse							**316.4** (313–319)	**1650.3** (1556–1783)	**2522.0** (2360–2640)	**184.5** (175–193)
15%						**5134.6** (4912–5619)	**479.7** (461–496)	**1879.2** (1867–1885)	**2663.0** (2500–2813)	**179.8** (174–185)
50%						**5477.7** (5290–5924)	**479.0** (464–490)	**1795.5** (1709–1876)	**2672.3** (2477–2809)	**170.9** (167–177)
85%						**5489.3** (5278–5740)	**482.0** (471–489)	**1708.8** (1570–1766)	**2631.2** (2454–2815)	**169.3** (166–172)

[a]Duration in milliseconds (ms); consonant duration refers to burst for velar and bilabial stops, to frication for lateral fricatives.
[b]Root mean square amplitude across full duration of consonant noise (frication or burst).
[c]Centroid frequencies of consonant noise at onset, middle, and offset of noise (frication or burst).
[d]Frequency in Hz at first pitch pulse and at 15%, 50%, and 85% into vowel for lowest three formants and $F0$ (LPC estimates).
[e]Minimum and maximum values, to nearest integer.
[f]First pitch pulse of vocalic portion.
[g]Percent into consonant noise (frication or burst) for centroid measures; percent into vocalic portion for formant and $F0$ measures.

We also directly compared performance on the two nonnative TC contrasts in a between-subject contrast×trial type ANOVA. There were no significant differences.

2. Assimilation patterns

Assimilation patterns were determined according to the experiment 1 criteria. As predicted, the great majority of participants assimilated the Tigrinya ejectives as a TC contrast ($n=16$). Of the 16 who showed TC assimilation, 12 reported hearing /p/-/t/, that is, their assimilations were consistent with the supralaryngeal articulators, constriction locations, and constriction degree. Consistent with the notion that listeners can detect within-category phonetic differences, i.e., between non-native phones and the native categories to which they are assimilated, some listeners noted deviant articulatory details involving throat and/or larynx for the /p/-/t/ assimilations (e.g., click in the throat; windy—a lot of breath behind it; swallowing the consonant; abruptly cut off; sucked in); others noted deviant supralryngeal articulations (spitting out the syllables; hard or pronounced P and T). Two other TC listeners reported /p/ vs /pt/ or /pb/, and two reported an isolated vowel (i.e., no consonant) vs vowel_t. Two of the remaining showed SC assimilation, one reporting ''EH''-''EH,'' the other /p/-/p/. Consistent with PAM expectations, the latter 2 participants showed substantially lower discrimination ($n=2$, $M=64.58\%$, s.e.$=13.22$) than the 16 who displayed TC assimilation ($M=94.09\%$, s.e.$=1.96$). The final participant failed to describe one Tigrinya consonant; his assimilation was not classifiable.

TABLE V. Mean percent correct discrimination for trial type×contrast in experiment 2.

	Trial type			
	Primacy		Recency	
Contrast	AAB	BBA	BAA	ABB
English [sɛ–zɛ]	97.66 (95.32)[a]	99.42 (98.83)	99.71 (99.42)	98.25 (96.49)
English [ʃɛ–ʒɛ]	98.24 (96.49)	98.54 (97.08)	99.71 (99.42)	98.54 (97.08)
Tigrinya [p′ɛ–t′ɛ]	91.81 (83.62)	88.59 (77.19)	92.11 (84.21)	92.98 (85.96)

[a]Values in ( ) are corrected for guessing/bias (Macmillan and Creelman, 1991; see text), i.e., corrected percent above chance.

C. Discussion

The findings from experiment 2 are straightforward. The TC assimilation pattern and its associated high level of discrimination clearly generalized to another non-native contrast. Experiment 2 involved a different type of non-native phonetic contrast, and a second unrelated language, than in experiment 1. Thus, the TC assimilation pattern apparently applies to constriction location contrasts as well as laryngeal gesture contrasts (voicing).

Note, however, that the TC assimilation of non-native contrasts reported in both experiments yielded modestly but significantly lower discrimination (low-mid 90% range) than do comparable native contrasts (near-100% range). Listeners appear to be sensitive, simultaneously, both to information that may be relevant to a native phonological contrast, and also to articulatory differences between nonnative phones and the most similar native phonemes.

Finally, as with the Zulu bilabial results in experiment 1, some individual differences were apparent in assimilation and discrimination of at least some non-native consonant contrasts. Two experiment 2 participants failed to note any phonetic or phonological differences between the Tigrinya ejectives, showing SC assimilation, with concomitantly poor discrimination.

IV. CONCLUSIONS

The results of the reported experiments support the notion that listeners perceptually assimilate and discriminate non-native consonants with respect to their phonetic similarity to native contrasts, in accordance with predictions from the perceptual assimilation model (PAM: Best, 1994a,b, 1995; Best *et al.*, 1988). Specifically, for non-native contrasts in which listeners perceived a correspondence to some native phonological distinction, deemed as two-category

(TC) assimilation, discrimination was excellent—above 90% correct. This TC pattern was evident for both a laryngeal gesture distinction and a constriction location distinction from two unrelated languages. By comparison, when the contrasting non-native consonants were heard as differing in goodness of fit to a single native consonant, indicating CG assimilation, discrimination was very good though significantly lower than in TC cases. Thus, while listeners detected variations in the details of items they perceived as variants of a single native consonant, this did not benefit discrimination as much as did the detection of phonologically contrastive information. Finally, when listeners perceived a non-native contrast as equally good variants of a single native consonant, displaying SC assimilation, discrimination was much poorer, as expected. The full set of findings is highly supportive of PAM's proposal of systematic relations between assimilation and discrimination, confirming the discrimination order of TC>CG>SC.

The assimilations of non-native phones to AE consonants corresponded well, for nearly all listeners, to the predictions we had developed from principles of articulatory phonology (see experiment 1 introduction; cf. Browman and Goldstein, 1986, 1989, 1990a,1990b, 1992). Those predictions focused on the use of the same articulators, constriction locations, and/or constriction degrees by non-native and native consonants. Consistent with those expectations, listeners assimilated the Zulu voiced bilabial stops to AE voiced consonants involving the same articulator, that is, lip gestures. Typically, the same location and constriction degree were involved (bilabial stop /b/), though sometimes constriction location and degree differed (labio-dental fricative /v/). Similarly, listeners assimilated the Zulu voiceless and ejective velar stops to the AE voiceless velar stop (/k/), thus to the same articulators (tongue dorsum and glottis) and the same supralaryngeal constriction location and degree, although glottal constriction degree differed for the ejective /k′/. The voiced and voiceless lateral fricatives strongly tended to be assimilated to the AE lateral approximant /l/ (same articulators and constriction locations, but different constriction degree), often combined with voiced or voiceless apical fricatives, respectively (same articulators and constriction degree). And the Tigrinya bilabial versus alveolar ejectives tended to be assimilated to AE voiceless stops using the same supralaryngeal articulators, constriction locations and degree (/p/-/t/).

At the same time, participants often noted additional phonologically irrelevant articulatory features in their written descriptions of the stimulus sets, distinguishing the non-native consonants from the native consonants they perceived as most similar. To illustrate, in addition to their English spellings, listeners sometimes described vocal tract sounds resulting from constrictions of the involved articulators and locations, including tongue body and pharynx (choking, gagging, gurgling, throat-clearing, throaty, guttural sounds), tongue tip (clucking, clicking, stronger /s/), lips (lip-pursing, [lip] muscle-tensing, harder /b/), etc. They only rarely noted nonarticulatory nonspeech sound properties such as clacking noise, broken up.

Three other findings offer several additional insights about nonnative speech perception. First, the native similarity effect, an asymmetry favoring discrimination when the target (X) is the more rather than less native-like member of a nonnative contrast, was found for TC, CG, and SC assimilation types alike in experiment 1. This suggests that familiarity with the typical phonetic form of native consonants aids rather than hinders discrimination, whether the listener is attending for information about phonological contrast, or phonetic goodness of fit to a single phoneme, or nonlinguistic stimulus variations. This suggests that native speech experience results in more stable perception of tokens that are more nativelike, regardless of overall performance level or type of information being discriminated. Evidence of the converse, that perception of less nativelike utterances is less stable, can be seen in the common experience that perception of foreign-accented utterances in the listener's L1, or of utterances in a late-learned L2, is more effortful and error-prone than perception of native L1 utterances.

A second, and perhaps related, finding is that although discrimination of TC contrasts was quite high and that of the SC contrast was quite low, performance differed significantly from ceiling and from chance, respectively. Both observations suggest that listeners retain greater sensitivity to articulatory-phonetic variants of non-native consonants, i.e., show lower perceptual stability or lower perceptual equivalence among tokens, than they do for for native consonants. This burdens discrimination somewhat in TC cases, where the parallel to native phonological contrasts should otherwise have yielded ceiling performance, but it *aids* discrimination in SC cases, where *lack* of correspondence to a native phonological contrast or to a phonetic difference in goodness of fit should otherwise have yielded chance performance. These discrepancies from the upper and lower performance extremes indicate that listeners detect not only the presence/absence of phonological contrast, but also detect phonologically irrelevant phonetic and/or nonlinguistic details. This finding, together with the native similarity effect, appears compatible with the notion discussed in experiment 1, that listeners are able to discriminate three types of information in speech: phonological, phonetic, and nonlinguistic (see also Hallé *et al.*, 1999; Hallé *et al.*, 1998, 2000; Whalen, 1984, 1991). As argued earlier, this could simply involve detection of certain types of information, and need not entail three qualitatively different cognitive processes (see Werker and Logan, 1985).

Third, only the SC contrast elicited recency effects in discrimination. This finding suggests a qualitative division between detection of linguistic (phonological, phonetic) and nonlinguistic information in speech. This memory effect, putatively auditory, occurred only when listeners failed to report hearing phonological or phonetic differences, presumably leaving only a nonlinguistic basis on which they could have discriminated. The implied relationship between recency effects and nonlinguistic auditory discrimination was supported by a reanalysis of earlier findings with Zulu clicks, which had been perceived as nonassimilable (NA) nonspeech sounds (Best *et al.*, 1988). Like the SC bilabials, the clicks were also discriminated on a nonlinguistic rather than a phonological or phonetic basis; however, they were discrimi-

nated much better than the bilabials. The new analyses revealed a recency effect for click discrimination. The apparent restriction of recency effects to SC and NA cases further supports the differentiation of discrimination of nonlinguistic versus phonological and/or phonetic information in speech. If no such difference existed, the recency effect should have been found across all assimilation types.

While we have interpreted our findings in terms of the PAM model, the differences in TC, CG, and SC discrimination may be, in some ways, reminiscent of classical findings on categorical perception (CP) with synthetic speech continua. Although the original CP claim had been that listeners discriminate speech stimuli only so well as they identify or label them differently, much evidence has indicated that within-category discrimination is usually significantly better than that predicted by labeling functions. In particular, discrimination of tokens near the category boundary, i.e., inconsistently labeled or ambiguous tokens, is above chance (cf. Best *et al.*, 1981). Thus, listeners typically display some sensitivity to within-category variations, though certainly less than that for between-category differences. Those observations might be extrapolated to the better discrimination of TC than CG contrasts, and of CG than SC contrasts. It is important to note, however, that CP findings typically involve unnatural synthetic stimulus variations, whereas the present research and other non-native speech studies involve perception of multiple tokens of natural utterances, a situation that better approaches the natural conditions involved in listening to and learning unfamiliar languages.

It is also important to consider whether and how the findings relate to the other non-native speech models discussed earlier: the speech learning model (SLM): Flege, 1986, 1989, 1995) and the native language magnet model (NLM: Grieser and Kuhl, 1989; Kuhl, 1991, 1992; Kuhl *et al.*, 1992). The SC findings might be seen as consistent with the NLM claim, and the SLM implication, that two non-native phones which are quite similar to a native phoneme should each be difficult to discriminate from it. By extension, they should also, presumably, be difficult to distinguish from each other. Both models might also be extended to account for the TC findings, as a case of non-native phones that are easy to discriminate because they are similar to two different native phonemes, and thus virtually identical to a native contrast. That account, however, is indistinguishable from PAM's explicit hypotheses about TC assimilations. Turning to CG assimilation, the results appear consistent with NLM (and SLM) in showing systematic differentiation of good (closer) versus poor (more distant) examples of a given native phoneme. The native similarity effect of experiment 1 is particularly relevant to NLM claims about asymmetries in discrimination of good versus poor exemplars of a native phoneme. However, as discussed there, it is uncertain whether the direction of our native similarity effect supports or conflicts with NLM. Further research would be needed to determine this. Moreover, NLM should predict a discrimination asymmetry for CG but not for TC or SC assimilations, yet significant asymmetries were observed for all three types of contrast. Thus, native speech experience aids categorial discrimination not only when a non-native contrast assimilates to a phonetic goodness difference within a native phoneme (CG), as emphasized in NLM predictions, but also when listeners fail to detect goodness differences and hear only some nonlinguistic difference (SC), as well as when they detect some phonological distinction (TC).

While certain results can be interpreted *a posteriori* as being compatible with both models, a more fundamental caveat with respect to how well NLM and SLM can address the present findings is that both models focus on the attributes of individual phonetic categories. PAM instead focuses on the functional organization of the native phonological system, specifically on the phonological distinctions between, and phonetic variations within, native phonological equivalence classes. Importantly, neither SLM nor NLM would have generated the current set of comparisons. And neither offers a singular, coherent account of the findings like PAM does. Thus, a key contribution of PAM is its provision of a theoretical motivation for systematic comparisons among diverse types of non-native contrasts within the broader context of phonological systems.

Another important theoretical issue, not directly examined here, is how native language effects on non-native speech perception emerge developmentally. Infant research indicates that some native language influences appear during the second half-year, with declining discrimination of at least some non-native consonants by 8–10 months (e.g., Best *et al.*, 1995; Werker, 1989; Werker *et al.*, 1981; Werker and Lalonde, 1989), and of some non-native vowels by 6–8 months (Polka and Werker, 1994; cf. Kuhl *et al.*, 1992; but see Polka and Bohn, 1996). Interestingly, there is no developmental decrease for nonnative Zulu click consonants, consistent with AE adults' very good discrimination and assimilation of them as NA nonspeech sounds (Best *et al.*, 1988, 1995). Even more intriguing, however, is that older infants' perception of both native and non-native speech still differs from that of adults in several nontrivial ways, suggesting that they do not yet perceive phonological contrasts like adults (e.g., Best 1991; Hallé and de Boysson-Bardies, 1996; Stager and Werker, 1997). Those developmental differences have led some to posit that infants are initially responsive to language-universal properties of speech, then begin to recognize language-specific phonetic patterns, and only later discover the contrastive phonological functions of native phonetic classes, perhaps in relation to increases in size of their early lexicon (e.g., Best, 1993; Stager and Werker, 1997; Werker and Pegg, 1992). Against the backdrop of our discussion of the three types of information that adults detect in speech (phonological, phonetic, nonlinguistic) we suggest that infants progress developmentally from detection of only nonlinguistic (or perhaps nonspecific phonetic) information in speech, to recognition of how phonetic variants fit into (or fail to) language-specific phonetic classes, to eventually discovering the phonologically contrastive functions those phonetic classes serve in distinguishing native words. Further research on infants' changing perceptions of diverse nonnative contrasts, assimilated by adults as TC vs CG vs SC (vs NA) contrasts, will be needed to test those speculations. We note, however, that the proposed developmental path is consistent with the classic direct realist view of Gibson and Gib-

son (1955) that perceptual learning involves the increasing differentiation of the lawful stimulus information provided by real-world events. In the case of speech, this differentiation is posited here to involve the emerging recognition of classes of articulatory gestures employed in native speech, followed by discoveries about how those gestural classes help to distinguish among native words.

Before closing, we must address some limitations of the present investigation. The primary methodological limitation lies in our assessment of assimilations. Having listeners give a single native spelling and description of each stimulus set, following the discrimination task, may bias them to search for some between-set difference they were not attending to in the discrimination task. For a more refined approach, forced-choice or perhaps even open-set spellings could be obtained for each token in an assimilation task involving multiple, randomized repetitions. Such data could be easily subjected to standard statistical analyses. Additionally, listeners could rate the goodness of fit between each nonnative token and their associated native-language spelling. Such ratings would be especially useful for differentiating between CG and SC assimilation patterns (see Best *et al.*, 1996; Calderón and Best, 1996). However, other task adjustments would be needed to evaluate listeners' perception of nonlinguistic properties.

Several other aspects of assimilation also deserve further examination, including the basis for predicting the most likely assimilations of a given nonnative contrast by listeners of a given language community. Another finding that calls for further study is the striking individual variation in assimilation patterns for some non-native contrasts, as we found for Zulu bilabials. Evidently, the phonetic properties of non-native phones reflect multiple dimensions of similarity to various native phonemes, and listeners may differ in their attention to specific dimensions.

Additional research is also needed to substantiate the proposed differences in perception of phonological, phonetic, and nonlinguistic information in non-native speech. For example, native phonotactic rules (phonological), such as constraints on the vowels permitted in open and closed syllables, versus native coarticulatory patterns (phonetic) such as anticipatory or carryover coarticulation between vowels and consonants, may influence categorization and discrimination of nonnative contrasts in different ways (see Avery and Best, 1995). Neuropsychological studies could also provide insights. To illustrate, in a recent dichotic listening study, although American English speakers and Zulu speakers displayed similar overall performance levels in judgments of Zulu click consonants, only the Zulu listeners perceived them as speech and showed a left hemisphere advantage (Best and Avery, 1999), indicating one crucial difference in perception of nonlinguistic versus phonological information in click consonants.

To sum up, the present findings are consistent with the hypothesis that non-native speech perception is based on detection of articulatory-phonetic similarities to the phonological units and contrasts of the native language. Discrimination performance levels are strongly linked to listeners' assimilations of non-native phones within their native phonological system. To a large extent, assimilation and discrimination of non-native consonants reflects listeners' sensitivity to phonetic and/or phonological similarities to native consonants. The detection of nonlinguistic properties in speech contributes minimally to non-native speech perception, being evident only when listeners reported hearing no phonetic or phonological differences between contrasting non-native consonants, in which case they showed fairly poor discrimination. The full set of results is most compatible with PAM predictions. While certain findings may be consistent with other views of non-native speech perception, PAM alone provided the motivation for the present cross-language comparisons, and it appears to offer the most coherent account.

ACKNOWLEDGMENTS

Support for this research came from NIH Grant Nos. DC00403 to the first author and HD01994 to Haskins Laboratories. We gratefully acknowledge the following colleagues for their helpful comments on an earlier version of the manuscript: Alice Faber, Carol Fowler, and Michael Studdert-Kennedy. We also thank Keith Kluender and three anonymous reviewers for their insightful comments on our original submission; the final version of the article benefited substantially from their input.

[1]Still, it is important to note that laboratory training effects are limited in magnitude in adults (Lively *et al.*, 1994), and that listeners' discrimination of the critical acoustic properties of nonnative contrasts presented in isolation may fail to generalize to good discrimination of them within speech contexts (Miyawaki *et al.*, 1975; Werker and Tees, 1984).

[2]Terminology for PAM predictions originally referred to ''phonetic categories.'' However, given the model's ecological theoretical perspective, it does not espouse cognitive processing assumptions about mental representations of categories, category formation, etc. Therefore, in this article we have discussed native phonological influences in terms of functional equivalence classes rather than in terms of phonetic categories. However, we have retained the assimilation terms used in earlier presentations of PAM, for consistency with previous publications.

[3]Based on Polka's summary of these subjects' descriptions, the latter assimilations may actually have been UU rather than NA types, given that they reported hearing consonants and/or vowels rather than nonspeech sounds.

[4]It should be noted, however, that Polka concluded that certain other aspects of her findings in these two studies may have been guided by acoustic attributes of the stimuli rather than by phonological/phonetic properties of the listeners' native language.

[5]Reading deficits were used as an exclusionary criterion because they are often associated with deficient phonological skills (e.g., Scarborough, 1998; Shankweiler *et al.*, 1995). Due to requirements of subject pool use, some screening factors had to be applied *after* subject participation. For the same reason, gender was not balanced in the sample. However, this was not deemed critical, as no sex differences have been reported for speech perception tasks such as those used here.

[6]Different vowels were used for the three contrasts because we also planned to use these stimulus materials for a within-subjects study with AE infants. The vowel difference was deemed necessary to maintain infants' attention across their three required tests (Best *et al.*, 1990).

[7]Zulu is a tone language with a differentiation between high and low tones on syllable nuclei.

[8]Note that Werker has often used an even longer ISI of 1500 ms in her investigations of non-native speech perception in infants and adults, as has Polka (1991, 1992).

[9]Comments that instead identified vowel qualities or intonational properties do not reflect perception of the consonants *per se*, and so were not factored into the consonant assimilation determinations.

[10]We also recorded, and collected perceptual data, for the same AE C's followed by /e^{i}/; data on the latter stimuli will not be reported here because the vowel environment differed from the Zulu lateral fricatives and

the Tingrinya stimuli. However, the perceptual results for those stimuli were virtually identical to those for the AE stimuli reported in experiment 2.

[11]Native similarity was not included as a factor because neither the non-native contrast nor the native contrasts of experiment 2 involved differences in English-likeness.

Abramson, A. S., and Lisker, L. (**1970**). ''Discriminability along the voicing continuum: Cross-language tests,'' in *Proceedings of the Sixth International Congress of Phonetic Sciences* (Academia, Prague), pp. 569–573.

Aslin, R. N., and Pisoni, D. B. (**1980**). ''Some developmental processes in speech perception,'' in *Child Phonology: Perception*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson (Academic, New York).

Avery, R. A., and Best, C. T. (**1995**). ''Phonological and phonotactic influences on perception of two nonnative vowel contrasts,'' J. Acoust. Soc. Am. **97**, 3362.

Best, C. T. (**1991**). ''Phonetic influences on the perception of nonnativce speech contrasts by 6–8 and 10–12 month-olds,'' presented at the Society for Research in Child Development, Seattle, April.

Best, C. T. (**1993**). ''Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development,'' in *Developmental Neurocognition: Speech and Face Processing in the First Year*, edited by B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, and J. Morton (Kluwer Academic, Dordrecht, The Netherlands).

Best, C. T. (**1994a**). ''Learning to perceive the sound pattern of English,'' in *Advances in Infancy Research*, edited by C. Rovee-Collier and L. P. Lipsitt (Ablex, Norwood, NJ).

Best, C. T. (**1994b**). ''The emergence of native-language phonological influences in infants: a perceptual assimilation model,'' in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, edited by H. C. Nusbaum (MIT, Cambridge, MA).

Best, C. T. (**1995**). ''A direct realist perspective on cross-language speech perception,'' in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research*, edited by W. Strange (York, Timonium, MD), pp. 167–200.

Best, C. T., and Avery, R. A. (**1999**). ''Left-hemisphere advantage for click consonants is determined by linguistic significance and experience,'' Psychological Science **10**, 65–69.

Best, C. T., and Strange, W. (**1992**). ''Effects of phonological and phonetic factors on cross-language perception on approximants,'' J. Phonetics **20**, 305–330.

Best, C. T., Faber, A., and Levitt, A. G. (**1996**). ''Perceptual assimilation of non-native vowel contrasts to the American English vowel system,'' J. Acoust. Soc. Am. **99**, 2602.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). ''Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants,'' J. Exp. Psychol. Hum. Percept. Perform. **4**, 45–60.

Best, C. T., Morrongiello, B., and Robson, R. (**1981**). ''Perceptual equivalence of acoustic cues in speech and nonspeecch perception,'' Percept. Psychophys. **29**, 191–211.

Best, C. T., McRoberts, G. W., LaFleur, R., and Silver-Isenstadt, J. (**1995**). ''Divergent developmental patterns for infants' perception of two nonnative consonant contrasts,'' Infant Behavior and Development **18**, 339–350.

Best, C. T., McRoberts, G. W., Goodell, E., Womer, J., Insabella, G., Klatt, L., Luke, S., and Silver, J. (**1990**). ''Infant and adult perception of nonnative speech contrasts differing in relation to the listeners' native phonology,'' presented at the International Conference in Infant Studies, Montreal, 19–22 April.

Browman, C. P., and Goldstein, L. (**1986**). ''Towards an articulatory phonology,'' Phonology Yearbook **3**, 219–252.

Browman, C. P., and Goldstein, L. (**1989**). ''Articulatory gestures as phonological units,'' Phonology **6**, 151–206.

Browman, C. P., and Goldstein, L. (**1990a**). ''Gestural specification using dynamically-defined articulatory structures,'' J. Phonetics **18**, 299–320.

Browman, C. P., and Goldstein, L. (**1990b**). ''Representation and reality: physical systems and phonological structure,'' J. Phonetics **18**, 411–424.

Browman, C. P., and Goldstein, L. (**1992**). ''Articulatory phonology: An overview,'' Phonetica **49**, 155–180.

Calderón, J., and Best, C. T. (**1996**). ''Effects of bilingualism on the perception of nonnative consonant voicing contrasts,'' J. Acoust. Soc. Am. **99**, 2602.

Crowder, R. G. (**1971**). ''The sound of vowels and consonants in immediate memory,'' J. Verbal Learn. Verbal Behav. **10**, 587–596.

Crowder, R. G. (**1973**). ''Representation of speech sounds in precategorical acoustic storage,'' J. Exp. Psychol. **98**, 14–24.

Diehl, R., and Kluender, K. R. (**1989**). ''On the objects of speech perception,'' Ecological Psychol. **1**, 1–45.

Dooling, R. J. (**1989**). ''Perception of complex, species-specific vocalizations by birds and humans,'' in *The Comparative Psychology of Audition: Perceiving Complex Sounds*, edited by R. J. Dooling and S. H. Hulse (Erlbaum, Hillsdale, NJ).

Eimas, P. D. (**1975**). ''Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants,'' Percept. Psychophys. **18**, 341–347.

Eimas, P. D. (**1991**). ''Comments: Some effects of language acquisition on speech perception,'' in *Modularity and the Motor Theory of Speech Perception*, edited by I. G. Mattingly and M. Studdert-Kennedy (Erlbaum, Hillsdale, NJ).

Flege, J. E. (**1986**). ''The production and perception of foreign language speech sounds,'' in *Human Communication and its Disorders*, edited by H. Winitz (Ablex, Norwood, NJ), Vol. 2, pp. 224–401.

Flege, J. E. (**1989**). ''Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training,'' J. Acoust. Soc. Am. **86**, 1684–1697.

Flege, J. E. (**1990**). ''Perception and production: The relevance of phonetic input to L2 language learning,'' in *Crosscurrents in Second Language Acquisition and Linguistic Theories*, edited by C. Ferguson and T. Heubner (J Benjamin, Philadelphia).

Flege, J. E. (**1995**). ''Second language speech learning: Theory, findings and problems,'' in *Speech Perception and Linguistic Experience: Issues in Cross-language Speech Research*, edited by W. Strange (York, Timonium, MD), pp. 233-272.

Fowler, C. A. (**1986**). ''An event approach to the study of speech perception from a direct-realist perspective,'' J. Phonetics **14**, 3–28.

Fowler, C. A. (**1989**). ''Real objects of speech perception: A commentary on Diehl and Kluender,'' Ecological Psychol. **1**, 145–160.

Fowler, C. A., Best, C. T., and McRoberts, G. W. (**1990**). ''Young infants' perception of liquid coarticulatory influences on following stop consonants,'' Percept. Psychophys. **48**(6), 559–570.

Francis, W. N., and Kucera, H. (**1982**). *Frequency analysis of English usage: Lexicon and grammar* (Houghton-Mifflin, Boston).

Frieda, E. M., Walley, A. C., Flege, J. E., and Sloane, M. E. (**1999**). ''Adults' perception of native and nonnative vowels: Implications for the perceptual magnet effect,'' Percept. Psychophys. **61**, 561–577.

Gibson, J. J., and Gibson, E. J. (**1955**). ''Perceptual learning: Differentiation or enrichment?'' Psychol. Rev. **62**, 32–41.

Goto, H. (**1971**). ''Auditory perception by normal Japanese adults of the sounds 'L' and 'R','' Neuropsychologia **9**, 317–323.

Grieser, D. L., and Kuhl, P. K. (**1989**). ''Categorization of speech by infants: Support for speech-sound prototypes,'' Developmental Psychology **25**, 577–588.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (**2000**). ''An investigation of second language speech perception: The case of Japanese adults' perception of English consonants,'' J. Acoust. Soc. Am. **107**, 2711–2724.

Hallé, P. A., and de Boysson-Bardies, B. (**1996**). ''The format of representation of recognized words in infants' early receptive lexicon,'' Infant Behavior and Development **19**, 463–481.

Hallé, P. A., Best, C. T., and Levitt, A. (**1999**). ''Phonetic versus phonological influences of French listeners' perception of American English approximants,'' J. Phonetics **27**, 281–306.

Hallé, P. A., Chéreau, C., and Segui, J. (**2000**). ''Where is the /b/ in 'absurde'? It is in French listeners' minds,'' J. Memory Language **43**, 618–639.

Hallé, P. A., Segui, J., Frauenfelder, U., and Meunier, C. (**1998**). ''The processing of illegal consonant clusters: A case of perceptual assimilation?'' J. Experimental Psychology: Human Perception and Performance **24**, 1–17.

Iverson, P., and Kuhl, P. K. (**1996**). ''Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/,'' J. Acoust. Soc. Am. **99**, 1130–1140.

Keller, E. (**1994**). *Signalyze: Signal Analysis for Speech and Sound: Version 3.0* (Infosignal, Inc., Lausanne, Switzerland).
Kluender, K. (**1994**). ''Speech perception as a tractable problem in cognitive science,'' in *Handbook of Psycholinguistics*, edited by M. A. Gernsbacher (Academic, San Diego, CA).
Kuhl, P. K. (**1988**). ''Auditory perception and the evolution of speech,'' Human Evolution **3**, 19–43.
Kuhl, P. K. (**1991**). ''Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not,'' Perception and Psychophysics **50**(2), 93–107 .
Kuhl, P. K. (**1992**). ''Speech prototypes: Studies on the nature, function, ontogeny and phylogeny of the ''centers'' of speech categories,'' in *Speech Perception, Production and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (Ohmsha, Tokyo), pp. 239–264.
Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (**1992**). ''Linguistic experience alters phonetic perception in infants by 6 months of age,'' Science **255**, 606–608.
Ladefoged, P. and Maddieson, I. (**1996**). *The Sounds of the World's Languages* (Blackwell, Malden, MA).
Liberman, A. M., and Mattingly, I. G. (**1989**). ''A specialization for speech perception,'' Science **243**, 489–494.
Liberman, A. M., Cooper, F. S., Shankweiler, D. S., and Studdert-Kennedy, M. (**1967**). ''Perception of the speech code,'' Psychological Review **74**, 431–461.
Lisker, L., and Abramson, A. S. (**1967**). ''The voicing dimension: Some experiments in comparative phonetics,'' Haskins Laboratories Status Report on Speech Research SR11, pp. 9–15.
Lively, S. E. (**1993**). ''An examination of the perceptual magnet effect,'' J. Acoust. Soc. Am. **93**, 2423.
Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). ''Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories,'' J. Acoust. Soc. Am. **94**, 1242–1255.
Lively, S. E., Logan, J. S., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (**1994**). ''Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories,'' J. Acoust. Soc. Am. **96**, 2076–2087.
Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). ''Training Japanese listeners to identify English /r/ and /l/: A first report,'' J. Acoust. Soc. Am. **89**, 874–886.
Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1998**). ''Depolarizing the perceptual magnet effect,'' J. Acoust. Soc. Am. **103**, 3648–3655.
Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1996**). ''Effects of language experience on perceptual organization of vowel sounds,'' presented at the Fifth Conference on Laboratory Phonology, Northwestern University, Chicago, 6 July.
MacKain, K. S. (**1982**). ''Assessing the role of experience on infants' speech discrimination,'' J. Child Language **9**, 527–542.
MacKain, K. S., Best, C. T., and Strange, W. (**1981**). ''Categorical perception of English /r/ and /l/ by Japanese bilinguals,'' Applied Psycholinguistics **2**, 369–390.
MacMillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge U. P., Cambridge, England).
Maddieson, I. (**1984**). *Patterns of Sounds* (Cambridge U. P., Cambridge).
Miller, G. A., and Nicely, P. E. (**1955**). ''Perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.
Miller, J. L. (**1994**). ''On the internal structure of phonetic categories,'' Cognition **50**, 271–285.
Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (**1975**). ''An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English,'' Percept. Psychophys. **18**, 331–340.
Mochizuki, M. (**1981**). ''The identification of /r/ and /l/ in natural and synthesized speech,'' J. Phonetics **9**, 283–303.
Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (**1989**). ''The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults,'' J. Speech Hear. Res. **32**, 120–132.
Pallier, C., Bosch, L., and Sébastian-Galles, N. (**1997**). ''A limit on behavioral plasticity in speech perception,'' Cognition **64**, 9–17.
Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (**1982**). ''Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants,'' J. Exp. Psychol. **8**, 297–314.
Polivanov, E. (**1931**). ''La perception des sons d'une langue étrangère,'' Travaux du Cercle Linguistique de Prague **4**, 79–96.
Polka, L. (**1991**). ''Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions,'' J. Acoust. Soc. Am. **89**, 2961–2977.
Polka, L. (**1992**). ''Characterizing the influence of native experience on adult speech perception,'' Percept. Psychophys. **52**, 37–52.
Polka, L., and Bohn, O-S. (**1996**). ''A cross-language comparison of vowel perception in English-learning and German-learning infants,'' J. Acoust. Soc. Am. **100**, 577–592.
Polka, L., and Werker, J. F. (**1994**). ''Developmental changes in perception of nonnative vowel contrasts,'' J. Experimental Psychology: Human Perception and Performance **20**, 421–435.
Ruhlen, M. (**1975**). *A Guide to the Languages of the World* (Stanford Univ., Palo Alto, CA).
Scarborough, H. S. (**1998**). ''Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and iQ,'' Annals of Dyslexia **48**, 115–136.
Shankweiler, D., Crain, S., Katz, L., Fowler, A. E., Liberman, A. M., Brady, S. A., Thornton, R., Lundquist, E., Dreyer, L., Fletcher, J. M., Stuebing, K. K., Shaywitz, S. E., and Shaywitz, B. A. (**1995**). ''Cognitive profiles of reading-disabled children: Comparison of language skills in phonology, morphology, and syntax,'' Psychological Science **6**, 149–156.
Stager, C. L., and Werker, J. F. (**1997**). ''Infants listen for more phonetic detail in speech perception than in word-learning tasks,'' Nature (London) **388**, 381–382.
Strange, W., and Dittmann, S. (**1984**). ''Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English,'' Percept. Psychophys. **36**, 131–145.
Sussman, J. E., and Lauckner-Morano, V. J. (**1995**). ''Further tests of the ''perceptual magnet effect'' in the perception of [i]: Identification and change-no-change discrimination,'' J. Acout. Soc. Am. **97**, 539–552.
Takagi, N., and Mann, V. A. (**1995**). ''Signal detection modeling of Japanese listeners' /r/-/l/ labeling behavior in a one-interval identification task,'' J. Acoust. Soc. Am. **97**, 563–574.
Tees, R. C., and Werker, J. F. (**1984**). ''Perceptual flexibility: Maintenance or recovery of the ability to discriminate nonnative speech sounds,'' Can. J. Psychol. **38**, 579–590.
Trubetzkoy, N. S. (**1939/1969**). *Principles of Phonology*, translated by C. A. M. Baltaxe (Univ. California, Berkeley, CA).
Volaitis, L. E., and Miller, J. L. (**1992**). ''Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories,'' J. Acoust. Soc. Am. **92**, 723–735.
Werker, J. F. (**1989**). ''On becoming a native listener,'' Am. Sci. **77**, 54–59.
Werker, J. F., and Lalonde, C. E. (**1989**). ''Cross-language speech perception: Initial capabilities and developmental change,'' Developmental Psychology **24**, 672–683.
Werker, J. F., and Logan, J. (**1985**). ''Cross-language evidence for three factors in speech perception,'' Percept. Psychophys. **37**, 35–44.
Werker, J. F., and Pegg, J. (**1992**). ''Infant speech perception and phonological acquistion,'' in *Phonological development: Models, research, implications*, edited by C. A. Ferguson, L. Menn, and C. Stoel-Gammon (York Press, Timonium, MD), pp. 285–311.
Werker, J. F., and Tees, R. C. (**1984**). ''Phonemic and phonetic factors in adult cross-language speech perception,'' J. Acoust. Soc. Am. **75**, 1866–1878.
Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. (**1981**). ''Developmental aspects of cross-language speech perception,'' Child Development **52**, 349–355.
Whalen, D. H. (**1984**). ''Subcategorical phonetic mismatches slow phonetic judgments,'' Percept. Psychophys. **35**, 49–64.
Whalen, D. H. (**1991**). ''Subcategorical phonetic mismatches and lexical access,'' Percept. Psychophys. **50**, 351–360.
Whalen, D. H., Wiley, E. R., Rubin, P. E., and Cooper, F. S. (**1990**). ''The Haskins Laboratories' pulse code modulation (PCM) system,'' Behav. Res. Methods Instrum. Comput. **22**, 550–559.
Yamada, R., and Tohkura, Y. (**1992**). ''The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners,'' Percept. Psychophys. **52**, 376–392.

Radiation impedance of a finite circular piston on a viscoelastic half-space with application to medical diagnosis

X. Zhang and T. J. Royston[a)]
University of Illinois at Chicago, Chicago, Illinois 60607

H. A. Mansy and R. H. Sandler
Rush Medical College, Chicago, Illinois 60612

(Received 6 May 2000; accepted for publication 16 October 2000)

In a recent study a new analytical solution was developed and validated experimentally for the problem of surface wave generation on a linear viscoelastic half-space by a rigid circular disk located on the surface and oscillating normal to it. The results of that study suggested that, for the low audible frequency range, some previously reported values of shear viscosity for soft biological tissues may be inaccurate. Those values were determined by matching radiation impedance measurements with theoretical calculations reported previously. In the current study, the sensitivity to shear viscoelastic material constants of theoretical solutions for radiation impedance and surface wave motion are compared. Theoretical solutions are also compared to experimental measurements and numerical results from finite-element analysis. It is found that, while prior theoretical solutions for radiation impedance are accurate, use of such measurements to estimate shear viscoelastic constants is not as precise as the use of surface wave measurements. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1334598]

PACS numbers: 43.80.Cs, 43.80.Ev, 43.80.Qf, 43.20.Bi [FD]

I. INTRODUCTION

In a recent study[1] a new analytical solution was developed for the problem of surface wave generation on a linear viscoelastic half-space by a rigid circular disk located on the surface and oscillating normal to it. See Fig. 1 for a description of this pedagogical problem. The new solution was an incremental advancement of theoretical work reported in seminal articles focused on seismology.[2,3] With an interest in medical diagnostics, the theory was verified experimentally using a viscoelastic phantom with material properties comparable to biological soft tissue. Findings suggested that prior estimates in the literature of the shear viscosity in human soft tissue may not be accurate in the low audible frequency range. The suspect values were determined by matching radiation impedance measurements for this same setup (see Fig. 1) with theoretical calculations reported previously in seminal papers by Oestreicher[4] and von Gierke *et al.*[5]

These previously reported discrepancies raise several questions, which it is the purpose of this follow-up article to address. While Oestreicher[4] derived an expression for radiation impedance, the prior derivation of the authors[1] was for surface wave propagation. To further compare these two formulations, it is desirable to extend numerical finite-element studies and experimental measurements to both radiation impedance and surface wave propagation. Particularly, we are interested in examining the sensitivity of the radiation impedance and surface wave propagation to changes in the viscoelastic Lame constants.

With these points in mind, the specific tasks of the present study are to:

(1) Review the formulation of Oestreicher[4] and Miller and Pursey[2,3] for radiation impedance, identifying any limiting assumptions. (While Miller and Pursey's theoretical developments formed the basis for the authors' previously reported analytical solution for surface wave propagation, they also derived an integral expression that can be numerically solved to approximate radiation impedance.)
(2) Compare the radiation impedance formulations for the following cases: (a) purely elastic half-space cases with material properties relevant to seismologic and structural (metallic) problems; and (b) a viscoelastic phantom half-space model that approximates soft biological tissue and for which experimental results and a finite-element simulation of radiation impedance and surface wave propagation are available.
(3) Assess the relative sensitivity of radiation impedance and surface wave propagation to viscoelastic Lame constant values via a parametric study to determine which, if either, type of measurement may be more precise in identifying values for these constants.

Measurement of skin surface vibration has been studied by a number of researchers for rapid, nonintrusive diagnosis of a variety of specific medical ailments.[6–18] For example, Lee[6] and Hong and Fox[7,8] have investigated noncontact characterization of cardiovascular dynamics using optical interferometry. Diagnosis of certain types of edema conditions and other skin diseases, including skin cancer, via alterations in surface and shear wave propagation has been studied by several groups.[9–16] Other research has focused on character-

a)Author to whom correspondence should be addressed. Electronic mail: troyston@uic.edu

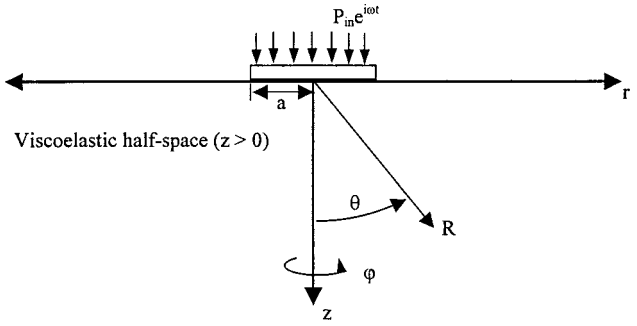


FIG. 1. Ideal viscoelastic half-space problem.

izing lung edema by studying surface wave propagation directly on the inflated pulmonary parenchyma[17] or on the chest wall.[18] It is hoped that the developments reported here will advance these techniques and also provide insight into related diagnostic methods, such as sonoelastic imaging[19–21] and other methodologies that utilize disease-related variations in soft tissue shear viscoelastic properties.

II. RADIATION IMPEDANCE DETERMINED BY OESTREICHER AND VON GIERKE *ET AL.*

For all cases considered here, the assumption is that of an isotropic, homogeneous, viscoelastic compressible medium (Voigt's body) for which one can use either of the following formulations of the equation of motion for small perturbations about an operating point:

$$(\lambda+\mu)\nabla\nabla\cdot\mathbf{u}+\mu\nabla^2\mathbf{u}=\rho\ddot{\mathbf{u}}, \tag{1a}$$

or

$$(\lambda+2\mu)\nabla\nabla\cdot\mathbf{u}-\mu\nabla\times\nabla\times\mathbf{u}=\rho\ddot{\mathbf{u}}. \tag{1b}$$

Here, $\lambda=\lambda_1+(\partial/\partial t)\lambda_2$ and $\mu=\mu_1+(\partial/\partial t)\mu_2$, where λ and μ are the linear viscoelastic Lame constants with λ_1 referring to volume compressibility, λ_2 referring to volume viscosity, μ_1 denoting shear elasticity, and μ_2 denoting shear viscosity. Density of the medium is denoted by ρ and displacements in the medium are denoted by $\mathbf{u}=[u_r,u_\phi,u_z]^T$ in terms of polar coordinates. Also, with respect to polar coordinates we have the following where j denotes the unit vector in the ϕ direction:

$$\nabla\cdot\mathbf{u}=\frac{1}{r}\frac{\partial}{\partial r}(ru_r)+\frac{\partial u_z}{\partial z}, \tag{2a}$$

and

$$\nabla\times\mathbf{u}=\left(\frac{\partial u_r}{\partial z}-\frac{\partial u_z}{\partial r}\right)j. \tag{2b}$$

In an infinite medium of this type, Oestreicher[4] considered small oscillatory motion of a rigid sphere of radius ''a'' in the z direction. Radiation impedance was defined as the ratio of the z component (vertical) of the force P acting on the sphere divided by the velocity of the center of the sphere $\dot{u}_0$. With the assumption of no cavitation, the following expression for this radiation impedance was derived:

$$\begin{aligned}Z(\omega)&=\frac{P}{i\omega u_0}\\&=-\frac{4}{3}\pi\rho\omega a^3 i\left[\left(1-\frac{3i}{ak_2}-\frac{3}{a^2k_2^2}\right)\right.\\&\quad\left.-2\left(\frac{i}{ak_2}+\frac{1}{a^2k_2^2}\right)\left(3-\frac{a^2k_1^2}{ak_1i+1}\right)\right]\\&\quad\div\left[\left(\frac{i}{ak_2}+\frac{1}{a^2k_2^2}\right)\frac{a^2k_1^2}{ak_1i+1}+\left(2-\frac{a^2k_1^2}{ak_1i+1}\right)\right].\end{aligned} \tag{3}$$

Here, $i=\sqrt{-1}$, ω refers to the circular frequency in radians per second, $k_1=\sqrt{\rho\omega^2/(2\mu+\lambda)}$ is the compression wave number, and $k_2=\sqrt{\rho\omega^2/\mu}$ is the shear wave number. To approximate the case of a sphere vibrating normal to the surface of a viscoelastic half-space, it logically follows that the impedance should be one-half of the above expression. As an approximation to a circular piston of radius ''a'' vibrating normal to the surface of a viscoelastic half-space, von Gierke *et al.*[5] suggested a further division by a factor of 1.18. This is based on the ratio of the static stiffness of a sphere on an elastic half-space to that of a piston of the same radius on an elastic half-space, which for both static cases exact solutions can be determined.

According to Oestreicher,[4] the assumptions of isotropy and homogeneity will hold for muscle tissue approximately up to 200 kHz. (While this is certainly a questionable statement, particularly as frequency increases, it is the authors' contention that, at least up to a hundred Hz, the range of interest in the present study, the isotropic, homogeneous model does have relevance to soft biological tissue.) The possibility of cavitation is entirely omitted in the analysis.

III. RADIATION IMPEDANCE DETERMINED BY MILLER AND PURSEY

In the papers of Miller and Pursey,[2,3] the problem of an oscillating circular piston of radius ''a'' on the surface of an elastic half-space was directly considered. Some discussion of the extrapolation of the result to the case of a viscoelastic half-space was also provided, though results for a specific viscoelastic example cases were not presented. The starting point was the same equations that were used to describe the medium in Oestreicher's derivation (1a)–(1b), that of a linear viscoelastic Voigt's body. However, several different assumptions were made. In particular, it was assumed that the piston imparted a uniform normal stress on the surface within the region covered by the disk. This approximation is likely to restrict the validity of the solution to the case of a small piston radius ''a'' relative to the surface wavelength. In other words, there will be an upper frequency limit for the solution's validity.

Consider the case of harmonic force excitation of the massless disk of amplitude per unit area P_{in} and circular frequency ω such that it applies a uniform stress on the surface of the medium in the circular region of $r<a$. Miller and Pursey[2] derive the following integral expression for wave

propagation in the r and z directions at any location in and on the surface of the medium, i.e., $z \geqslant 0$. Note that time dependence $e^{i\omega t}$ is omitted where $i=\sqrt{-1}$.

$$\frac{u_z}{P_{in}} = \frac{a}{\mu}\int_0^\infty \frac{J_1(\zeta a k_1)\sqrt{\zeta^2-1}}{F_0(\zeta)}\{2\zeta^2 e^{-zk_1\sqrt{\zeta^2-\eta^2}} + (\eta^2-2\zeta^2)e^{-zk_1\sqrt{\zeta^2-1}}\}J_0(\zeta r k_1)d\zeta, \tag{4a}$$

$$\frac{u_r}{P_{in}} = \frac{a}{\mu}\int_0^\infty \frac{J_1(\zeta a k_1)\zeta}{F_0(\zeta)}\{2\sqrt{\zeta^2-1}\sqrt{\zeta^2-\eta^2}e^{-zk_1\sqrt{\zeta^2-\eta^2}} + (\eta^2-2\zeta^2)e^{-zk_1\sqrt{\zeta^2-1}}\}J_1(\zeta r k_1)d\zeta, \tag{4b}$$

where

$$F_0(\zeta) = (2\zeta^2-\eta^2)^2 - 4\zeta^2\sqrt{\zeta^2-\eta^2}\sqrt{\zeta^2-1}, \tag{4c}$$

$$\eta = k_2/k_1, \tag{4d}$$

$$k_1 = \omega\sqrt{\rho/(\lambda+2\mu)}, \tag{4e}$$

and

$$k_2 = \omega\sqrt{\rho/\mu}. \tag{4f}$$

In the above equations, J_0 and J_1 refer to Bessel functions of the first kind. The dummy variable ζ is used to denote integration over the wave number domain that has been normalized with respect to k_1.

From the above expression, Miller and Pursey[2] derived an integral expression for the mobility (reciprocal of impedance) that could be solved numerically. Unlike Oestreicher,[4] the definition of impedance in this case is the ratio of the applied stress P_{in} (Pa) on the piston to the mean vertical velocity $\bar{u}_z(m)$ on the circular region of the surface located underneath the piston. Here, all length dimensions are nondimensionalized by the compression wave number k_1 ($a_1 = ak_1$, $r_1 = rk_1$, $z_1 = zk_1$, and $u_{z1} = u_z k_1$)

$$\frac{\overline{u_{z1}}}{P_{in}} = \frac{2\eta^2 a_1}{\mu}\int_0^\infty \frac{\zeta\sqrt{\zeta^2-1}}{F_0(\zeta)}\frac{\{J_1(\zeta a_1)\}^2}{\zeta a_1}d\zeta. \tag{5}$$

Equation (5) requires numerical evaluation. For the elastic case, there are three singularities along the path of integration at 1, η, and p with $1<\eta<p$. The lower two singularities, 1 and η, are branch points and are inversely related to the phase speed of longitudinal (compression) and bulk shear wave motion, respectively. The root of $F_0(\zeta)$, i.e., the pole of the integrand, is denoted as p; it is inversely related to the phase speed of surface waves. Miller and Pursey[2] describe a solution methodology for the above integral for the elastic case. For the viscoelastic case, η and p will be complex, residing below the real axis in the fourth quadrant. For this case, numerical integration of the above expression can be accomplished by separating the integration into three parts, from 0 to 1, from 1 to α (where α is a finite large number), and from α to ∞. The last integral, α to ∞, is performed as described in the reference, including the lower bounds of α which will still yield an accurate approximation.[2] For the viscoelastic cases considered in this study, evaluation of the first two integrals was performed using the ''NIntegrate'' command in MATHEMATICA® 4.0.1.0 software with the ''default'' options.

TABLE I. Comparison of theoretical impedance predictions for $\upsilon = 1/4$ and $\eta = \sqrt{3}$.

	$\overline{Z}$	
a_1	Oestreicher (Refs. 4, 5)	Miller and Pursey (Ref. 2)
0.05	1.1823−14.5256i	0.66−10.46i
0.1	1.1825−7.2595i	0.66−5.22i
0.2	1.1835−3.6232i	0.66−2.58i
0.3	1.1852−2.4084i	0.67−1.68i
0.4	1.1875−1.7994i	0.67−1.22i
0.5	1.1903−1.4328i	0.68−0.94i

IV. COMPARISON OF THEORIES FOR ELASTIC EXAMPLE CASES

To compare the theoretical derivations for radiation impedance of the finite disk, elastic and viscoelastic example cases are considered. The two elastic case examples given in Miller and Pursey[2] are considered where for case 1: $\nu = 1/4$, $\eta = \sqrt{3}$ and for case 2: $\nu = 1/3$, $\eta = 2$; here, ν refers to Poisson's ratio and $\eta = k_2/k_1$. The first case is typical of geological materials and the second case is typical of structural (metallic) materials. Calculated values for the two cases are given in Tables I and II. Here, Eq. (3) is nondimensionalized by dividing impedance by disk area, medium density, and compression wave speed to match the format of Miller and Pursey; also, adjustments are made per von Gierke *et al.*[5] (Sec. II)

$$\overline{Z} = -\frac{4}{3}a_1 i\left[\left(1-\frac{3i}{a_1\eta}-\frac{3}{a_1^2\eta^2}\right)-2\left(\frac{i}{a_1\eta}+\frac{1}{a_1^2\eta^2}\right)\times\left(3-\frac{a_1^2}{a_1 i+1}\right)\right]\div\left[\left(\frac{i}{a_1\eta}+\frac{1}{a_1^2\eta^2}\right)\frac{a_1^2}{a_1 i+1} + \left(2-\frac{a_1^2}{a_1 i+1}\right)\right]\div 2 \div 1.18. \tag{6}$$

Oestreicher's theory for the oscillating sphere should be applicable to a wide frequency range as long as cavitation does not occur; however, geometrically it is an approximation. Miller and Pursey's[2] integral expression is limited in frequency due to the assumption of a uniform normal stress region under the disk surface. For the example cases considered, Oestreicher's theory consistently yields both a larger resistance value (the real part of the impedance), about 75%

TABLE II. Comparison of theoretical impedance predictions for $\upsilon = 1/3$ and $\eta = 2$.

	$\overline{Z}$	
a_1	Oestreicher (Refs. 4, 5)	Miller and Pursey (Ref. 2)
0.05	1.0673−11.2960i	0.64−8.82i
0.1	1.0677−5.6428i	0.64−4.39i
0.2	1.0691−2.8111i	0.64−2.16i
0.3	1.0716−1.8630i	0.65−1.40i
0.4	1.0750−1.3861i	0.65−1.01i
0.5	1.0792−1.0982i	0.66−0.76i

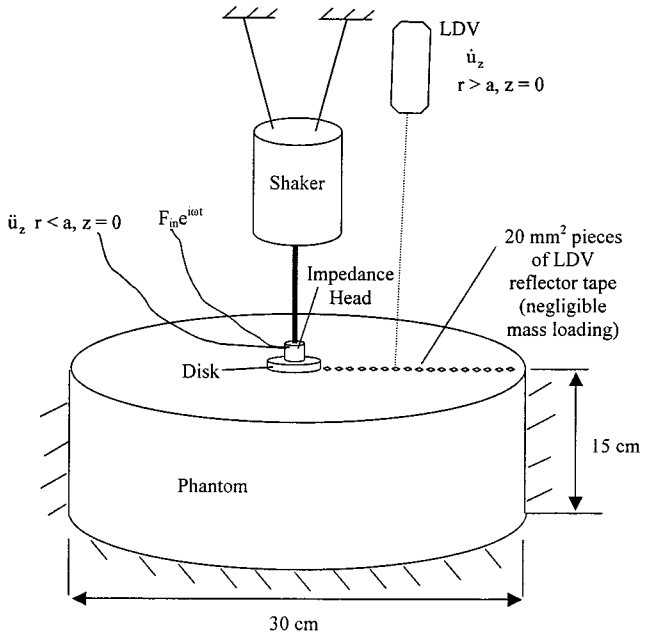


FIG. 2. Schematic of experimental setup.

and 65% larger than Miller and Pursey's theory in Tables I and II, respectively, and a larger reactance value (the imaginary part of the impedance), about 40% and 30% larger than Miller and Pursey's theory in Tables I and II, respectively. Note that as the disk radius increases, while the relative difference between resistances is fairly constant, the reactance value of Oestreicher increases further over the value of Miller and Pursey. Perhaps this is due to the limitation of the uniform normal stress assumption of the Miller and Pursey theory.

V. COMPARISON OF THEORETICAL, EXPERIMENTAL, AND NUMERICAL SOLUTIONS FOR A VISCOELASTIC EXAMPLE CASE

A. Description of the experiment

To compare the theoretical derivations presented above for a case with material properties comparable to soft biological tissue, an experiment was constructed that is schematically depicted in Fig. 2. This is the same experimental setup discussed in Royston *et al.*[1] The viscoelastic medium consists of a gel mixture of the following composition (per liter of water): 70 grams gelatin, 40 grams *n*-propanol, and 4 grams formaldhyde (37% solution). This ''recipe'' for a soft tissue phantom is based on prior investigations.[22] By taking velocity measurements using the laser Doppler vibrometer (Polytec model CLV-800-FF/1000), vertical velocity $\dot{u}_z$ can be measured for any $r>a$ at the surface, $z=0$. An impedance head (PCB model 288B02) records acceleration and force input to the Plexiglas disk at $z=0$ and $0<r<a$. The vibratory excitation is delivered through a stinger by an electromagnetic shaker (Labworks model ET-132-2) that is flexibly suspended above the phantom. Sensor output signals are recorded with a Hewlett Packard (Agilent Technologies) 35670 dynamic signal (FFT) analyzer, which also performs preliminary analysis. The analyzer also provides the chirp excitation signal to the amplifier that drives the shaker. Chirp (rapidly swept sinusoidal) excitation signals were used to enable extracting frequency-amplitude-phase-time information. This allowed detection of outgoing waves from the source independent from reflections due to the finite boundaries of the phantom. Beyond basic frequency response functions available in real time using the 35670, more detailed analyses of the experimental data were conducted using MATLAB® software.

Finite-element simulations (see the next section) were conducted with the same size and larger overall model dimensions, twice the depth (30 cm) and twice the radius (30 cm). Negligible differences in results suggested that for the impedance and surface wave studies conducted here, the experimental model was sufficiently large enough to simulate an infinite viscoelastic half-space.

B. Description of finite-element model and solution technique

ANSYS® Version 5.5 finite-element analysis (FEA) software running on an HP 9000/800 V2250 was used to numerically simulate the system depicted in Fig. 2. For the results depicted in the following figures, Structural Solid axisymmetric elements were used with dimensions of 5 $\times$5 mm. (Some simulations were conducted with reduced element dimensions of 2.5$\times$2.5 mm with no noticeable difference in the types of results obtained, which suggests that the model has sufficient resolution.)

In ANSYS, Young's modulus E and Poisson's ratio ν were specified based on the following equations:

$$E=\frac{\mu(3\lambda+2\mu)}{\lambda+\mu}, \tag{7}$$

$$\nu=\frac{\lambda}{2(\lambda+\mu)}, \tag{8}$$

where, for harmonic excitation at circular frequency ω we have $\lambda=\lambda_1+i\omega\lambda_2$ and $\mu=\mu_1+i\omega\mu_2$.

For the viscoelastic case, there will be a nonzero imaginary part of E that is related to linear viscous (rate-dependent) damping. In ANSYS, this component is specified via the damping parameter β (damping proportional to stiffness), which equals the imaginary part of E (not including ω) divided by the real part of E. (There is an imaginary part to ν but it is negligible.) As an example, given complex Lame constants of $\lambda_1=2.6\times10^9$ N/m^2, $\lambda_2=0$ Ns/m^2, $\mu_1=4.5\times10^3$ N/m^2, and $\mu_2=4$ Ns/m^2, we have $E\approx13\,500+12\omega i$, $\nu\approx0.499\,999\,1-3.8\times10^{-10}\omega i$, $\beta\approx0.000\,89$, and at 60 Hz $\eta=730.5-119.2i$.

To calculate the radiation impedance of the disk, a harmonic response analysis was conducted in ANSYS, specifying a harmonic displacement condition on the surface within the region $r<a$. Impedance was calculated as the total force applied in the disk region divided by the velocity. To calculate surface wave propagation away from the disk, a harmonic analysis was also conducted. The same displacement load was applied on the surface within the region $r<a$.

TABLE III. Comparison of theoretical impedance predictions for $\mu_1 = 4.5 \times 10^3$ N/m^2, $\mu_2 = 4$ N s/m^2, $\lambda_1 = 2.6 \times 10^9$ N/m^2, $\lambda_2 = 0$ N s/m^2, and $\rho = 1000$ kg ($v \approx 0.499\,991$). Here, the dimensionalized disk radius is kept constant at $a = 1.0$ cm.

$\omega/2\pi$ (Hz)	η	a_1	$\bar{Z}$	
			Oestreicher (Refs. 4, 5)	Miller and Pursey (Ref. 2)
30	$752-62.6i$	0.001 17	$0.003\,99-0.003\,15i$	$0.002\,97-0.002\,74i$
40	$746-82.4i$	0.001 56	$0.004\,00-0.002\,01i$	$0.002\,97-0.001\,61i$
50	$739-101i$	0.001 95	$0.004\,01-0.001\,25i$	$0.002\,97-0.000\,821i$
60	$731-119i$	0.002 34	$0.004\,02-0.000\,669i$	$0.002\,98-0.000\,205i$
70	$721-136i$	0.002 73	$0.004\,04+0.000\,201i$	$0.002\,99+0.000\,315i$
80	$710-152i$	0.003 12	$0.004\,05+0.000\,199i$	$0.003\,01+0.000\,773i$
90	$699-166i$	0.003 51	$0.004\,07+0.000\,552i$	$0.003\,05+0.001\,19i$
100	$687-179i$	0.003 40	$0.004\,09+0.000\,873i$	$0.003\,09+0.001\,57i$

C. Results and discussion

In a prior study by the authors,[1] values for the complex Lame constants of the experimental setup of Fig. 2 were estimated based on matching theoretical predictions of the surface wave motion with experimental measurements, in terms of amplitude, decay rate, wavelength, and wave speed. In the present article the values previously determined for $\mu = \mu_1 + i\omega\mu_2$ are used to predict radiation impedance theoretically and to determine radiation impedance and surface wave motion numerically using ANSYS finite-element analysis. [As in previous studies[1,4] it is assumed that $\lambda_1 = 2.6 \times 10^9$ N/m^2 (water) and $\lambda_2 = 0$. Also, $\rho = 1000$ kg/m^3.] Additionally, theoretical and numerical parametric studies are conducted for different values of μ_1 and μ_2.

Radiation impedance calculations using Oestreicher's and Miller and Pursey's formulations are compared in Table III in a similar format to that of Tables I and II. However, note that for a fixed (dimensionalized) disk radius of 1 cm the nondimensional disk radius $a_1 = ak_1$ varies as a function of frequency. Also, $\eta = k_2/k_1$ varies with frequency. It is seen that viscoelastic Lame constants comparable to soft biological tissue result in substantially different values for η and ν as compared to Tables I and II. A significant imaginary part to η is due to shear viscous effects. The Poisson ratio ν is very close to that of an incompressible medium. Despite these differences as compared to the elastic example cases of Sec. IV, there are similar trends in the relative values for radiation impedance based on Oestreicher's and Miller and Pursey's theories. In this case, Oestreicher's resistance value is consistently 33% larger than that of Miller and Pursey's. It is more difficult to assess the relative difference for reactance as, in this case, a natural frequency is traversed, with the the impedance going from negative (stiffness-dominated) to positive (mass-dominated) values. Oestreicher's and Miller and Pursey's theories are similar in that they both predict the sign change to occur between 60 and 70 Hz. Additionally, in progressing from 30 to 100 Hz, the reactance increases by 0.004 and 0.004 27 for Oestreicher's and Miller and Pursey's theories, respectively. Hence, overall it appears that there is a closer, but not perfect, match between the two theories for this particular viscoelastic example case (akin to soft biological tissue material properties) relative to the elastic example cases (seismological and metallic structure properties) of Sec. IV.

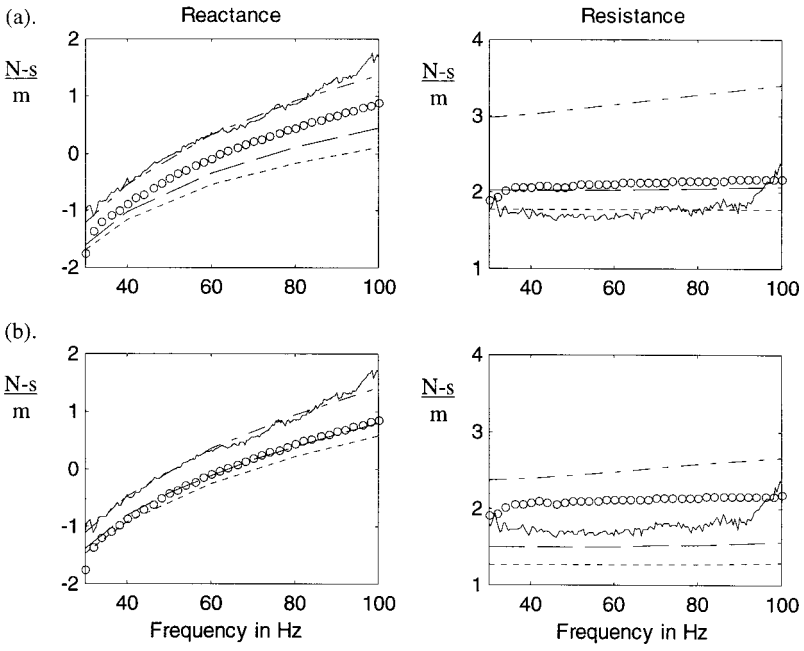


FIG. 3. Comparison of theoretical predictions, finite-element calculations, and experimental measurements of the impedance for the experiment depicted in Fig. 2. Resistance and reactance are the real and imaginary parts of the impedance, respectively. Key: ——— experiment, o o o finite element ($\mu = 4500 + 4i\omega$ N/m^2), - - - theory ($\mu = 4500 + 1i\omega$ N/m^2), –––– theory ($\mu = 4500 + 4i\omega$ N/m^2), –·– theory ($\mu = 4500 + 15i\omega$ N/m^2). (a) Theory of Oestreicher (Ref. 4) and von Gierke *et al.* (Ref. 5). (b) Theory of Miller and Pursey (Refs. 2, 3).

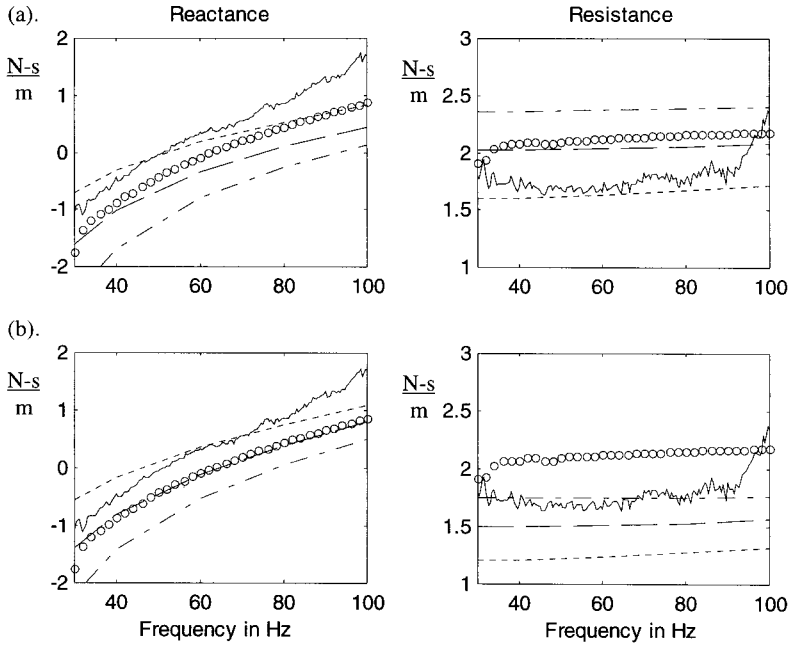


FIG. 4. Comparison of theoretical predictions, finite-element calculations, and experimental measurements of the impedance for the experiment depicted in Fig. 2. Resistance and reactance are the real and imaginary parts of the impedance, respectively. Key: ——— experiment, o o o finite element ($\mu = 4500 + 4i\omega$ N/m^2), - - - theory ($\mu = 2500 + 4i\omega$ N/m^2), ––– theory ($\mu = 4500 + 4i\omega$ N/m^2), –·– theory ($\mu = 6500 + 4i\omega$ N/m^2). (a) Theory of Oestreicher (Ref. 4) and von Gierke *et al.* (Ref. 5). (b) Theory of Miller and Pursey (Refs. 2, 3).

In Figs. 3 and 4 theoretical predictions using the two radiation impedance formulas are compared to numerical finite-element predictions and experimental measurements of radiation impedance, reactance, and resistance, for a range of values for μ_1 and μ_2. The approximate agreement of ANSYS finite-element results and experimental measurements suggests that the selected values for μ_1 and μ_2 based on the authors' surface wave theory are reasonable, though a decreased value for μ_1 may be slightly more accurate (note the trends in Fig. 4). Predictions using Oestreicher's formula appear to more closely approximate the finite-element results, particularly with respect to resistance. While Miller and Pursey's reactance predictions more accurately match finite-element results, the discrepancy with respect to resistance is substantial. Hence, overall, Oestreicher's prediction is judged superior.

In terms of identifying values for μ_1 and μ_2 based on comparing theoretical predictions to experiment, one would need to iterate both μ_1 and μ_2 based on reactance and resistance measurements. Increasing μ_1 results in *decreased reactance* and *increased resistance* (e.g., about 45% increase when μ_1 increases from 2500 to 6500 N/m^2). Increasing μ_2 results in *increased reactance* and *increased resistance* (e.g., about 15% increase when μ_2 increases from 1 to 4 N s/m^2). Hence, a reasonable match could be obtained; but, it may only be accurate for a narrow frequency range. Additionally, some experimental studies have shown that radiation impedance measurements can be sensitive to the bias preload applied to the disk to ensure its contact with the surface at all times.[5,23] This, in part, may be responsible for the range of values reported for μ_1 and particularly μ_2 that are based on variations of the experimental radiation impedance technique compared to Oestreicher's theoretical formula.

In Figs. 5 and 6 theoretical and numerical predictions are compared to experimental measurements of surface wave motion radiating away from the disk. The theoretical predictions are based on the theory previously developed by the authors.[1] The same perturbations in μ_1 and μ_2 that were shown for radiation impedance in Figs. 3 and 4 are used here for theory and ANSYS finite-element numerical simulation. The following observations are made: (1) Theoretical and numerical predictions are both in reasonable agreement with experimental measurements. (While the finite-element results indicate periodic dips in the response, it is believed that these are numerical artifacts that may be related to modal truncation or other effects. They are not evident in the experimental results or theoretical solution. Their location is dependent on the elastic modulus but not on the mesh reso-

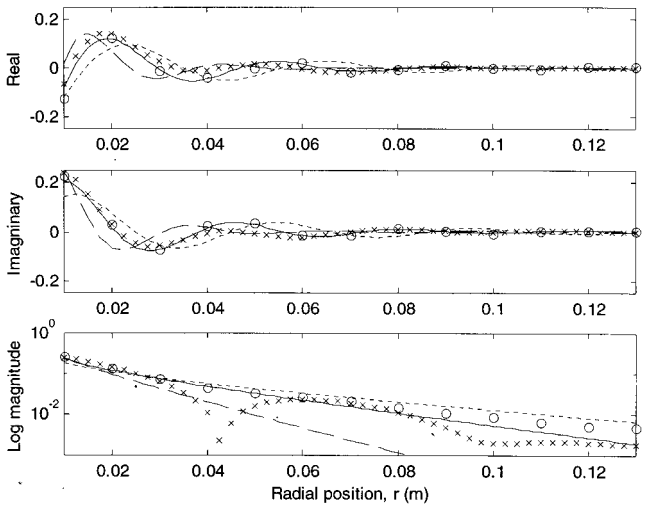


FIG. 5. Comparison of gel phantom experimental measurements with theoretical predictions [Royston *et al.* (Ref. 1)] and finite-element calculations of vertical particle velocity for different values of μ_1 ($\mu_2 = 4$ Ns/m^2). Velocity is taken with respect to excitation force input $\dot{u}_z/F_{\text{in}}$ (m/sN) on the surface as a function of radial position at excitation frequency $\omega/2\pi = 60$ Hz. Key: o o o experiment, × × × finite-element with $\mu_1 = 4500$ N/m^2, ——— theory with $\mu_1 = 4500$ N/m^2, ––– theory with $\mu_1 = 2500$ N/m^2, - - - theory with $\mu_1 = 6500$ N/m^2.

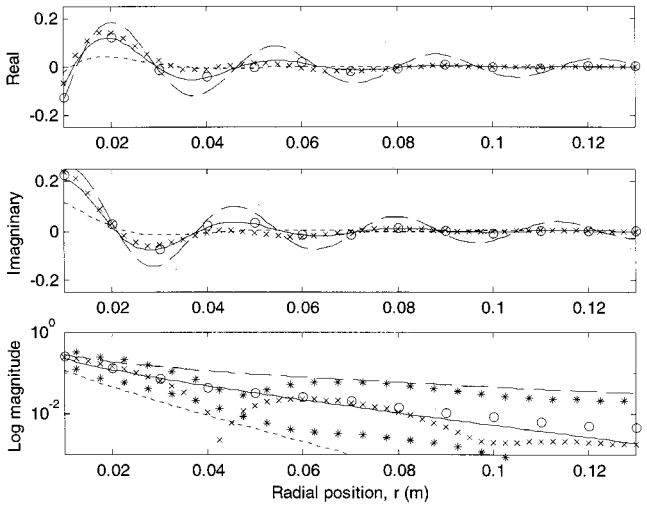


FIG. 6. Comparison of gel phantom experimental measurements with theoretical predictions [Royston *et al.* (Ref. 1)] and finite-element calculations of vertical particle velocity for different values of μ_2 ($\mu_1 = 4500\ \mathrm{N/m^2}$). Velocity is taken with respect to excitation force input $\dot{u}_z / F_{\mathrm{in}}$ (m/sN) on the surface as a function of radial position at excitation frequency $\omega/2\pi = 60$ Hz. Key: o o o experiment, $\times\ \times\ \times$ finite-element with $\mu_2 = 4\ \mathrm{Ns/m^2}$, *** finite element analysis with $\mu_2 = 1$ and 15 $\mathrm{Ns/m^2}$, ——— theory with $\mu_2 = 4\ \mathrm{Ns/m^2}$, −−− theory with $\mu_2 = 1\ \mathrm{Ns/m^2}$, - - - theory with $\mu_2 = 15\ \mathrm{Ns/m^2}$.

lution or overall phantom dimensions.) (2) It appears that surface wave motion is more sensitive to values of μ_1 and μ_2 in distinct ways relative to radiation impedance. Consider the following for $\omega/2\pi = 60$ Hz. The shear viscosity μ_2 strongly affects the rate of attenuation (greater than 300% change in dB/cm attenuation between $\mu_2 = 1$ and 4 $\mathrm{N\,s/m^2}$) but does not influence the surface wavelength that much (only a 4% change between $\mu_2 = 1$ and 4 $\mathrm{N\,s/^2}$). The shear elasticity μ_1 significantly affects both the wavelength of the surface waves (47% change between $\mu_1 = 2500$ and 6500 $\mathrm{N/m^2}$) and attenuation (300% change in dB/cm attenuation between $\mu_1 = 2500$ and 6500 $\mathrm{N/m^2}$). Thus, μ_1 may first be adjusted to predict the correct wavelength. Then, μ_2 can be adjusted to predict the correct attenuation rate. Thus, given the greater and more unique sensitivity of surface waves to the shear viscoelastic Lame constants, it is concluded that estimates of μ_1 and μ_2 can be made more precisely using surface waves measurements as opposed to impedance measurements. Of course, in practice surface wave measurements will require more time and instrumentation relative to impedance measurements; hence, there is a trade-off.

VI. CONCLUSION

The original motivation for the work reported in this paper was a discrepancy in predictions of shear elasticity and viscosity for materials like soft biological tissue. The radiation impedance theory of Oestreicher *et al.*[4,5] yielded predictions over a wide range of values for the shear viscosity μ_2, in particular. The surface wave theory of the authors[1] suggested that the most appropriate value for μ_2 might be at the lower end of the range reported by Oestreicher. The work reported in the present article offers logical explanations for the apparent discrepancies. While the radiation impedance theory of Oestreicher and the surface wave theory of the authors are both shown to be reasonably accurate based on experimental and numerical finite-element studies, the surface wave theory is more sensitive to values of shear elasticity and viscosity in distinct ways. Consequently, use of it will lead to more precise and accurate estimates of the complex Lame viscoelastic material constants.

Another contribution of this study is the direct comparison of the radiation impedance theories of Miller and Pursey[1] and Oestreicher *et al.*[4,5] Elastic and viscoelastic example cases were considered that approximated seismologic, metallic, and soft biological tissue materials. Similar differences between the theories were observed in each of these cases. In particular, Oestreicher's theory predicted a larger resistance value in all three cases. Using a finite-element solution as the benchmark for the biological tissue case only, it would appear that Oestreicher's theory is the more accurate one of the two, though further study over a wider range of example case material conditions is warranted.

Future work of the authors includes extensions of these studies to a wider range of material conditions and to more geometrically complex situations that are relevant to biological and other applications, including orthotropy in the half-space medium (e.g., layering) and other types of vibration excitation sources (e.g., buried sources).

ACKNOWLEDGMENTS

The financial support of the Whitaker Foundation (Grant No. RG 97-113) and the National Institutes of Health (NCRR Grant No. 14250) is acknowledged. S.-H. Kim is acknowledged for assistance with the experimental study.

[1] T. J. Royston, H. A. Mansy, and R. H. Sandler, ''Excitation and propagation of surface waves on a viscoelastic half-space with application to medical diagnosis,'' J. Acoust. Soc. Am. **106**, 3678–3686 (1999).
[2] G. F. Miller and H. Pursey, ''The field and radiation impedance of mechanical radiators on the free surface of a semi-infinite isotropic solid,'' Proc. R. Soc. London, Ser. A **223**, 521–541 (1954).
[3] G. F. Miller and H. Pursey, ''On the partition of energy between elastic waves in a semi-infinite solid,'' Proc. R. Soc. London, Ser. A **224**, 55–69 (1955).
[4] H. L. Oestreicher, ''Field and impedance of an oscillating sphere in a viscoelastic medium with an application to biophysics,'' J. Acoust. Soc. Am. **23**, 707–714 (1951).
[5] H. E. von Gierke, H. L. Oestreicher, E. K. Franke, H. O. Parrack, and W. W. von Wittern, ''Physics of vibrations in living tissues,'' J. Appl. Physiol. **4**, 886–900 (1952).
[6] J. Lee, ''On the coupling and detection of motion between an artery with a localized lesion and its surrounding tissue,'' J. Biomech. **7**, 403–409 (1974).
[7] H.-D. Hong and M. D. Fox, ''No touch pulse measurement by optical interferometry,'' IEEE Trans. Biomed. Eng. **41**, 1096–1099 (1994).
[8] H.-D. Hong and M. D. Fox, ''Noninvasive detection of cardiovascular pulsations by optical Doppler techniques,'' J. Biomed. Opt. **2**, 382–390 (1997).
[9] R. O. Potts, D. A. Chrisman, and E. M. Buras, ''The dynamic mechanical properties of human skin *in vivo*,'' J. Biomech. **16**, 365–372 (1983).
[10] M. Mridha and S. Odman, ''Characterization of subcutaneous edema by mechanical impedance measurements,'' J. Invest. Dermatol. **85**, 575–578 (1985).
[11] M. Mridha, S. Odman, and P. A. Oberg, ''Mechanical pulse wave propagation in gel, normal and oedematous tissues,'' J. Biomech. **25**, 1213–1218 (1992).

[12] V. V. Kazakov and B. N. Klochkov, ''Low frequency mechanical properties of the soft tissue of the human arm,'' Biophysics (Engl. Transl.) **34**, 742–747 (1989).
[13] J. M. Pereira, J. M. Mansour, and B. R. Davis, ''Analysis of shear wave propagation in skin; application to an experimental procedure,'' J. Biomech. **23**, 745–751 (1990).
[14] J. M. Pereira, J. M. Mansour, and B. R. Davis, ''Dynamic measurement of the viscoelastic properties of skin,'' J. Biomech. **24**, 157–162 (1991).
[15] B. N. Klochkov and A. V. Sokolov, ''Waves in a layer of soft tissue overlying a hard-tissue half-space,'' Acoust. Phys. **40**, 244–248 (1994).
[16] A. Sarvazyan, ''Shear acoustic properties of soft biological tissues in medical diagnostics,'' J. Acoust. Soc. Am. **93**, 2329 (1993).
[17] S. Ganesan, C.-S. Man, and S. J. Lai-Fook, ''Generation and detection of lung stress waves from the chest surface,'' Respir. Physiol. **110**, 19–32 (1997).
[18] C. Man, M. Jahed, S. J. Lai-Fook, and P. K. Bhagat, ''Effect of pleural membrane on the propagation of Rayleigh-type surface waves in inflated lungs,'' J. Appl. Mech. **58**, 731–737 (1991).
[19] L. Gao, K. J. Parker, S. K. Alam, and R. M. Lerner, ''Sonoelasticity imaging: Theory and experimental verification,'' J. Acoust. Soc. Am. **97**, 3875–3886 (1995).
[20] S. Catheline, F. Wu, and M. Fink, ''A solution to diffraction biases in sonoelasticity: The acoustic impulse technique,'' J. Acoust. Soc. Am. **105**, 2941–2950 (1999).
[21] Y. Yamakoshi, J. Sato, and T. Sato, ''Ultrasonic imaging of internal vibration of soft tissue under forced vibration,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 45–53 (1990).
[22] T. Hall, M. Bilgen, M. F. Insana, and T. A. Krouskop, ''Phantom materials for elastography,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1355–1365 (1997).
[23] O. Amundsen, K. Gjaevenes, and T. Langeland, ''The acoustic impedance at the surface of the human body in relation to auscultation,'' Acustica **25**, 89–94 (1971).

Bioacoustic spatial perception by humans: A controlled laboratory measurement of spatial resolution without distal cues

Leslie Kay[a)]
P.O. Box 124, Russell 0255, New Zealand

(Received 27 March 2000; accepted for publication 27 October 2000)

The angular spatial resolution of a wide-angle air sonar using a continuous transmission frequency-modulated radiation, with the output coupled binaurally to the auditory system of a user, was measured under restrained controlled conditions. This was done to determine the effect of adding a narrow central field of view of 9 deg to a wide-angle sonar. The target objects were three equidistant vertical rods initially spaced apart by 10 deg. This was varied down to a spacing of 4 deg. Ten nonvisual subjects achieved an angular resolution of 6 deg. Four of these ten subjects continued learning to achieve an unexpected spatial resolution of 4 deg within the 9 deg central field. A mean error of approximately 1 deg in direction accuracy was achieved. It is inferred that the unique variations in the octave band ultrasonic echoes within the narrow field, and the invariance of the on-axis echo as one's head is turned, enables this angular resolution and accuracy to be achieved within the wide binaural field of view of 50 deg. This ability to resolve specula objects within a narrow angular resolution element of 9 deg is linked to the bat's ability to seemingly resolve object glints within a distal resolution element of less than 2 wavelengths. [DOI: 10.1121/1.1336138]

PACS numbers: 43.80.Ev, 43.80.Jz, 43.80.Ka [WA]

I. INTRODUCTION

Head-worn wide-angle field-of-view ultrasonic spatial sensors—or sonars with an audible output—have been developed so as to aid blind persons to better perceive their environment.[1] Blind persons are reported as being able to recognize landmarks, particularly trees and plants.[2,3] Through recent robotic studies with the sonar information being computer processed, confirmation of this remarkable recognition ability has been obtained.[4] Interest is focused on quantifying this ''spatial resolution''—necessary for object recognition—that blind persons seemingly enjoy. Imaging of their object space is carried out at a cortical level and evidence of spatial resolution depends entirely on behavioral responses—as with animals using sonar. It was reported that spatial resolution—as measured by the ability to count a number of rods on a table—was greatly increased when both angular and distal cues were available.[1] This was compared with the much poorer resolution when only angular spacing was a discrimination cue. In that experiment, free head scanning was allowed. A special feature of the latest high-resolution sensor (identified as the Trisensor) is the noticeable variation in echo structure when the target—a rod—is off the sensor axis. However, when a rod is viewed orthogonally on axis the echo is invariant and enables a user to recognize this specific condition which aids correct counting. If the scanning head does not seek this condition, the ability to discriminate between a rod and the space between rods is reduced. Subjects learning to use the sensor by learning to count rods, and thereby determine the spatial resolution, did not fully learn to seek the optimum condition for correct counting. This task may be likened to the familiar one of counting straight parallel white lines on a black background when careful control of the fovea is essential. Also, if a line is missed in the counting process, one has to start again because there is no specific cue with which to discriminate one line from another. Sensory system designers with experience in using the sensor were able to achieve better results through having learned to use the fine nuances in the echo signal signatures. The experiment reported here was designed to find out if novice subjects could quickly learn to use the same fine nuances in the signal signatures as those with long experience.

II. MATERIALS AND METHOD

The experimental sensory system comprised a headband-mounted cluster of ultrasonic sensor elements shown in Fig. 1. The headband was coupled to a box housing the electronics for generating the special continuous transmission frequency-modulated (CTFM) sonar range code in the ultrasonic echo signals being sensed. These ultrasonic signals were converted to the audible frequency range in a binaural form. (A detailed description of the sensor operation can be found in Ref. 1.) The specially matched cluster of sensor elements mounted in the headband produced the field-of-view plots shown in Fig. 2. These plots represent the variation in sound intensity as perceived binaurally by the user when turning one's head relative to the test target. The center channel is simply added to the left and right channels. The distance to a stationary object—one that produces a single specula reflection—is determined from the pitch of the audible tone it produces at the earphones, the frequency of which is proportional to distance (in this case, 2500 Hz rep-

[a)]Electronic mail: kaysonic@voyager.co.nz

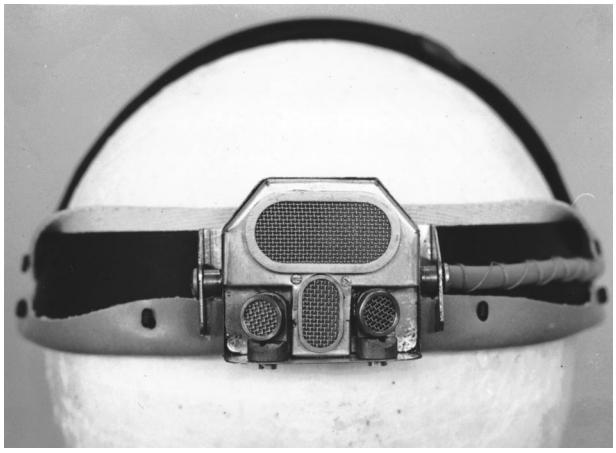


FIG. 1. This shows the head-mounted spatial sensor used in the experiment. The ultrasonic transducer cluster is specially designed to provide wide-angle spatial information about the surrounding object space. The small oval transducer radiates the CTFM-coded ultrasonic signal over an arc of approximately 50 deg at the mean frequency of 75 kHz. The remaining transducers act as receivers. Note the angle at which the circular transducers are mounted so as to provide binaural information from which direction of an object is determined. The large oval transducer receives signals from over a narrow central field of 9 deg.

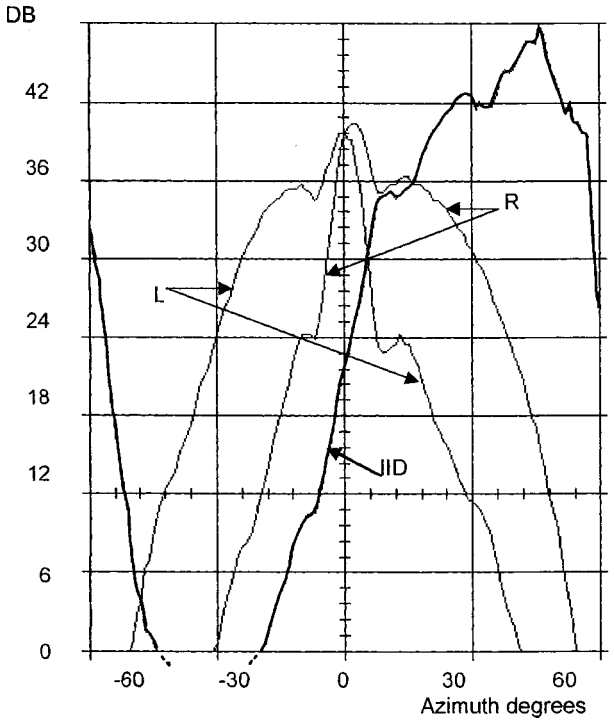


FIG. 2. Plot of the convolution of the radiation field and the two peripheral receiver fields, with the central field added, resulting from a reflection of the radiated signal from a vertical 40-mm-diameter plastic pipe acting as a standard target. The left and right plots are spaced in angle by 23 total degrees. The central field is 4 dB more sensitive on axis than the peripheral fields on their axes. Also plotted is the interaural intensity difference (IID) that produces the binaural lateral shift from which direction is determined. This plot models the degree of lateral shift sensed by a user. Note: the zero value of intensity difference occurs on the sensor axis when both the left and right signals are equal. The IID plot reverses when the sensitivity is very weak ($>\pm 40$ deg) and merges into the system noise. This reversal is not perceived. The dB scale applies, but not the *Y*-ordinate values. Accurate location of an object on the axis using the central field characteristics enables fixation on an object for discrimination and possible recognition.

resents 1 m). In general, under stationary conditions, echoes appear as almost continuous tone complexes that represent objects with discontinuities in both shape and texture. Periodic brief interruptions due to the sawtooth frequency modulation in the radiated signal produce a transient pulsation in the tone. This enables object discrimination through signal characterization. Object distance can then be related to a specific feature of an object, such as a reflecting surface, or from a corner. Using a digital frequency sweep, the distal resolution of this system is approximately 1.6 λ at the mean operating frequency.[5]

The center channel introduces interference within the narrow field. This causes the binaural direction cue in the azimuth plane to be magnified by a factor of 2 within the narrow central field.[1] The audible character of an object echo on the sensor axis is not modulated by the varying width of the fields during the frequency sweep. The echo signature on the axis then most closely represents the object shape as presented by the object pose.

The bioacoustic sensing process seems to be optimally carried out for object detection and recognition purposes, modeling correlation-receiver operation.[6] Here, signal processing in a large time-bandwidth CTFM sonar sensing system is that employing frequency-domain correlation. This is achieved in a biophysical acoustic sense through the combination of an electronic multiplier in each receiver channel followed by the biointegration process determined by the critical bandwidth of the human cochlea and the neural system. This bioacoustic process acts as an adaptive ''spectrum analyzer'' with the spatial information being tonotopically mapped along the frequency axis in the cortex so as to represent distance.[7] The binaural processing maps direction orthogonally with the frequency axis at the cortex. Head motion—which automatically includes any body motion—results in the receiver array continuously sampling the multiplicity of reflected and scattered waves from the insonified object space as these waves flow past the head. This head motion synchronously modifies the pitch of the echo tone complexes producing real-time varying sound ''signatures'' at the audio output for each tone complex, whose binaural fusion moves laterally to represent relative motion. The sound signatures include information on the frequency sensitive size, shape, and texture of objects, modulated by the constantly changing sensing characteristics of the three sensory fields of view as the ultrasonic frequency varies. The correlation process minimizes masking noise, as in an ideal receiver, so that maximum object information is made available for the discrimination and even the recognition processes. The object signatures have precise repeatability with direction of view, or the pose of the object. This motion invariance leads to a rapid learning process. The direction of different objects is perceived by the left or right lateral shift of each complex binaural fusion that each object produces. Reflected and scattered waves arriving on axis at the central receiver produce signals with time-varying frequency and amplitude. They contain, within the time period for a frequency sweep, the maximum information available for object categorization as the head scans a small arc—modeling in a simple way operation of the fovea of the eye. Off axis, it

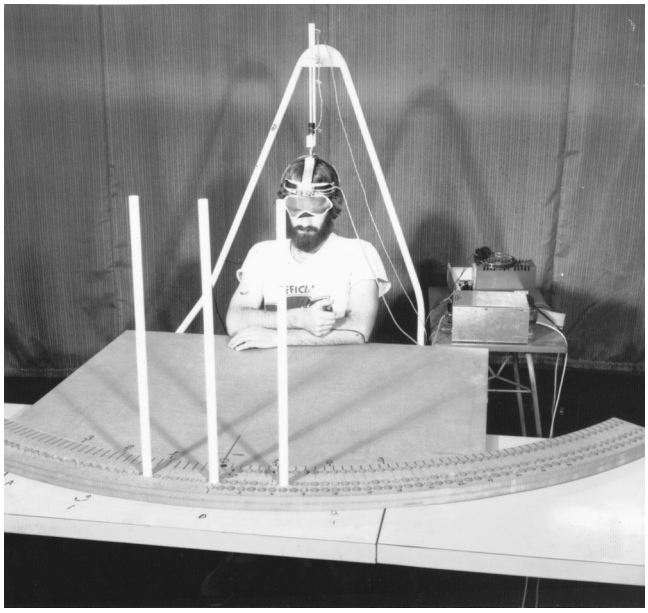


FIG. 3. The arrangement used for determining the spatial resolution of the sensor. A nonvisual subject is required to locate a rod with the sensor and press a button to record the head position on the X axis of an X–Y recorder when it is thought the sensor is looking directly at the desired rod. The three rods produce the same distance cue in the form of similar-pitch tones. The several characteristics of the tones vary as the head is turning. When a rod is on the sensor axis the tone is more pure, the binaural lateral shift is central, and the sound is loudest enabling accurate fixation to take place. (The arrangement was used for a variety of sensory experiments.)

is motion perception through binaural ''acoustic flow'' that is dominant, rather than absolute direction determination.

A. Experimental method

An arrangement was used as shown in Fig. 3 for testing a blind or blindfolded subject's ability to locate, bioacoustically, a vertical rod in the presence of two other masking rods at the same radial distance. The angular spacing of the rods could be increased or decreased till a threshold criterion of location and discrimination was established. This is the most difficult arrangement for viewing rods since there is no distal pitch cue with which to separate one rod from the other, as was shown to be helpful in counting rods.[1] The spacing of the rods was such that when the head was directed at the central rod all rods were in the wide field of view and each produced a sound having the same pitch but with a discriminable character. The subject's head was clamped in the sensor headband. This was attached to the direction-measuring sensor via a coupling. The head had freedom to turn only in azimuth so as to ensure that the sensor azimuth axis was orthogonal to the vertical rod. As the head turned the coupling that was attached to a potentiometer, a voltage was produced that was proportional to the angle of rotation of the head. A pen on an X–Y recorder than recorded the angular position of the head on the X axis. The subject was required to make a vertical mark of constant length on the Y axis of the recorder to represent the angular position of the rod. This was done by pressing a button when he or she ''looked'' at the rod and thought that the sensor axis was passing exactly through the rod in question; in other words, when the subject was looking at (fixating on) the rod orthogonally. The other two rods would be making a tonal sound of similar pitch that had amplitude modulation imposed by the off-axis frequency-sensitive field response of the sensor. If given special attention, these two rods could be perceived to be to the left or right of the rod being fixated and would have a different time varying tonal character.

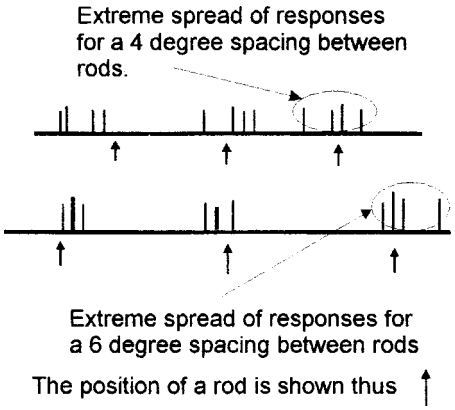


FIG. 4. This is an example of the records obtained from individual subjects of the position of their ''head-looking'' position when they thought that the sensor is directed exactly at the appropriate rod. Each short vertical mark simply indicates the angle of the sensor axis relative to the rod's angular position at the moment that the user presses a button. Most subjects achieved a spatial resolution of 6 deg. Four of these subjects achieved a resolution of 4 deg. That is well within the field of view of the central beam and this is of special interest.

III. THE EXPERIMENT

Three rods were placed in a slider with the central one being on the axis of the ahead-looking sensor position and the other two rods were initially spaced 10 deg to the left and right of the central rod. Care was taken to ensure that the rods were vertical so that they produced a maximum reflection along the horizontal sensor axis. Correct positioning of the headband on the subject's head was critical, as small angular errors were to be recorded. Fourteen nonvisual (one blind) subjects in turn were required to turn their heads to locate and look directly at each of three equally spaced distant rods in turn. When they were satisfied that they were looking directly at a rod, they were to make a mark on the recorder by pressing a button. This process was repeated three more times producing a record having three clusters, formed from four close-together marks as shown in Fig. 4. Each cluster represented a rod. An experienced experimenter monitored the sounds through headphones. During the learning process the experimenter, by auditory signal monitoring, was able physically to guide the subject's head so as to look directly at a rod. Through this learning process the subject could become assured that the correct axial sound was being obtained when making a response. With an angular spacing of 15 deg it was known from the earlier reported experiments[1] that 100%-correct discrimination between rods was highly likely. The degree to which a subject could accurately look at a rod was of interest. More importantly, the experiment was to determine the minimum angular spacing before angular resolution was found to be no longer possible.

Angular resolution in this experiment is defined as the angular spacing needed in order that there is seen to be clear

TABLE I. Trial analysis determined from the results of each subject. ''Mean error'' is the mean error of location of the rods in degrees. S.D. is the standard deviation of the errors of all the subjects.

Rod spacing		Left rod		Center rod		Right rod	
		Mean error	S.D.	Mean error	S.D.	Mean error	S.D.
10 deg	S=12	0.44	1.94	0.07	1.18	0.57	2.43
8 deg	S=14	0.65	1.34	0.36	1.41	0.53	1.87
6 deg	S=10	0.85	1.96	0.92	1.41	0.10	3.23
4 deg	S=4	1.13	1.55	0.39	0.38	0.30	0.70

separation between the three groups of responses representing four looks per rod, such as shown for 6-deg rod spacing in Fig. 4. We were concerned here with obtaining a measure of the limit of perceptual discrimination for the most difficult controlled sensory task when a reliable decision has to be made during an action. The concept of perceptual resolution is subjective, but is related to the standard deviation of the angular errors in the responses. After obtaining a very clear angular spacing resolution for a 10-deg spacing, using the first 12 subjects, the spacing was reduced to 8 deg and all 14 subjects then became involved. A period of training was again provided, then the test was repeated. This procedure was followed reducing the spacing in steps of 2 deg until all subjects felt that their responses were erratic and gave up further learning.

IV. RESULTS

The analysis of the many trials is presented in Table I, determined from the results of each subject—such as shown in Fig. 4 for one subject. The mean angular error in degrees was determined for each subject for each test. For each value of angular spacing, the overall mean error and the standard deviation of the errors was then determined. The mean error of location does not significantly exceed 1 deg. The separation of 6 deg was the closest separation for resolution to be obtained for the ten subjects who continued to this level of difficulty. Only four of these subjects were prepared to continue to attempt resolution for a 4-deg separation. The standard deviation of the errors shows that at 4-deg separation, resolution was possible. One of these four subjects was a 10-year-old blind girl who had, at the age of 6, learned very effective use of the low-resolution Canterbury Child's Aid.[8]

V. DISCUSSION

Dominant signal amplitude variation with angular change near the sensor axis due to the central beam of 9 deg width at the −6-dB level cannot alone explain the spatial resolution of 4 deg—as determined by the standard deviation of angular errors being less than half the angular spacing. At least one additional rod was producing an amplitude-modulated signal of similar pitch to that being viewed and this tends to mask the wanted signal. It is seen from Fig. 2 that angular movement of 4 deg about the central axis produces little more than a 1.0-dB loudness change. This is much less than the just noticeable difference (jnd) for amplitude change.[9] Two other factors played a more important part. The rod echo signal on axis was of a noticeably different characteristic—in the form of a 50-ms constant amplitude echo envelope—from the off-axis signal, which was tapered due to the frequency change. This noticeable change in characteristic takes place in real time with angular change. Also, the interaural intensity difference (IID) rate, in dB per degree, producing the binaural direction cue changed rapidly within the central field. Awareness of the change from the peripheral IID to the more accentuated central IID depends on how controlled the sensor axis is caused to approach the direction of the object from the left or right as the head is turned. When the fused binaural signal was centralized, when it correlated with the desired sound signature, and was also thought to be loudest, the subjects in this experiment were confident about making a judgment and pressed the button. As the spacing was reduced the novice subjects found the task mentally exhausting and all but four dropped out at the 6-deg spacing. The process became competitive.

This experiment is the first dynamically interactive attempt at measuring what might be called ''perceptual location'' accuracy and object spatial resolution, using a human with the head physically coupled to a recording instrument for monitoring azimuthal head turning, so as to test a high-resolution ultrasonic spatial sensor. This method is very different from the customary ''hand-pointing'' or ''hand-grasping'' methods used by independent psychological experimenters who failed to provide confirmatory information that the indicated direction of the hand pointing is closely aligned with the cognitive perception.[10,11] Here, it is believed that more meaningful results were obtained, the accuracy of ''measured head pointing'' being at least six times and up to 20 times superior to measured hand pointing that used no corrective feedback. It was evident from observing subjects that the operation of the biophysical system, with the central field of view in use, makes this system operable like the fovea of the eye for locating and discrimination between objects. This is enabled by the immediate feedback in the change of sound as the head is moved, providing good control of neck motion. The controlled feedback is quickly learned but skill increases with experience. With a little experience one can readily return to ''look'' at the base position, from which an object scan commenced, by matching with the memorized sounds and repeat the operation. Observation of school-age blind children using the sensor to locate a rod led to the realization that their auditory neural feedback system operated like a heavily damped control system, which briefly oscillates about the zero error and becomes steady.[12] This is considered to be equivalent to fixating on the rod. The importance, of course, is that the sound from the acoustic visual substitution process represents three-dimensional Euclidean space—as is normally seen by eyes.

This sensory performance is obtained only using information from earphone inputs to the two auditory channels and the associated effects of the head motion of a human. A 4-deg spatial resolution normally requires use of an aperture of 12+ wavelengths that produces a complex near field some 70 cm long. The field plots in Fig. 2 were taken at a distance of only 40 cm. The possible alternative use of FM echo pulse compression and time correlation for range coding, in place of the CTFM range coding and frequency correlation, could

not match the spectral analysis of the human auditory neural system, and multiple object perception would be destroyed. Object recognition would be lost in any kind of audible click that was produced, without this first being accompanied by a times-50 time-stretched signal.[13] The human memory of sound signatures representing object shape and their variation in real time with motion and object pose would also be lost. The Trisensor processing plays a very important part in the fixation process for object recognition. It suggests an explanation for the Saillant *et al.*[14] hypothesis that the bat enjoys better distal resolution than is predicted by the ideal receiver theory—as was demonstrated by the recognition of two glints within the receiver resolution limits. They refer to the bat having hyper-resolution. Kuc[15] examined this hyper-resolution using a USA penny as an object that was expected to produce multiple glints. He used a range-gated wideband ultrasonic pulse system, operating at an angle of 45 deg to the surface and at a distance of 15 cm from the coin. It was found possible for his biomimetric machine, after many integrations, to recognize the difference between the ''head'' and the ''memorial'' sides of the USA penny for a particular pose. This required extra high resolution because of the constant presence of the echo from the rim. Kuc's experiment was repeated using the Trisensor system but at 30 cm, with the audible output bandpass filtered to pass echoes from within a range bracket proportional to that used by Kuc. Three system designers experienced in listening to the Trisensor sounds found that the same task could be done in perceptual real time, as the coin was rotated lying on the center of a turntable. Only the central channel was used; the binaural facility was found to be unnecessary. With only the audible output there was a clearly distinctive sound complex arising from the two glints from (a) the closest memorial step—when it became orthogonal to the sensor axis—and (b) the near-side rounded coin rim, acting in combination. This pose sounded quite different from the sound produced from the head of the coin. The distance between the rear-rounded (not sharp as drawn by Kuc) coined edge of the penny and the first lip of the memorial steps—both producing a comparable echo signature—is about 1 mean wavelength. The application of ideal receiver theory employing an octave band radiation enables a distal resolution limit of 1.6 wavelengths.[5] In practice, the effective bandwidth of the Trisensor is significantly greater than 1 octave. Thus, two-glint object recognition actually occurred within these theoretical resolution limits, just as found by Saillant *et al.* The human has shown, through a simple validation of Kuc's experiment, that when presented with sensory spatial information resulting from a very wideband radiation and with a signal matched to the hearing process, he/she can do similar to the bat and better than information theory predicts. This seems to justify a claim of hyper-resolution bioacoustic spatial perception using the Trisensor.

Importantly, the ability to recognize orthogonality of a rod and fixate on it in the presence of similar masking sounds, as demonstrated in this experiment, depends on the same ability to recognize find sound nuances from two or more glints—as required for recognizing the pose of the memorial steps. The hyper-resolution—as it might be called—in angle is achieved through similar variation in signal nuances and discrimination as produced by the rotating coin. The coin glints were not perceived as separate sources—like points of light—but instead the sound change they produced *represented* sources spaced in distance at a mean radial distance *represented* by their combined envelope-modulated pitch. The envelope modulation in Trisensor for two close glints is similar to that shown in Ref. 14, Fig. 11(a). It is inferred that the concept of resolution when using an ideal receiver[14] is limiting when considering auditory neural processing and cognition for object recognition. Much consideration needs to be given to the exceptionally complex neuronal signal processing occurring in conjunction with a powerful memory and its part in the process of object recognition. The signal processing used by Kuc may have destroyed some important information.

The human auditory system cannot do better than allowed by the critical bandwidth theory.[9] This has variously been said to determine auditory frequency resolution. But, Do and Kay[16] showed that a frequency difference that enabled clear simultaneous resolution between two tones (like we observe between two points of light) required a frequency separation greater than 40% of the mean frequency. This is much wider in frequency separation than the critical bandwidth of the auditory system of man. Two-tone resolution (separation) was determined in their experiment by testing the ability of subjects to correctly say which of two simultaneous pulsating tones produced a left or right binaural lateral shift—that was when an interaural amplitude difference was applied to one of the two tones. It was found that the seeming poor spatial resolution was very greatly increased when the frequency of one of the two tones was changing and was dependent on the rate of change. If, instead, a subject is required to say if there are one or two tones present, then many different interpretations result but they do not determine resolution.

A simple experiment was tried using the slightly displaced edges of two sheets of writing paper that were orthogonal to the sensor axis to produce two strong echoes (glints). These were echoes that were the first derivative of acoustic impedance with respect to range and produces an 180-deg phase shift.[15] It was established that fine discrimination between two different double-glint situations (slightly different degrees of modulation) is possible using the Trisensor—even when the space between them was made so small as to approach zero. Exact alignment of the pages was possible by recognizing the particular state when a single specula reflection was obtained. This kind of discrimination may only be done nonvisually by humans when using such an octave band sensor as employed in this experiment, or seemingly by a bat using a wide-bandwidth sonar. The echoes in both systems have similar modulations imposed by the changing wavelength. The auditory systems and the cognitive processes of both the bat and the human are capable of making these very fine discriminations related to Euclidean space. The SCAT model of Saillant *et al.* may not be the only one, even though it does demonstrate feasibility.

VI. CONCLUSION

The formal experiment reported here, in conjunction with that earlier reported,[1] may be recognized primarily as part of an attempt to determine the value of the sensor to a blind person through a spatial resolution test, akin to an optical test of spatial resolution for a driving license. General usability may then be inferred—as in optical vision. It must be said, however, that sonocular perception (as this form of ultrasonic spatial perception is being called) is very much poorer in spatial information than optical sensor perception, and such inference may not be possible. The information made available is nevertheless very much richer than that provided by natural echolation, upon which blind persons are expected to rely. Few do so. Also, it is believed that some visually impaired persons would not have been able to achieve results similar to those obtained here using their defective vision, suggesting that persons with a severe visual impairment may find sonocular perception additionally useful. In particular, determining the distance to a specific object or environmental feature using a sensor of this type may overcome the distance perception problem the low-vision person experiences. Since natural echolation with its limitations remains the best means readily available for blind persons to detect and locate objects around them, it would seem worthwhile to study sonocular perception further, so as to gain a greater understanding of its capability under real dynamic daily living conditions. It is a more complex problem than the use of a hearing aid. The controlled static conditions in a laboratory, as mainly used by collaborators employing conventional psychological testing of Trisensor use, have not provided much useful information about a highly dynamic process that is learned *only* by blind persons over a significant period of time. No sighted person has learned a form of sonocular perception like those blind persons using the Sonicguide™ over long periods of time (25 years). It has also been found that translating the better spatial information from sonocular perception into teachable experiences of daily use is not obvious to teachers of the blind. The acoustic insights do not exist. It would seem, however, that sonocular perception with the degree of spatial resolution now available would enable blind persons to move about more naturally, and in a more relaxed way, then when relying entirely on using a cane and natural echolocation as their only means for environmental object perception. It is suggested that maybe a degree of natural self-learning from the dynamic use of the sensor through specially designed home exercises would lead to effective use. Further studies are needed to obtain some degree of correlation between sensor performance—as obtained here—with sensor usefulness to blind persons. Can a developing blind child naturally self-learn sonocular perception and develop cognition and action more quickly? Blind neonate macaques self-learned to use the Trisensor from birth.[17] This subject seems to offer a very challenging opportunity.

ACKNOWLEDGMENTS

I am pleased to acknowledge the very useful comments of the reviewers, which have enabled me to focus on the major issue addressed in this paper.

[1] L. Kay, ''Auditory perception of objects by blind persons, using a bioacoustic high resolution air sonar,'' J. Acoust. Soc. Am. **107**, 3266–3275 (2000).

[2] F. Gissoni, ''My cane is twenty feet long,'' *The New Outlook for the Blind* (American Foundation for the Blind, New York, 1966).

[3] P. W. Araisan, ''Evaluation of the binaural sensory aid,'' Research Bulletin No. 26, pp. 51–71 (American Foundation for the Blind, New York, 1977).

[4] P. J. McKerrow and N. L. Harper, ''Recognizing leafy plants with in-air sonar,'' Sens. Rev. **19**, 202–206 (1999).

[5] P. T. Gough, A. de Roos, and M. J. Cusdin, ''A continuous transmission FM sonar with one octave bandwidth and no blind time,'' Proc. IEE **131**, 270–274 (1984).

[6] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Part 1 (Wiley, New York, 1968), p. 249.

[7] C. K. Henkel, ''The auditory system,'' in *Fundamental Neurosciences*, edited by D. E. Haines (Churchill Livingstone, New York, 1997).

[8] E. R. Strelow, N. Kay, and L. Kay, ''Binaural sensory aid: Case studies of its use by two children,'' J. Visual Impair. Blindness **72**, 1–9 (1978).

[9] E. Zwicker and H. Fasti, *Psychoacoustics—Facts and Models* (Springer, London, 1990).

[10] D. H. Warren and E. R. Strelow, ''Training in the use of artificial displays,'' in *Electronic Spatial Sensing for the Blind*, edited by D. H. Warren and E. R. Strelow (Martinus Nijhoff, Boston, 1985).

[11] B. Bentzon, ''Spatial updating in the blind: Effects of training with air sonar versus sounding objects,'' Ph.D. thesis, Boston College, 1991.

[12] G. Hornby, L. Kay, M. Satherly, and N. Kay, ''Spatial awareness training of blind children using the Trisensor,'' in *Electronic Spatial Sensing for the Blind*, edited by D. H. Warren and E. R. Strelow (Martinus Nijhoff, Boston, 1985).

[13] W. W. L. Au and D. Martin, ''Insights into dolphin sonar discrimination capabilities from human listening experiments,'' J. Acoust. Soc. Am. **86**, 1662–1670 (1989).

[14] P. A. Saillant, J. A. Simmons, S. P. Dear, and T. A. McMullen, ''A computational model of echo processing and acoustic imaging in frequency-modulated echolating bats: The spectrogram correlation and transformation receiver,'' J. Acoust. Soc. Am. **94**, 2691–2712 (1993).

[15] R. Kuc, ''Biomimetric sonar recognizes objects using binaural information,'' J. Acoust. Soc. Am. **102**, 689–696 (1997).

[16] M. A. Do and L. Kay, ''Resolution in an artificially generated multiple object auditory sensations space using new auditory sensations,'' Acustica **36**, 9–15 (1976/77).

[17] E. R. Strelow, D. H. Warren, B. J. Sonnier, A. H. Riesen, L. Kay, and J. Sinton, ''Behavioral observations of sensory substitution in neonate macaques (*Macaca arctoides*),'' Behav. Neurosci. **101**, 738–741 (1987).

Coding of concurrent vocal signals by the auditory midbrain: Effects of stimulus level and depth of modulation

Deana A. Bodnar
Department of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853

Andrew H. Bass
Department of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853 and UC Bodega Marine Laboratory, Bodega Bay, California 94923

(Received 6 July 2000; accepted for publication 22 November 2000)

The segregation of concurrent vocal signals is an auditory processing task faced by all vocal species. To segregate concurrent signals, the auditory system must encode the spectral and temporal features of the fused waveforms such that at least one signal can be individually detected. In the plainfin midshipman fish (*Porichthys notatus*), the overlapping mate calls of neighboring males produce acoustic beats with amplitude and phase modulations at the difference frequencies (dF) between spectral components. Prior studies in midshipman have shown that midbrain neurons provide a combinatorial code of the temporal and spectral characteristics of beats via synchronization of spike bursts to dF and changes in spike rate and interspike intervals with changes in spectral composition. In the present study we examine the effects of changes in signal parameters of beats (overall intensity level and depth of modulation) on the spike train outputs of midbrain neurons. The observed changes in spike train parameters further support the hypothesis that midbrain neurons provide a combinatorial code of the spectral and temporal features of concurrent vocal signals. [DOI: 10.1121/1.1340646]

PACS numbers: 43.80.Lb, 43.64.Tk [WA]

I. INTRODUCTION

The detection and discrimination, i.e., segregation, of individual vocalizations that overlap in time are auditory processing tasks faced by all vocal species. Because of the nature of sound, the waveforms of simultaneous acoustic signals fuse into a single stimulus at a receiver's ear. When the overlapping signals are multi-harmonic, the resultant waveform contains acoustic beats with amplitude and phase modulations at the difference frequencies (dF) between the fundamental frequencies (F_0's) and harmonic components of the individual signals. To segregate the concurrent signals, the auditory system must encode the resultant waveform such that at least one of the signals can be independently detected. The elucidation of the neural mechanisms and computations utilized in the coding of concurrent vocal signals is essential to understanding the processes underlying their segregation.

A number of psychoacoustic studies in humans have shown that small differences in the F_0's of overlapping vowels (multi-harmonic signals) facilitates their segregation (Brokx and Nooteboom, 1982; Chalikia and Bregman, 1989). Modeling studies of the possible mechanisms underlying concurrent vowel segregation by humans suggest that the temporal encoding of the individual F_0's (Assman and Summerfield, 1990; Meddis and Hewitt, 1992; de Cheveigne, 1993, 1997) or the beat waveform itself (Culling and Darwin, 1994) can serve as the basis for signal segregation. Neurophysiological studies of mammalian auditory afferent encoding of concurrent vowels show that the F_0's are temporally coded (Palmer, 1990; Cariani and Delgutte, 1996a, b). Within the ventral cochlear nucleus of cats, the temporal coding of concurrent vowel F_0's is maintained by chopper-like neurons (Keilson *et al.*, 1997). However, the encoding of concurrent vocal signals by higher auditory centers in mammals and the neural computations utilized in their segregation remain unexplored.

In a vocal teleost, the plainfin midshipman (*Porichthys notatus*), overlapping vocalizations are a common occurrence in the course of its natural social reproductive behavior. During the breeding season, nesting male midshipman congregate in localized regions of the intertidal zone and produce long duration (>1 min), multi-harmonic signals (''hums'') during courtship of females (Brantley and Bass, 1994; Bass *et al.*, 1999). Hence, vocal signals often overlap. Behavioral two-choice phonotaxis experiments have demonstrated that midshipman can detect and localize a single humlike tone when presented with a choice between two concurrent tones that originate from separate underwater loudspeakers (McKibben and Bass, 1998). Hence, midshipman possess the neural mechanisms necessary to segregate concurrent signals.

In two recent studies, we examined the encoding of the temporal envelope and spectral composition of concurrent vocal signals by the midshipman auditory midbrain (Bodnar and Bass, 1997, 1999). Within the auditory midbrain, neurons synchronized their spike bursts to beat dFs and in most cases were tuned to a specific dF (Bodnar and Bass, 1997). In addition, the spike rates and interspike intervals (ISIs) of midbrain neurons were sensitive to the spectral composition of beats. In response to beats with the same dF, but that differed in a single frequency component, 60% of the neurons showed significant differences in their spike rates while 97% showed significant differences in their ISIs over a spe-

cific range of intervals (Bodnar and Bass, 1999). The differences in spike rate accounted for only 30% of the variance of the differences in ISIs. Together these data indicate that differences in synchronization to dF, spike rate, and ISIs of midshipman midbrain spike trains may all play a role in concurrent signal encoding, i.e., together they could provide a combinatorial code of the dF and spectral composition of concurrent signals.

Our previous studies of the coding of concurrent signals focused on assessing responses to signals with a standard set of parameters, namely beats that were 12 dB above threshold and with 100% depth of modulation (equal intensity primary tones). However, in an animal's natural habitat, both the overall intensity of beats and the relative intensities of the beat's individual components will vary with the receiver's distance from signalers and their individual sound outputs. Thus understanding how changes in signal parameters are reflected in changes of spike train coding parameters is necessary to derive the exact nature of the neural code of concurrent signals and how it may be utilized in beat detection and signal segregation. In this study, we present beat and AM stimuli at different intensity levels and depths of modulation and examine the changes in spike rate, synchronization to dF, and ISI spectral sensitivity of auditory midbrain neurons. AM stimuli were included because midbrain units respond differentially to the modulation of AM and beat signals (Bodnar and Bass, 1997); midshipman also generate brief duration AM-like signals called ''grunts'' in agonistic encounters (Brantley and Bass, 1994; Bass *et al.*, 1999). AM stimuli also provide an interesting comparison to beats because while both signals are similar in their amplitude modulations, they differ in their spectral and fine temporal properties such as phase modulations. We assessed the effects of changes in stimulus parameters on both the responses of individual neurons and across the population of animals tested. Our results show that changes in any one signal parameter produce changes in more than one spike train parameter and that changes in signal power alone do not provide an adequate explanation for the observed data. Instead, it appears that changes in the constellation of multiple spike train attributes specify a particular set of signal parameters. This lends further support to the hypothesis that midbrain neurons provide a combinatorial code of the temporal and spectral features of acoustic signals.

II. METHODS

A. Neurophysiological recordings and acoustic stimuli

Animal preparations and recording methods are identical to those reported in Bodnar and Bass (1997, 1999). Single unit recordings were obtained from the auditory midbrain of male ($n=24$) and female ($n=12$) midshipman fish. For surgery, animals were anaesthetized by immersion in 0.2% ethyl *p*-amino benzoate (Sigma Chemical, Inc., St. Louis, MO) in seawater from their housing unit. The midbrain was exposed, and a plastic dam was attached to the skin surrounding the opening which allowed for submersion of the fish below the water surface. During recording, pancuronium bromide (0.5 mg/kg) was used for immobilization and fentanyl (1 mg/kg) for analgesia. The animal was respired via constant water flow across its gills. Acoustic signals were synthesized using custom software (CASSIE designed by J. Vrieslander at Cornell Univ.) and delivered through an UW30 underwater speaker positioned beneath the fish in a 32-cm-diameter tank (design after Lu and Fay, 1993). The frequency response of the speaker was measured with a Bruel and Kajer 8103 mini-hydrophone, and sound pressure was equalized using CASSIE software. Hydrophone recordings of acoustic stimuli verified that reflections from the tank walls and water surface did not alter the sound pressure waveform of the signals. The research reported here was performed within the guidelines of the Cornell University Animal Care and Use Committee and the National Institutes of Health.

Beat stimuli were composed of two tones ($F1$ and $F2$) near the F_0's of natural hums (see Bass *et al.*, 1999). $F1$ was held constant at 90 or 100 Hz which is close to the characteristic frequency of most auditory midbrain units; $F2$ varied from $F1$ up to ± 10 Hz in 2-Hz increments and in some cases additional dFs were recorded out to ± 15 Hz, spanning the range of characteristic frequencies (Bodnar and Bass, 1997). The order of presentation of beat stimuli was either with increasing dF, decreasing dF, or random dF. Stimuli were 1 s in duration and data were collected for ten repetitions at each dF.

To assess the effects of changes in stimulus levels, we typically recorded a neuron's spike train responses to beats with 100% depth of modulation at 6 dB and 12 dB above threshold. In some cases, responses were measured at 12 dB and 18 dB above threshold, while in others, responses to all three stimulus levels were recorded. To test the effects of changes in depth of modulation, responses to beats with 50% and 100% depth of modulation were recorded. In some neurons, responses to 75% depth of modulation were also recorded. We did not compensate for decreases in the overall stimulus level with decreased depth of modulation.

Responses to amplitude modulated (AM) signals with a carrier frequency of 90 Hz and modulation frequencies (modF) of 2, 4, 6, 8, and 10 Hz were recorded at depths of modulation of 50%, 80%, and 100%. A depth of 80% AM was used because this depth was used in previous comparisons of beat and AM signals (Bodnar and Bass, 1997).

B. Spike train analysis

1. Spike rate and vector strength of synchronization to beat dF (VS_{dF})

The mean spike rate was calculated over ten repetitions at each dF or modF. The analysis of VS_{dF} follows the method described previously (Bodnar and Bass, 1997). Briefly, VS_{dF} was calculated for each dF by determining the vector strength of spike times for the cycle period equal to dF. In the case of AM signals, vector strength was computed for the modF, VS_{modF}. All other analyses of spike train responses to AM signals described below follow those of beats, using modF and VS_{modF} rather than dF and VS_{dF}.

Traditionally, a single vector strength value is calculated over all the spike train data. However, to quantify and compare beat responses between different stimuli, we computed VS_{dF} over 1-s stimuli and statistical measures over ten repetitions. Using this method, Bodnar and Bass (1997) showed that midbrain neurons exhibit selectivity at one or more dFs.

The Rayleigh *Z*-value tests whether synchronization to the dF period signal is significant (Batschelet, 1981). Here, we computed the *Z*-value using the mean VS_{dF} and mean spike rates:

$$Z = (VS_{dF})^2 * \text{mean spike rate}$$

$$p < 0.05 \quad \text{for} \quad Z > 3. \tag{1}$$

2. Gain of the modulation transfer function (MTF) and the magnitude of the Fourier component (FC)

We also computed the gain of the modulation transfer function (MTF) and the magnitude of the Fourier component (FC) for each neuron. These values were computed following Rees and Palmer (1989):

$$\text{Gain} = 20 * \log \frac{\%\ \text{modulation of the period histogram}}{\%\ \text{modulation of the waveform}}$$

$$\text{where \% modulation of the period histogram} = 200 * VS_{dF}, \tag{2}$$

$$\text{FC} = 2 * \text{mean spike rate} * VS_{dF}. \tag{3}$$

The gain of the MTF indicates how much the temporal modulation of a neuron's spike train amplifies or attenuates the actual modulation in the stimulus waveform. In the case of stimuli with 100% modulation, this value directly reflects VS_{dF}. The FC gives a measure of the combined effects of the changes in spike rate and VS_{dF} at a given dF with changes in stimulus parameters.

3. Statistical comparisons of spike rate, VS_{dF}, gain, and FC data

The effects of changes in the intensity or depth of modulation on spike train parameter responses were assessed for both individual neurons and across the population of animals tested. We assessed the effects of changes in stimulus level and depth of modulation on the spike rate and VS_{dF} at each dF between ±10 Hz for individual neuron responses using a two-way factorial ANOVA. The ANOVA was computed across responses to individual stimulus repetitions with dF and the stimulus parameter (intensity or depth of modulation) as the two factors. This analysis was not performed for the gain and FC as determination of these values was based on the mean spike rate and VS_{dF} values; hence, statistical comparisons could not be made.

To determine the effects of stimulus parameters on the spike rate, VS_{dF} and FC across the population, we computed the relative change in each spike train parameter for an increase in intensity (e.g., 12 dB/6 dB) or depth of modulation (e.g., 100%/50%) at each dF. We determined a mean value for the relative change in any one spike train parameter in two ways: (1) a mean was computed across all dFs for the

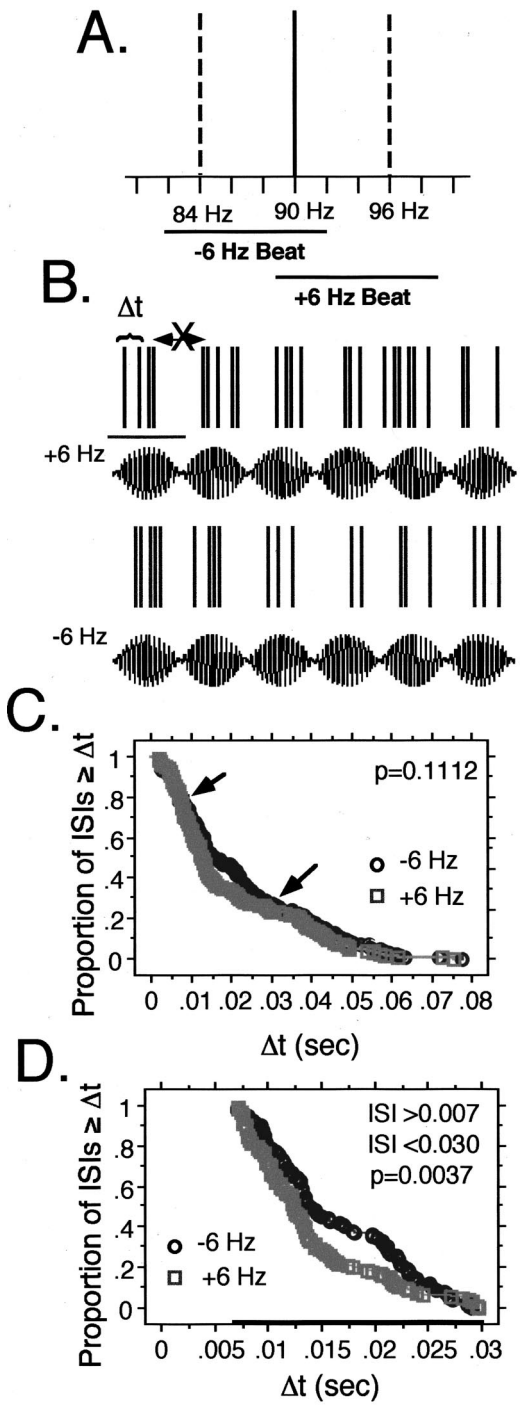


FIG. 1. Spike train analysis of ISI spectral sensitivity (modified from Bodnar and Bass, 1999). (A) A schematic of an example of a power spectrum of negative and positive beat dFs (±6 Hz) used in this study. For most stimuli, one component was always held constant at 90 or 100 Hz, while the other component was varied from −10 Hz to +10 Hz in 2-Hz increments. (B) A schematic of spike rate and interspike interval (ISI) measurements used in this study. ISIs (Δt) were measured within a beat cycle and not between beats (designated by *X* between the first and second beat cycles). ISIs were compared between negative and positive dF beats so that the beat stimuli differed in only one spectral component. (C) A plot of the inverse cumulative distribution functions of a unit which does not show significant differences in its mean ISIs for a ±6-Hz beat over the entire range of intervals. The arrows mark the beginning and end of where the two plots show a divergence. (D) A plot of the inverse cumulative distribution functions of the same unit over the intervals marked by the arrows in (C). Within this limited range of intervals (ISI), there is a significant difference in the mean ISIs of the ispike train responses to a ±6-Hz beat. In all cases, a unit was considered to show ISI spectral sensitivity only if ⩾50% of its ISIs fell within the spectrally sensitive range.

TABLE I. Summary of the changes in midbrain neuron spike train parameters with changes in stimulus intensity of beats. Mean values and statistical tests are on individual animals, i.e., data for multiple neurons in a single animal are averaged together.

	N (neurons/animals)	VS_{dF}	Spike rate	Gain	FC
12 dB/ 6 dB					
all dFs	34/19	1.02±0.15	1.42±0.34[b]	0.05±1.11	1.44±0.33[b]
dFs with $Z>3$ at 6 dB	29/17	0.94±0.07[a]	1.40±0.31[b]	−0.60±0.69[a]	1.32±0.26[b]
18 dB/12 dB					
all dFs	12/10	0.92±0.09	1.26±0.18[b]	−0.95±0.92	1.17±0.16[a]
dFs with $Z>3$ at 12 dB	12/10	0.93±0.09	1.26±0.19[b]	−0.94±0.90	1.17±0.16[a]

[a]$p<0.05$ one sample sign test.
[b]$p<0.01$ one sample sign test.

entire population of single units sampled; and (2) a mean value for relative change was determined for only those dFs with a significant VS_{dF} at either the lower intensity or depth of modulation. We carried out the latter computation because our previous studies have shown that VS_{dF} is the most salient parameter for coding beat dFs (Bodnar and Bass, 1997, 1999). Hence, we wanted to determine the effects of stimulus parameter changes on significant VS_{dF} coding.

Because more than one unit was often recorded from the same animal, we averaged the values of units within the same animal in order to maintain strict independence of the data for statistical tests. A one sample sign test (nonparametric one sample *t*-test) against a distribution with mean equal to one was used to test for significant differences ($p<0.05$). If there was no change in the mean relative spike train parameter, then the distribution should have a mean that is not significantly different from one (null hypothesis). This approach was the most conservative in that statistical power was reduced (lower sample size of animals rather than neurons and nonparametric test).

Because gain measurements inherently denote a relative change (i.e., +6 dB always indicates twice the value), absolute differences in the gain values were measured and averaged across dFs and individual animals. In this case, a one sample sign test against a hypothesized distribution with a mean of zero was used to test significance ($p<0.05$).

In our experiments, the number of stimulus parameters for each test varied between neurons. In addition, there is not a nonparametric repeated measures ANOVA. Hence, we compared cumulative data across the population separately for 6 dB vs 12 dB and 12 dB vs 18 dB, as well as 50% vs 100% and 75% vs 100%.

4. ISI spectral sensitivity

The sensitivity of spike train ISIs to the spectral composition of beat signals (ISI spectral sensitivity) was assessed following methods described in detail in Bodnar and Bass (1999); we outline this relatively new procedure again in Fig. 1. Comparisons were made between the first order ISIs within beat periods of spike train responses to ±dF stimuli [Figs. 1(A) and (B)]. Because the dF of the two signals is the same, differences in a neuron's spike train responses will reflect differences in the spectral composition of the two beats of opposite sign. To compare first-order ISI probability distributions, we examined the inverse cumulative distribution functions, also known as survival functions (using StatView 4.5), of the ISIs of spike train responses to positive and negative dF beats [Fig. 1(C)]. A cumulative probability distribution (F) shows the probability of event occurrences less than or equal to a designated value [e.g., Δt, Fig. 1(B)]; the inverse cumulative distribution ($1-F$) shows the probability of event occurrences greater than or equal to the designated value. These plots were used because they facilitate identification of differences in the ISI distributions. The points of divergence in the ISIs is easily observed in the inverse cumulative ISI distributions and hence, the range over which the ISIs appear to exhibit sensitivity to differences in the spectral composition of a beat can be directly assessed [Fig. 1(C)]. A unit is designated as exhibiting ISI

TABLE II. Summary of the changes in the mean relative change in midbrain neuron spike train parameters with changes in the depth of modulation of beats. Mean values and statistical tests are on individual animals, i.e., data for multiple neurons in a single animal are averaged together.

	N (neurons/animals)	VS_{dF}	Spike rate	Gain	FC
100%/50%					
all dFs	39/20	1.25±0.10[b]	1.10±0.17	−4.33±0.58[b]	1.35±0.20[b]
dFs with $Z>3$ at 50%	33/18	1.14±0.10[b]	1.09±0.25	−4.83±0.97[b]	1.22±0.22[a]
100%/75%					
all dFs	18/10	1.08±0.07[a]	1.12±0.15[a]	−1.98±1.00[a]	1.23±0.18[a]
dFs with $Z>3$ at 75%	15/9	1.05±0.07	1.12±0.13	−2.26±0.80[a]	1.16±0.12[a]

[a]$p<0.05$ one sample sign test.
[b]$p<0.01$ one sample sign test.

TABLE III. Summary of the changes in the mean relative change in midbrain neuron spike train parameters with changes in the depth of modulation of AM signals. Mean values and statistical tests are on individual animals, i.e., data for multiple neurons in a single animal are averaged together.

	N (neurons/animals)	VS_{modF}	Spike rate	Gain	FC
100%/50%					
all modFs	20/12	1.36 ± 0.14^{b}	1.12 ± 0.09^{b}	-3.45 ± 0.93^{b}	1.51 ± 0.18^{b}
modFs $w/Z>3$ at 50%	17/10	1.26 ± 0.12^{b}	1.12 ± 0.07^{a}	-4.23 ± 1.01^{b}	1.39 ± 0.22^{a}
100%/80%					
all modFs	27/13	1.07 ± 0.06^{a}	1.06 ± 0.12	-1.26 ± 0.79^{b}	1.17 ± 0.17^{b}
modFs $w/Z>3$ at 80%	27/13	1.05 ± 0.05	1.07 ± 0.12	-1.35 ± 0.77^{b}	1.15 ± 0.16^{a}

[a] $p<0.05$ one sample sign test.
[b] $p<0.01$ one sample sign test.

spectral sensitivity if it shows significant differences ($p<0.05$, Mann–Whitney U-test) over at least 50% of its total ISIs in spike output responses to ±dF beats [e.g., Fig. 1(D)].

To compare ISI spectral sensitivity for different parameter settings, its presence or absence was scored for each intensity or depth of modulation level. We designated ISI spectral sensitivity as being maintained, if ISI spectral sensitivity (i.e., significant differences in ISIs) was present at both the lower and the higher parameter setting. ISI spectral sensitivity was considered lost, if ISI spectral sensitivity was present at the lower parameter value and absent (i.e., no significant differences in ISIs) at the higher parameter value. Finally, we designated that there was a gain in ISI spectral sensitivity, if ISI spectral sensitivity was absent at the lower parameter value, but appeared for the higher parameter setting.

A single unit may exhibit ISI spectral sensitivity for some dFs and not others, as well as over different ranges for different dFs. Thus there was no obvious way to combine data for different dFs for a single neuron or individual animals. Hence, cumulative data are presented in terms of percent cases which maintained, lost, or gained spectral sensitivity and no statistical comparisons were made: each case represents a single dF. Because ISI spectral sensitivity is determined by the comparison of ±dF signals, this measure cannot be obtained for AM signals as only one stimulus exists for each modulation frequency.

III. RESULTS

Complete data sets were recorded in 42 neurons for intensity comparisons, 41 neurons for depth of modulation of beat comparisons, and 27 neurons for depth of modulation of AM stimuli in a total of 36 animals. Thresholds ranged from 85 to 120 dB *re*: 1 μPa. A data set was considered complete if recordings were obtained for a minimum of the same eight dFs or modFs for at least two levels of intensity or depth of modulation. Hence, the number of units for parameter comparisons was limited by the time required to hold a stable recording and obtain complete data sets. For units with exceptionally stable isolation, data were obtained at an additional intensity (18 dB) or depth of modulation (75%).

As mentioned in the Introduction, a primary focus of our previous studies has been on the temporal coding of beat dF or AM modF via synchronization, i.e., VS_{dF} or VS_{modF}, respectively. We compared the effects of intensity and depth of modulation both for all dFs and for only dFs which showed significant synchronization at the lower intensity or depth of modulation; we also compared the effects of changes in AM depth of modulation on the encoding of all modFs and for only modFs which showed significant synchronization at the lower depth of modulation. The statistics from both analyses of the population data are summarized in Tables I–III. We also show representative responses of individual neurons as well as the frequency distributions of the population data for all animals.

A. Effects of changes in overall intensity of beats

A total of 42 neurons were tested; 30 neurons for changes in intensity from 6 dB to 12 dB, 4 neurons for 6 dB, 12 dB, and 18 dB, and 8 neurons for changes in intensity from 12 dB to 18 dB above threshold. Examples of the effects of changes in overall stimulus intensity on the spike train parameters of individual neurons are shown in Fig. 2. The observed changes in VS_{dF} of two representative midbrain neurons at different dFs with a 6-dB increase in intensity are shown in Fig. 2(A). The neuron on the left showed a significant decrease in VS_{dF} with increased intensity (two-way ANOVA $p<0.0001$, df=1, $F=185.1$), while the unit on the right showed no significant change across all dFs. The majority of units (62%, $n=26/42$) showed significant changes in the VS_{dF} with changes in intensity level ($p<0.05$ ANOVA effect of intensity). Changes in intensity produced changes in the dF tuning in only 17% of the neurons, i.e., the dFs that showed the maximum VS_{dF} shifted ($n=7/42$, ANOVA effect of dF* intensity, $p<0.05$). Thus dF tuning remained relatively stable with changes in this parameter.

The effects on the spike rate responses for the same neurons in Fig. 2(A) are shown in Fig. 2(B). In both of these neurons, increases in intensity produced significant increases in spike rate [two-way factorial ANOVA, left panel: $p<0.0001$, degrees of freedom (df=1, $F=234.2$; right panel: $p<0.0001$, df=1, $F=26.2$]. At the single unit level, increases in intensity produced significant changes in spike rate in 88% ($n=37/42$) of the neurons tested (ANOVA effect of intensity $p<0.05$). There was a significant interaction between dF and intensity in 55% ($n=23/42$) of the neurons (two-way ANOVA dF* intensity $p<0.05$) indicating that

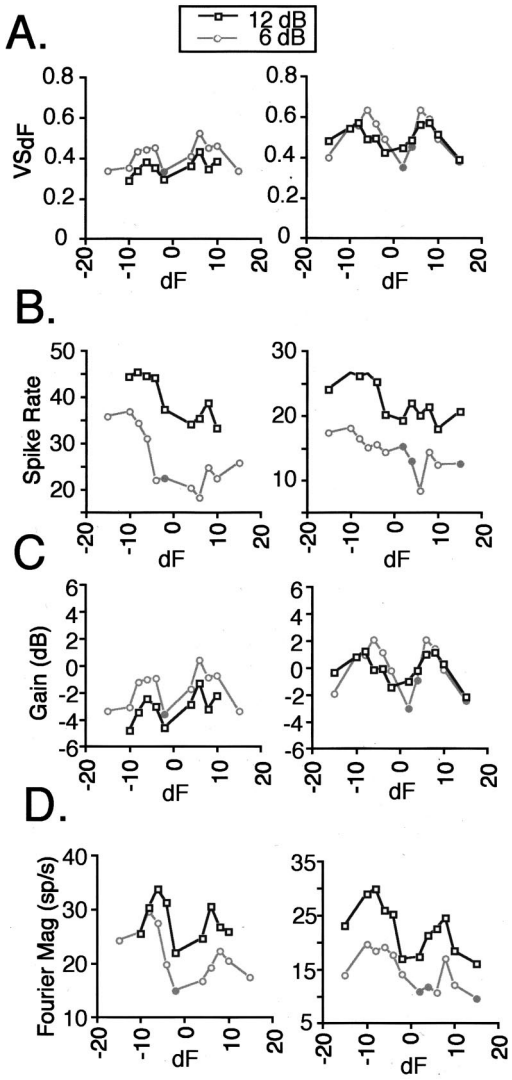


FIG. 2. Effects of changes in stimulus intensity on spike train parameters for individual neurons in response to beat stimuli. The graphs show plots of the designated parameter versus beat dF for two representative neurons at 6 dB (circles and gray lines) and 12 dB (squares and black lines) above threshold. dFs at which VS_{dF} was **significant** have open symbols while those with **insignificant** synchronization have filled symbols. (A) VS_{dF}; (B) Spike rate (SR); (C) Modulation gain; (D) Magnitude of the Fourier Component (FC).

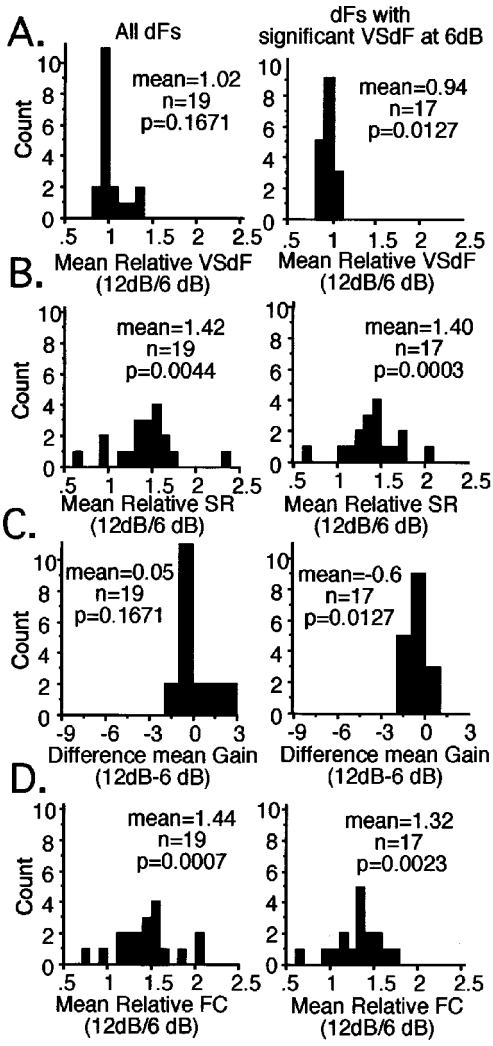


FIG. 3. Effects of changes in stimulus intensity on spike train parameters across the population. The plots on the left show the frequency distribution of the mean relative change (12 dB/6 dB) of the population for the designated parameter across all dFs. The plots on the right show the frequency distribution of the mean relative change of the population for designated parameter across dFs at which VS_{dF} was significant at 6 dB. *P*-values are for a one-sample sign test. (A) VS_{dF}; (B) Spike rate (SR); (C) Modulation gain; (D) Magnitude of the Fourier Component (FC).

the spike rate responses to beat signals were dependent on both stimulus intensity and beat dF for many neurons.

The changes in the gain of the MTF for the same neurons from Fig. 2(A) are shown in Fig. 2(C). Because the depth of modulation in these experiments was 100%, changes in the gain of the MTFs reflected changes in VS_{dF}. The changes in the magnitude of the FC of the same two neurons are shown in Fig. 2(D); in both cases, the FC increased with increased stimulus level.

A summary of the effects of all changes in stimulus intensity on the population spike rate, VS_{dF}, MTF gain, and FC are shown in Table I. The frequency distributions of changes from 6 dB to 12 dB above threshold are shown in the histograms of Fig. 3. The histograms on the left show the mean relative changes across all dFs in the entire population, while those on the right show the mean changes only for dF values that showed significant synchronization at 6 dB. At 6 dB above threshold, 85% ($n = 29/34$) of the neurons showed significant synchronization for at least one dF.

The distribution of changes in mean relative VS_{dF} across the population for increases in intensity from 6 dB to 12 dB are shown in Fig. 3(A). As a population, there was no significant change in VS_{dF} across all dFs with an increase in stimulus level. However, there was a small but significant decrease in the mean relative VS_{dF} at dFs that showed significant synchronization at 6 dB above threshold. There were comparable, although nonsignificant, decreases in VS_{dF} for increases from 12 dB to 18 dB in intensity. Nevertheless, the trend was consistent with what was observed for increases from 6 dB to 12 dB.

The distributions of the effects of increases in intensity from 6 dB to 12 dB on spike rate across the population are shown in Fig. 3(B). The mean change in relative spike rates was significant both across all dFs and dFs with significant VS_{dF} at 6 dB. Spike rate changes were also significant for

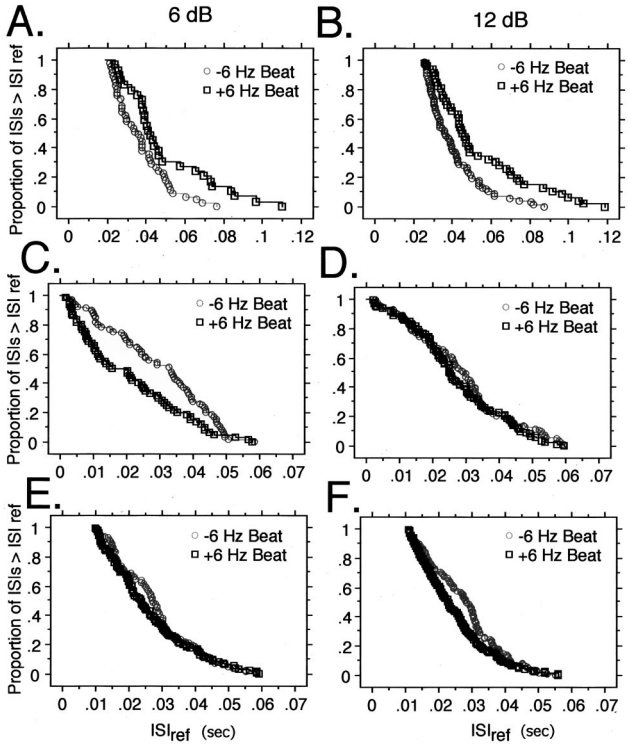


FIG. 4. Examples of comparisons of ISI spectral sensitivity between ±dF beats at different stimulus levels. The plots in each panel show the inverse cumulative ISI distributions for ±6-Hz beat stimuli (84+90 Hz, 90+96 Hz) across a specific range of intervals. (A),(B) An example of a case in which the unit exhibits ISI spectral sensitivity at 6 dB above threshold (A), and it is maintained when the stimulus level is increased to 12 dB above threshold (B). (C),(D) An example of a case in which a unit that showed ISI spectral sensitivity at 6 dB (C), and lost it at 12 dB (D) above threshold. (E),(F) An example of a case in which a unit does not show significant ISI spectral sensitivity at 6-dB (E) above threshold, but gains it at 12-dB above threshold (F).

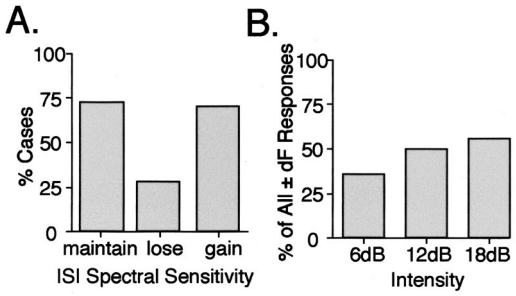


FIG. 5. Cumulative data for comparisons of ISI spectral sensitivity between ±dF beats. (A) The percentage of cases of ISI spectral sensitivity that are maintained, lost or gained with a 6 dB increase in overall stimulus level. (B) The percentage of all comparisons of ±dF beats that show ISI spectral sensitivity at each of the intensity levels above threshold.

increases in intensity from 12 dB to 18 dB (see Table I).

The distribution of changes in mean difference in the gain of the MTFs is shown in Fig. 3(C). As noted in the methods section, gain values already denote a relative measure (e.g., a 6 dB increase in gain always indicates a doubling in gain). Hence, an assessment of changes in gain across the population were measured in terms of absolute rather than relative gain. An increase in stimulus level did not produce any significant change in MTF gains across the population at all dFs. At dFs with significant synchronization at 6 dB, the decrease in gain was however significant. These changes directly reflected those observed for the changes in the mean VS_{dF} because both stimuli had 100% depth of modulation.

The population data for the mean relative change in the magnitude of the FC is shown in Fig. 3(D). As a population at all dFs, midbrain neurons showed a significant increase in the magnitude of their FC. Similarly, at dFs with significant VS_{dF} at 6 dB, there was a significant increase in the magnitude of the FC with an increase in intensity. This indicates that the increases in spike rate with increased stimulus level outweighed the decreases in VS_{dF}. A similar effect was observed for increases in intensity from 12 dB to 18 dB above threshold. Hence, spike rate appears to be the more salient feature for encoding changes in the stimulus intensity of beats.

To assess the effects of increased intensity on the ISI spectral sensitivity of neurons, we measured ISI spectral sensitivity for ±dFs at two different intensity levels. At 6 dB above threshold, 63 cases of ISI spectral sensitivity were observed in a possible 168 possible ±dF pairs in 34 neurons tested at 6 dB and 12 dB. For increases in intensity from 6 dB to 12 dB above threshold ISI spectral sensitivity was maintained in 73% ($n=46/63$) of the cases of ISI sensitivity observed at 6 dB [e.g., Figs. 4(A), (B); 5(A)], lost in 27% ($n=17/63$) of the cases [e.g., Figs. 4(C), (D), 5(A)], but gained in an additional 71% ($n=45/63$) of the cases [e.g., Figs. 4(E), (F); 5(A)]. For increases in intensity from 12 dB to 18 dB above threshold 26 cases of ISI spectral sensitivity were observed in a possible 58±dF pairs in 12 neurons at 12 dB. The results were similar to those observed for increases from 6 dB to 12 dB; 69% ($n=18/26$) maintained, 31% ($n=8/26$) lost, and 50% ($n=13/26$) gained ISI spectral sensitivity. Comparisons of the occurrence of ISI spectral sensitivity at each intensity level are shown in Fig. 5(B); ISI spectral sensitivity was observed in 37% ($n=63/168$) of ±dF pairs tested at 6 dB, 51% ($n=107/207$) at 12 dB (includes all 12 dB data), and 57% ($n=33/58$) at 18 dB above threshold.

B. Effects of changes in depth of modulation of beats

Changes in the depth of modulation of beat waveforms are produced by changing the relative amplitudes of the primary tones. When the relative amplitude of one signal is half that of the other (6 dB lower in amplitude), the depth of modulation is 50%. When the relative amplitudes are equal, the depth of modulation is 100%. At 75% depth of modulation, one signal is 25% (3 dB) lower in amplitude. A total of 41 neurons were tested; 23 neurons at 50% and 100%, 16 neurons at 50%, 75% and 100%, and 2 neurons at 75% and 100% depth of modulation.

Examples of the effects of increases in the depth of modulation on the VS_{dF} of two neurons are shown in Fig. 6(A). Both units showed significant increases in their VS_{dF} with increased depth of modulation (left panel: $p<0.0001$, df=1, $F=43.7$; right panel: $p<0.0001$, df=2, $F=10.1$). The changes in VS_{dF} with changes in depth of modulation

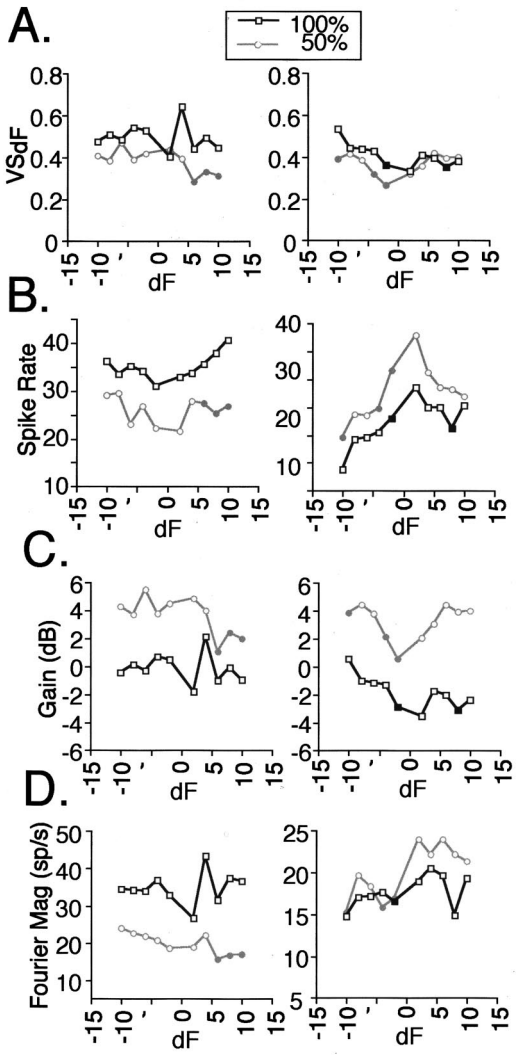


FIG. 6. Effects of changes in depth of modulation on spike train parameters for individual neurons in response to beat stimuli. The graphs show plots of the designated parameter versus beat dF for two representative neurons at 50% (circles and gray lines) and 100% (squares and black lines) depth of modulation. dFs at which VS_{dF} was **significant** have open symbols while those with **insignificant** synchronization have filled symbols. (A) VS_{dF}; (B) Spike rate (SR); (C) Modulation gain; (D) Magnitude of the Fourier Component (FC).

were significant in 90% ($n=37/41$) of the units (ANOVA effect of dF, $p<0.05$). At 50% depth of modulation, all but six neurons (16%) showed significant VS_{dF} for at least one dF. In addition, dF tuning of a neuron changed significantly in 27% of the neurons tested, i.e., the best dF changed ($p<0.05$ ANOVA effect of dF* depth, $n=11/41$ neurons). Thus changes in dF tuning were more dramatic compared to increasing stimulus intensity (only 18%).

Figure 6(B) shows the changes in spike rate with increased depth of modulation for the neurons in Fig. 6(A). There was a wide range in variation of observed changes in spike rate with some neurons showing a significant increase in spike rate (e.g., left panel: $p<0.0001$, df=1, $F=197.0$), some showing a significant decrease in spike rate across dFs (e.g., right panel: $p<0.0001$, df=2, $F=105.9$) and others showing both increases and decreases in spike rate and hence, no significant overall change in spike rate (not shown). The changes in spike rate were significant in 70% ($n=29/41$) of the units tested (two factor ANOVA $p<0.05$). A significant interaction between spike rate and depth of modulation was observed in 63% ($n=26/41$) of the neurons. This indicated that, for the majority of neurons, spike rate responses to beats were dependent on both their depth of modulation and dF.

Examples of the effects of the depth of modulation on the gain of midbrain neuron responses to beats are shown in Fig. 6(C) [same neurons from Fig. 6(A)]. Both neurons exhibited large decreases in their gains with increases in the depth of modulation from 50% to 100%. The changes in the FC for the same neurons are shown in Fig. 6(D).

A summary of the effects of all increases in the depth of modulation of beats on the mean relative change in spike train parameters is shown in Table II, while plots of the frequency distributions of individuals for 50% to 100% depths are shown in Fig. 7. Histograms on the left show data for all dFs across the entire population, while those on the right show data for dFs with significant VS_{dF} at 50% depth of modulation.

Both for all dFs and dFs with significant VS_{dF} at 50%, showed significant increases in the mean relative VS_{dF} with increases in depth of modulation from 50% to 100% [Fig. 7(A)]. Significant increases in VS_{dF} were also observed for increases in depth of modulation from 75% to 100% for all dFs. However, the changes were not significant for dFs in which VS_{dF} was significant at 75% depth of modulation. Thus a subpopulation of neurons appeared to reach their maximum VS_{dF} at 75% depth of modulation.

Increases in the depth of modulation from 50% to 100% produced no significant increases in the mean relative spike rate of the population with one-third of the animals showing decreases in their mean relative spike rate with increased depth of modulation [Fig. 7(B)]. For increases in depth of modulation from 75% to 100%, there was a significant increase in the mean relative spike rate across the population for all dFs, yet spike rate changes were insignificant for dFs with significant VS_{dF} at 75% depth of modulation.

The mean differences in gain across the population for increases in depth of modulation are shown in Fig. 7(C). There was a very large and significant decrease in gain at all dFs and dFs with significant VS_{dF} at 50% depth of modulation. This indicates that although the synchronization to dF was lower at 50% depth modulation, the temporal modulation in the spike train relative to the stimulus was much higher. Decreases in gain were also significant although lower for increases in depth of modulation from 75% to 100% for all dFs and for dFs with significant VS_{dF} at 75% depth of modulation.

The changes in the FC across the population are shown in Fig. 7(D). Across all dFs and dFs with significant VS_{dF} at 50% depth of modulation, there was a significant increase in the FC. Hence, the increases in VS_{dF} with increased depth of modulation offset any decreases in spike rate. Similarly, for increases in depth of modulation from 75% to 100%, increases in FC were significant in all cases.

The distribution of changes in ISI spectral sensitivity with increases in depth of modulation of beats from 50% to

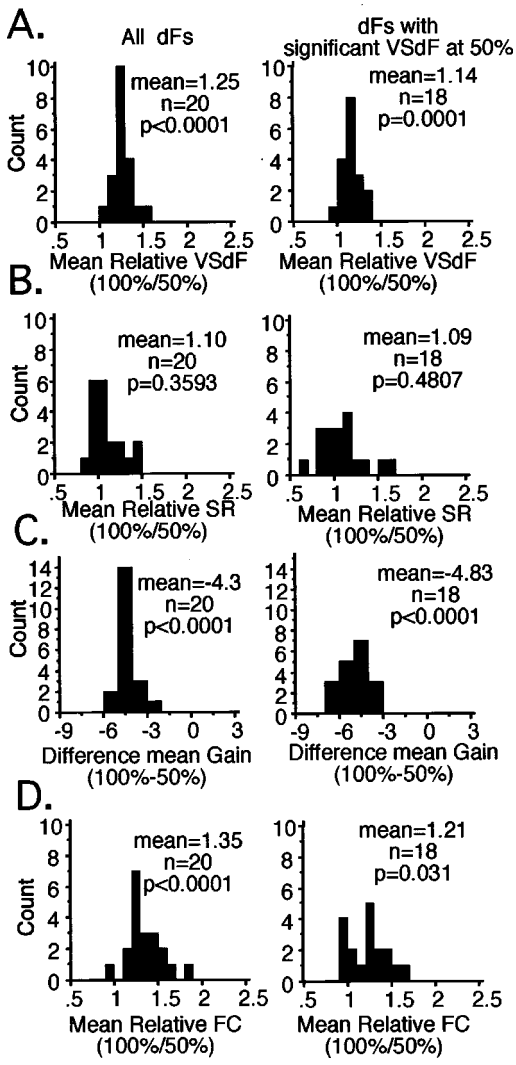


FIG. 7. Effects of changes in depth of modulation of beats on spike train parameters across the population. The plots on the left show the frequency distribution of the mean relative change (100%/50%) of the population for the designated parameter across all dFs. The plots on the right show the frequency distribution of the mean relative change of the population for the designated parameter across dFs at which VS_{dF} was significant at 50%. *P*-values are for a one-sample sign test. (A) VS_{dF}; (B) Spike rate (SR); (C) Modulation gain; (D) Magnitude of the Fourier Component (FC).

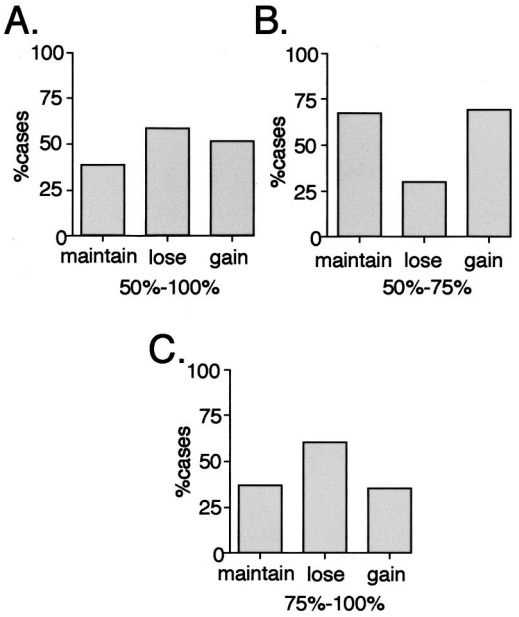


FIG. 8. Cumulative data for comparisons of ISI spectral sensitivity between ±dF beats for beats with different depths of modulation. The plots show percentage of cases of ISI spectral sensitivity that are maintained, lost, or gained with an increase in depth of modulation: (A) 50% to 100%; (B) 50% to 75%; (C) 75% to 100%.

100% is shown in Fig. 8(A) ($n=73$ cases of ISI spectral sensitivity at 50% depth of modulation). Only 40% ($n=29/73$) of the cases of ISI spectral sensitivity observed at 50% depth of modulation were maintained at 100% depth of modulation, while 60% ($n=44/73$) were lost. However, another 53% ($n=39/73$) of ISI spectral sensitivity cases were gained. For units in which responses to stimuli with 75% depth of modulation were also recorded, we compared ISI spectral sensitivity for increases from 50% to 75% [Fig. 8(B)] and 75% to 100% [Fig. 8(C)]. Interestingly, for increases of depth of modulation from 50% to 75%, only 31% ($n=10/32$) of the cases of ISI spectral sensitivity were lost, 69% ($n=22/32$) were maintained, and another 71% ($n=23/32$) were gained. In contrast, increasing the depth of modulation from 75% to 100% resulted in a greater percentage of cases of ISI spectral sensitivity being lost (62%, $n=28/45$), and fewer either gained (36%, $n=16/45$) or maintained (38%, $n=17/45$). This suggested that the greatest degree of ISI spectral sensitivity within the auditory midbrain existed when there was a 75% depth of modulation of beat signals.

C. Effects of changes in depth of modulation of AM

Responses to AM stimuli were measured in 27 neurons with 20 neurons tested at 50%, 80%, and 100% depth of modulation in 12 animals, and 7 neurons at 80% and 100% in 5 animals. Examples of the effects of changes in the depth of modulation on the spike train parameters of individual neurons are shown in Fig. 9. The effects of changes in the depth of modulation of AM signals on VS_{modF} on two neurons are shown in Fig. 9(A). Both neurons exhibited significant increases in their VS_{modF} with increases in the depth of modulation (left: $p=0.0078$, dF=2, $F=5.0$; right: $p<0.0001$, df=2, $F=18.0$). There was a significant effect of the depth of modulation of AM signals on the VS_{modF} of individual neurons in 78% of the units tested ($n=21/27$, $p<0.05$ two-way ANOVA). Five of the six neurons that showed no significant change were units tested at 80% and 100% depths of modulation only. Sixteen neurons (80%) showed significant VS_{modF} for at least one modF at 50% depth of modulation. In 22% of the neurons ($n=6/27$), there was a significant interaction between modF and depth of modulation ($p<0.05$ modF* depth of modulation two-factor ANOVA).

The changes in spike rate of the same two neurons are shown in Fig. 9(B); both neurons showed significant increases in their spike rates with increased depth of modulation (left: $p=0.0056$, df=1, $F=5.4$; right: $p<0.0001$, df=1, $F=27.22$). Across individual neurons, 81.5% ($n=22/27$) showed significant increases in their spike rates in response

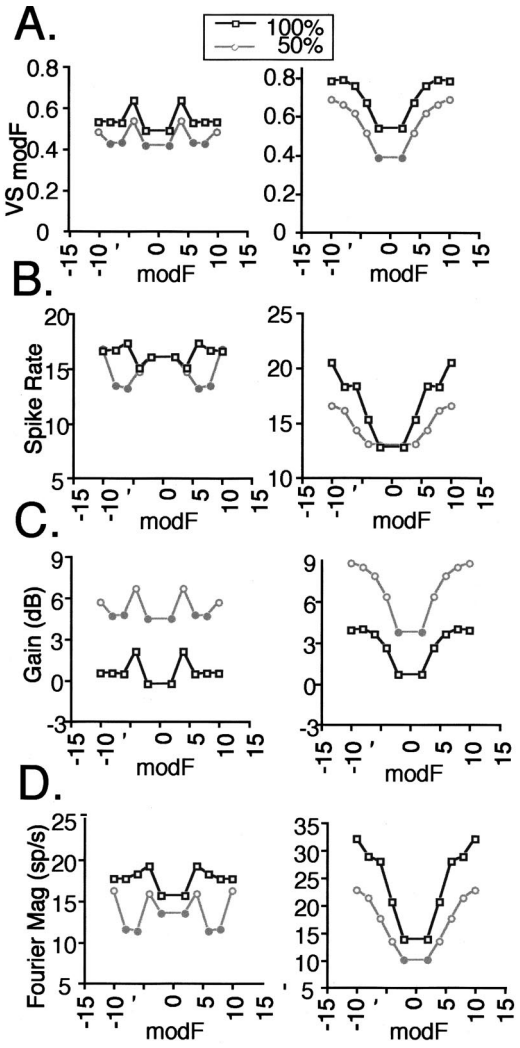


FIG. 9. Effects of changes in depth of modulation on spike train parameters for individual neurons in response to AM. The graphs show plots of the designated parameter versus AM modF for two representative neurons at 50% (circles and gray lines) and 100% (squares and black lines) depth of modulation. ModFs at which VS_{modF} was **significant** have open symbols while those with **insignificant** synchronization have filled symbols. (A) VS_{modF}; (B) Spike rate (SR); (C) Modulation gain; (D) Magnitude of the Fourier Component (FC).

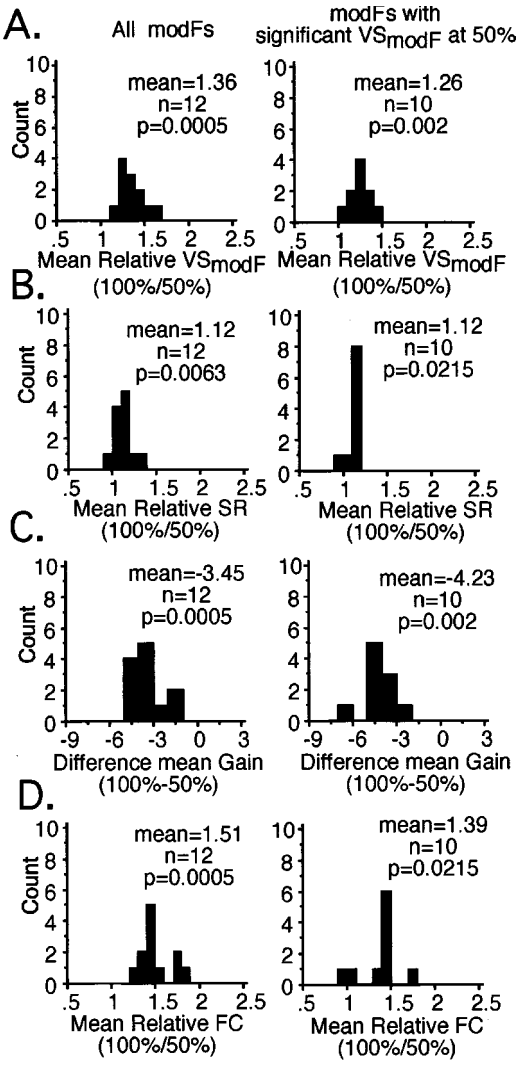


FIG. 10. Effects of changes in depth of modulation of AM signals on spike train parameters across the population. The plots on the left show the frequency distribution of the mean relative change (100%/50%) of the population for the designated parameter across all modFs. The plots on the right show the frequency distribution of the mean relative change of the population for the designated parameter across modFs at which VS_{modF} was significant at 50%. *P*-values are for a one-sample sign test. (A) VS_{modF}; (B) Spike rate (SR); (C) Modulation gain; (D) Magnitude of the Fourier Component (FC).

to increased depth of modulation of AM signals ($p<0.05$ effect of depth of modulation two-factor ANOVA). Four of five of the units which showed no significant increase were tested at only at 80% and 100% depths of modulation. In 37% of the neurons ($n=10/27$), there was a significant interaction between modF and depth of modulation ($p<0.05$ modF* depth of modulation two-factor ANOVA). Hence, spike rate profiles of AM signal modFs were dependent on their depth of modulation for approximately a one-third of the neurons.

Comparisons of the changes in gain in response to increased depth of modulation of AM signals of the same two neurons are shown in Fig. 9(C); both neurons showed large decreases in their gain. The effects of increases in the depth of modulation of AM signals of the FC are shown in Fig. 9(D); both neurons showed an increase in their FC with increased depth of modulation.

A summary of the changes in spike train parameters across the population for both 50% to 100% and 80% to 100% are shown in Table III, while frequency distributions of individual mean relative changes from 50% to 100% modF are shown in Fig. 10. Histograms on the left show data for all modFs across the entire population, while those to the right show data for modFs with significant VS_{modF} at 50% depth of modulation.

For all cases examined, the population showed significant increases in the mean relative VS_{modF} with increases in depth of modulation from 50% to 100% [Fig. 10(A)]. For increases in depth of modulation from 80% to 100%, the population continued to show significant increases in VS_{modF} for all modFs although the relative increase was lower. However, for modFs with significant VS_{modF} at 80% increases in VS_{modF} were no longer significant.

In contrast to beats, increases in the depth of modulation of AM signals from 50% to 100% produced significant in-

creases in the mean relative spike rate [Fig. 10(B)] across the population both for all modFs and modFs with significant VS_{modF} at 50% depth of modulation. While the total number of animals is considerably smaller than for tests of depth of modulation of beats, the variance in changes in spike rate was low with only one animal showing a decrease in its mean relative spike rate. For increases in the depth of modulation of AM signals from 80% to 100%, the population did not show any significant change in spike rate for all modFs or modFs in which VS_{modF} was significant at 80% depth of modulation. Thus the spike rates of responses to AM signals appear to reach their maximum at 80% depth of modulation.

The mean differences in gain across the population for increases in depth of modulation from 50% to 100% are shown in Fig. 10(C). There was a significant decrease in gain both at all modFs and modFs with significant VS_{modF} at 50%. Thus similar to increases in the depth of modulation of beats, the temporal modulation in the spike train relative to the stimulus is much higher at lower depths of modulation. For increases in depth of modulation from 80% to 100%, significant decreases in gain were still observed at both all modFs and at modFs with significant VS_{modF} at 80%.

The changes in the FC across the population are shown in Fig. 10(D). Across all modFs, there was a significant increase in the FC as well as at modFs with significant VS_{modF} at 50%. The large increases in the FC resulted from the combined increases in both the spike rate and VS_{modF} with increased depth of modulation of AM signals. For increases in depth of modulation of AM signals from 80% to 100%, there was also a significant increase in FC for all modFs and for modFs with significant VS_{modF} at 80%.

As noted in the methods, because ISI spectral sensitivity is determined by the comparison of ±dF signals, this analysis was not carried out for AM signals as only one stimulus exists for each modulation frequency.

IV. DISCUSSION

Our focus has been on assessing the encoding of the temporal and spectral features of beats and AM signals within the auditory midbrain of midshipman fish. These signals approximate concurrent and individual signals that are part of the natural repertoire of midshipman males; overlapping hums (multi-harmonic mate calls) produce beats while grunts (agonistic calls) have an AM-like structure (Bass *et al.*, 1999). Previous studies of the coding of concurrent vocal signals within the midshipman auditory midbrain showed that neurons provide a combinatorial code of the temporal and spectral features of concurrent vocal signals (Bodnar and Bass, 1997, 1999). Specifically, VS_{dF} provides an explicit temporal code of beat dF, while changes in spike rates and ISIs encode a beat's spectral information.

In this study, we examined how changes in the overall intensity of beats as well as the depth of modulation of beats and AM signals influence spike rate, VS_{dF}, ISI spectral sensitivity, MTF gain, and the FC. Our goal was to assess how these stimulus parameters influenced spike train parameters of both individual neurons and the population as a whole; a summary of the results are shown in Table IV. For increases in intensity from 6 dB to 12 dB and increases in depth of modulation from 50% to 100% changes in spike train parameters were in general similar both for all dFs and only those with significant VS_{dF} or VS_{modF} at the lower stimulus parameter. Only in the case of the effects of increase in intensity was there a small difference (across all dFs there was no change, and for dFs with significant VS_{dF} there was a small decrease in VS_{dF}). Hence, our results show that these levels of intensity and depth of modulation are encoded consistently by VS_{dF}, spike rate and/or ISIs. Moreover, no change in a single spike train parameter is correlated with any one change in a stimulus parameter. This suggests that stimulus parameters are encoded by a constellation of spike train parameters.

TABLE IV. Summary of the changes in midbrain neuron spike train parameters with changes in stimulus intensity and depth of modulation. NC=No Change.

Stimulus parameter	Spike rate	VS_{dF} or VS_{modF}	ISI spectral sensitivity
		Beats	
Intensity (6 dB to 12 dB)	Increase	NC/Decrease	Maintain & Gain
Depth of Mod (50% to 100%)	NC	Increase	Lose & Gain
		AM signals	
Depth of Mod (50% to 100%)	Increase	Increase	

Our data represent the mean values for individual animals and hence, provide a conservative estimate of the effects of stimulus parameters as low statistical power may in some cases obscure spike train parameter changes that may be present across the population. Nevertheless, for many spike train parameters, we observed consistent and significant changes across the population for a given change in stimulus intensity or depth of modulation (e.g., increases in VS_{dF} with increases in depth of modulation). However, for some spike train parameters, there was a large degree of variability in the data. This was most notable for changes in spike rate and ISI spectral sensitivity with increases in the depth of modulation of beats; both increases and decreases in spike rate were observed as well as the losses and gains of ISI spectral sensitivity.

Variability in the data could arise from a number of sources including the location of the recording site and the depth of anesthesia. In addition, variability in the temporal responses of midbrain neurons to pure tone stimuli (e.g., phasic, pauser, chopper response types) has been observed in the auditory midbrain of frogs (Gooler and Feng, 1992; Bibikov and Nizamov, 1996) and mammals (Rees *et al.*, 1996), and the medulla of fish (Kozloski and Crawford, 2000). Bibikov and Nizamov (1996) found that different response types showed variation in their degree of synchronization to the modF of AM signals depending on the depth of modulation. We made no attempts to classify midbrain neurons according to their responses to pure tones, but rather focused on their coding of beats and AM signals as a single population. Thus some of the observed variability in our data for changes in a spike train parameter could arise from pool-

ing together data from different response types. Nevertheless, as shown in our previous studies, a large percentage of midbrain neurons show significant synchronization to dF (90%) as well as spike rate (62%) and ISI spectral sensitivity (97%) in response to beats (Bodnar and Bass, 1997, 1999). Thus regardless of the possible presence of different response types within the midbrain, the majority of neurons appear to utilize a common encoding scheme for the temporal and spectral features of beats. Furthermore, some spike train parameters show very little variability in their changes (e.g., increases in VS_{dF} with increased depth of modulation). Variability in the changes of some spike train parameters and not others may also reflect the neural mechanisms underlying combinatorial coding of the temporal and spectral features of beats by midbrain neurons.

A. Combinatorial coding of concurrent vocal signals

Discerning how signals are encoded within neural spike trains entails, in part, determining whether different spike train parameters provide adequate information for accomplishing specific behavioral tasks, and whether predictions of coding hypotheses are upheld under different stimulus conditions. With regard to the representation of the spectral composition of vocal signals within CNS neural spike trains, the most widely held view of auditory coding is that frequency information is coded by the spike rate. Because auditory neurons are tuned in their spike thresholds to specific frequencies, their spike rates will vary depending on the spectral composition of a signal. However, such a coding scheme poses a severe limitation for providing a neural representation of signals with small differences in F_0's. In the case of the concurrent hums of midshipman, like the concurrent vowels of humans, the F_0's differ by only 2–10 Hz with the majority of natural dFs ranging from 1 to 4 Hz (Bodnar and Bass, 1997). Results of behavioral experiments suggest that midshipman can segregate signals with dFs of 5 and 2 Hz (McKibben and Bass, 1998; Bodnar and Bass, unpublished data). Thus the mechanisms of coding of concurrent vocal signals must permit this segregation. While midshipman midbrain neurons do exhibit low-pass and bandpass frequency tuning (Bodnar and Bass, 1997), very few units encode frequency differences in beats with dFs less than 6 Hz based on spike rate alone (Bodnar and Bass, 1999). This suggests that some other coding mechanism must also contribute to the representation of spectral information within neural spike trains.

The majority of midbrain neurons show significant changes in their ISI distributions over a specific range of intervals for beats with small dFs (Bodnar and Bass, 1999). Thus some of the information necessary for resolving small frequency differences in beats may be contained within the ISIs of midbrain neurons. Changes in the stimulus level and depth of modulation of beat signals provide tests of the hypothesis that the ISIs of midbrain neurons encode spectral information. First, if frequency information is encoded within midbrain ISIs, then the ISI spectral sensitivity of neurons observed at lower intensities should be maintained with increases in stimulus level since the spectral composition of the signal remains unchanged. Furthermore, because the number of spikes increases with intensity and consequently the number of ISIs, it is also likely that neurons will improve and/or gain ISI spectral sensitivity. Our results in fact show that in the case of increases of beat stimulus level from 6 dB to 12 dB above threshold, 73% of the cases of ISI spectral sensitivity were maintained while an additional 71% were gained. Second, a change in the depth of modulation of a signal results from a change in its spectral composition; the relative amplitude of one signal is decreased. Thus if spectral information is contained within midbrain ISIs, then changes in a signal's depth of modulation should produce changes in the ISIs of midbrain neurons. In terms of observations of significant differences in the ISIs of spike train responses to $\pm$dF beats, changes in the depth of modulation of signals would result in the loss or gain of differences in ISIs depending on how the spectral composition of the signal influences a neuron's ISI distribution. Our results are consistent with these predictions. For example, with an increased depth of modulation from 50% to 100%, only 40% of the cases of ISI spectral sensitivity were maintained, while larger percentages were lost or gained. Thus both the stimulus level and depth of modulation data sets support the hypothesis that the ISI distributions of midbrain neurons contain information related to the spectral composition of concurrent vocal signals.

Another stimulus dimension that is most often associated with spike rate coding is stimulus energy; in general, increases in stimulus energy produce increases in spike rate. Here, stimulus energy was increased both with increases in stimulus level (+6 dB) and increases in depth of modulation (+3 dB). Thus if this coding scheme holds for the coding of concurrent vocal signals, then increases in stimulus level as well as depth of modulation should produce increases in spike rate. Spike rate did in fact consistently increase with increased stimulus level for both the entire population sampled as well as the subpopulation with significant VS_{dF} at 6 dB. However, there was no significant change in spike rate with changing depth of modulation from 50% to 100% across the population with spike rate actually decreasing in one-third of the animals with increases in depth of modulation of beats from 50% to 100%. In contrast, similar increases in the depth of modulation of AM signals almost never produced decreases in spike rate. Hence, increased stimulus energy does not always give rise to an increase in spike rate, indicating that a more complex coding scheme is utilized.

With regard to the encoding of the temporal envelope of concurrent signals, midshipman midbrain neurons synchronize spike bursts to the dF (VS_{dF}) of beats (Bodnar and Bass, 1997). Furthermore, different neurons exhibit selectivity in their VS_{dF} for different dFs. Across the population, increases in intensity from 6 dB to 12 dB produced a small but significant decrease in VS_{dF} for dFs with significant VS_{dF} at 6 dB; across all dFs these changes were insignificant. Increases in depth of modulation from 50% to 100% produced far more robust significant increases in VS_{dF} for both all dFs and dFs with significant VS_{dF} at 50%. Hence, because VS_{dF} depends on dF, stimulus intensity, and depth of modulation, VS_{dF} must be combined with other spike train code parameters to uniquely specify the temporal envelope of a particular signal.

Similarly, the results for the AM stimuli used in this study also support this hypothesis; thus, both spike rate and VS_{modF} encode the temporal envelope.

Taken together, the data in this study suggest that no single spike train parameter specifies a particular stimulus parameter. Instead, the data suggest that acoustic signals are uniquely represented by coding schemes based on a combination of spike train parameters and/or population correlations. These more complex coding schemes would increase the number of combinations of signal parameters that could be individually stipulated. Hence, such a coding scheme would improve a system's coding capacity substantially. For signal processing tasks that require fine levels of discrimination such as in the segregation of concurrent vocal signals or the discrimination of beats from AM signals with similar modulation rates, increased specificity of signal coding would be particularly advantageous.

B. Transformations in neural coding within the midshipman auditory system

Midshipman auditory afferents exhibit strong synchronization to both components of beat waveforms but poor synchronization to dF (McKibben, 1998; McKibben and Bass, unpublished observations). Thus afferents provide an explicit temporal code of the individual components of a beat and/or the phase modulations of the beat waveform. In contrast, midbrain neurons show significant synchronization to dF but insignificant synchronization to the individual components (Bodnar and Bass, 1997). However, spectral differences in concurrent signals produce differences in the spike rate and ISI distributions of midbrain neurons (Bodnar and Bass, 1999). Thus there is a transformation in the coding of concurrent vocal signals within the auditory system from a temporal code of the individual components to a combinatorial code of dF and spectral information.

The period histograms of midshipman auditory afferents indicate that an afferent's probability of firing during a frequency cycle is proportional to the amplitude of the waveform during that cycle (McKibben, 1998). Preliminary data show that increases in stimulus level and depth of modulation both produced consistent increases in the spike rates and VS_{dF} of afferents (McKibben, 1998). In contrast, many midbrain neurons in the present study showed decreases in VS_{dF} with increased stimulus level and decreases in spike rate with increased depth of modulation. This suggests that there is also a peripheral to central transformation of the effects on spike train coding with respect to changes in stimulus level and depth of modulation of beats.

In a previous study, we proposed a model by which the afferent temporal code of the beat components could be transformed into the midbrain combinatorial ISI and VS_{dF} code (Bodnar and Bass, 1999). This model follows that originally proposed by Langner (1983) to explain midbrain responses to AM signals at different carrier frequencies. Under this model, midbrain combinatorial ISI and VS_{dF} responses arise from coincident inputs of two parallel pathways arising in the medulla [Fig. 11(A)]. One path is predicted to synchronize strongly to the individual components (F_0's) of a beat, while the other population synchronizes strongly to the beat dF. Convergent input of these two codes onto a coincidence detector will result in spike trains that are both synchronized to dF and contain spectral information within their ISIs. Recent neuroanatomical mapping of the acoustic circuitry in midshipman has shown that there are two distinct populations of neurons in the medulla that project to the midbrain, thus providing a structural basis for parallel coding pathways within the medulla (Bass *et al.*, 2000).

The above model could account for transformations in coding that are observed between afferent and midbrain spike trains with changes in intensity. A schematic of the spike train output of the proposed neuron populations in the medulla along with the resulting midbrain spike trains for changes in intensity is shown in Fig. 11(B). This example presumes significant VS_{dF} at the lower intensity value. As afferents produce fewer spikes in response to lower intensity signals, so too will both medullary populations; however, their spikes are more likely to occur near the waveform peaks. Hence, midbrain responses resulting from coincident spike inputs will give rise to the stronger synchronization to dF and higher MTF gains observed at lower intensities. As the intensity is increased, afferent spike rates increase, in turn increasing the number of spikes produced by both medullary populations. This leads to an increased probability of simultaneous, convergent spikes at the coincidence detector and more spikes over a longer portion of the dF period. Furthermore, because the ISIs of one population are related to the frequency composition, this constrains the minimum ISIs at the output of the coincidence detector, i.e., the ISI encoding frequency cannot be less than $1/F_0$. Thus there will be a decrease in the measured VS_{dF} at the output of the coincidence detector as more spikes will occur over a longer duration of the dF period spaced at intervals related to the spectral composition of the signal. For cases when VS_{dF} is insignificant at the lower value, VS_{dF} will generally be low due to random firing times, and hence, increase when it becomes significant at a higher intensity level.

With regard to increases in depth of modulation of beats and AM signals, the model can account for the observed increases in VS_{dF} and spike rate. In the case of lower depths of modulation, the range of changes in the peak-to-peak amplitude of the waveform is smaller, and thus the dF population will more likely produce spikes over a longer portion of the dF period [Fig. 11(C)]. The overall spike rate of the F_0 coding population will also be less because the maximum peak is lower. Hence, the resultant output of the coincidence detector will be fewer and less well synchronized dF-related spikes. Increases in the depth of modulation cause an increase in the range of the peak-to-peak amplitude of the waveform, i.e., depth of amplitude modulation. This will increase synchronization in the dF population and the overall number of spikes in the F_0 population. Thus the consequent output at the coincidence detector is a greater number of spikes centered near the peak amplitude of the signal, i.e., increased spike rate and VS_{dF}.

With increases in the depth of modulation of beats from 50% to 100%, approximately a third of the population exhibited decreases in their spike rate. Although our dataset for

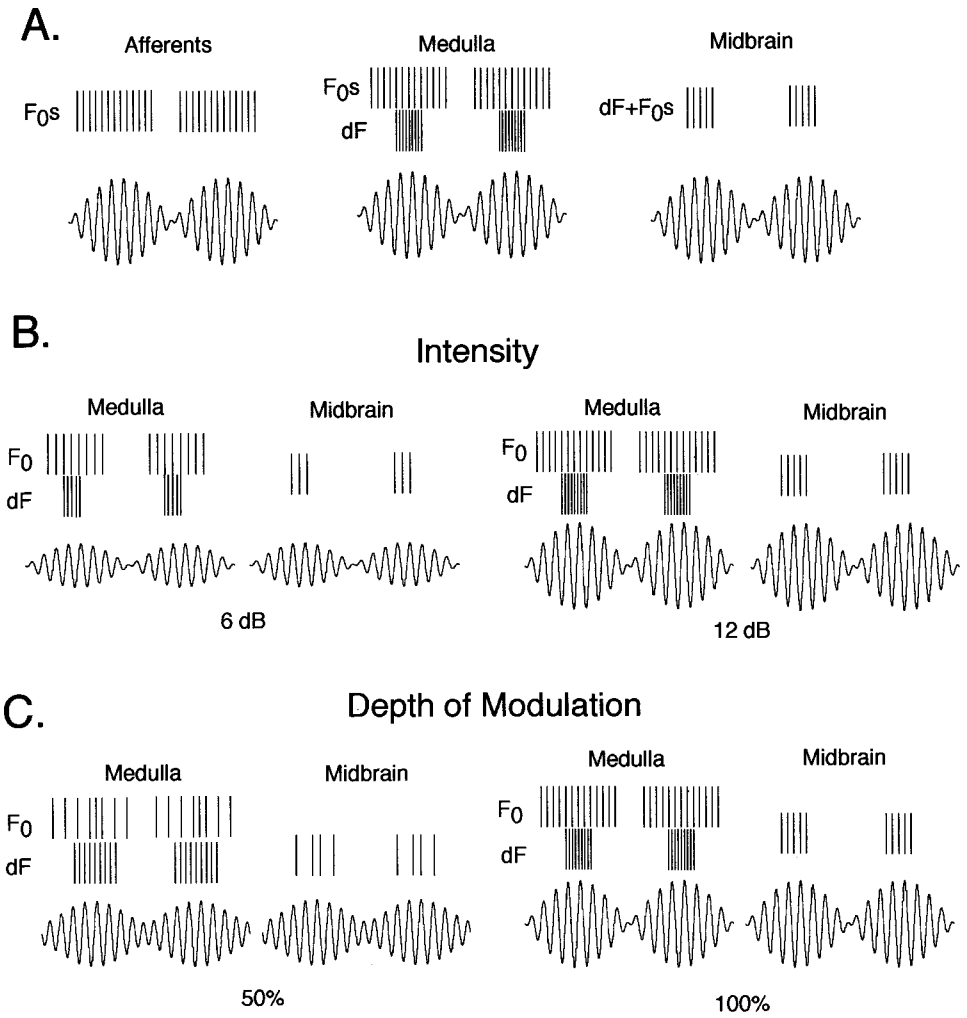


FIG. 11. Model for the temporal coding of concurrent vocal signals (modified from Bodnar and Bass, 1999). (A) A schematic of a model for the combinatorial coding of F_0 and dF information. Midshipman afferents synchronize their spike outputs to the F_0's of beat components (McKibben, 1998). Our model proposes the presence of two populations of units in the medulla, one which like afferents synchronizes to the F_0's of the beat and the other which synchronizes to dF (center panel). Midbrain neuron responses result from the convergent of these two populations. Hence, dF information is encoded by the degree of phase locking of bursts of spikes to dF (VS_{dF}), while information about F_0 is contained within the ISIs of spikes occurring within a beat period (right). (B) A schematic of the predicted spike outputs of two medullary neurons and a midbrain neuron for the model with an increase in stimulus intensity. The left panel shows responses to one signal and the right panel to a 6 dB higher stimulus level. The model predicts that at the higher intensity the midbrain units will show increases in spike rate and decreases in VS_{dF} (see text for more detailed explanation). (C) A schematic of the predicted spike outputs of two medullary neurons and a midbrain neuron for the model with an increase in depth of modulation. The left panel shows responses to a 50% depth of modulation signal and the right panel to a 100% depth of modulation signal. The model predicts that at the higher depths of modulation the midbrain units will show increases in spike rate and VS_{dF} (see text for more detailed explanation).

AM signals is admittedly small, decreases in spike rate were less frequent with increases in their depth of modulation from 50% to 100% [Fig. 10(B)]. Thus additional mechanisms would need to account for both increases and decreases in spike rate with increased depth of modulation of beats and primarily increases of spike rate with increased depth of modulation of AM signals.

A fundamental difference between beats and AM signals is that beats contain phase modulations within their waveforms while AM signals do not. As the depth of modulation of beats is increased, so too are the phase modulations. In the case of AM signals, there is no change in the phase structure of the signals. Thus neural mechanisms sensitive to the phase structure of a signal would produce differential responses to increases in the depth of modulation of beats and AM signals.

As discussed earlier, a variety of neurons with different temporal response properties such as phasic and chopper have been identified in the auditory midbrain of other species. Hence, the question arises as to whether this model can account for the presence of such response types. Because this model is a simple coincidence detector, any of the temporal features of inputs from either channel input will be preserved to some degree at the output. Thus if a medullary neuron has a phasic response to beats, the resulting midbrain neuron will also have phasic characteristics. Similarly, any intrinsic oscillations in a neuron's response, such as in the case of choppers, will also be maintained at the midbrain level. However, this mechanism is incapable of creating such response types de novo from sustained inputs. The primary characteristic of this mechanism is that it combines spectral and temporal

information from separate neuron populations into a single combinatorial code.

C. Comparisons with other vertebrates

The coding of concurrent vocal signals has been examined in only a few other studies. In cats and guinea pigs, auditory afferents temporally code the F_0's of concurrent vowels via synchronization (Palmer, 1990; Cariani and Delgutte, 1996a, b). Within the ventral cochlear nucleus (VCN) of cats, chopper and primarylike neurons exhibit synchronization to the F_0's of concurrent vowels (Keilson *et al.*, 1997). In addition, the spike rate responses of chopper neurons appear to reflect the power spectrum of the stimulus as seen through the frequency tuning curve of a unit. Hence, it is proposed that within the VCN, the stimulus energy of a signal is encoded by the discharge rates of chopper neurons while the temporal patterns of their spike train outputs encode the F_0's of the overlapping vowels. As discussed earlier, we found that the relationship between spike rate and stimulus energy is not straightforward in the midshipman midbrain.

A recent study of the auditory response properties of medullary neurons in another vocal fish, *Pollimyrus*, has revealed the presence of two distinct populations, primarylike and chopperlike neurons (Kozloski and Crawford, 2000). The primarylike neurons exhibit strong synchronization to pure tones, while the chopper neurons show other periodicities in their spike train responses. In addition, the choppers show strong synchronization to pulse-train signals. While this study did not examine the coding of concurrent vocal signals, the results do indicate that transformations in the coding of vocal signals occur within the acoustic medulla of vocal fish.

To date, there have been no studies of the coding of concurrent vocal signals within the midbrain of any other species. However, a number of studies have explored the coding of AM signals. Across a wide range of species, tuning to the modulation frequency or pulse rate of acoustic signals has been observed (mammals: Langner and Schreiner, 1988; Rees and Palmer, 1989; birds: Langner, 1983; frogs: Rose and Capranica, 1985; fish: Lu and Fay, 1993; Crawford, 1993, 1997; Bodnar and Bass, 1997). In frogs, increasing depth of modulation produces increases in synchronization to modF as well as spike rate, while increases in overall intensity results in increases in spike rate and decreases in synchronization (Rose and Capranica, 1985). In the present study, we found similar changes in spike rate and synchronization in responses to increases in the depth of modulation of AM signals and increases in the overall intensity of beat signals.

In guinea pigs, increases in stimulus intensities from 10 to 20 dB above threshold consistently produce increases in the mean spike rate and magnitude of the modF Fourier component; however, increases in intensity to higher levels often produce decreases in these spike train parameters as well as in the gain of the temporal MTFs (Rees and Palmer, 1989). In addition, at lower intensity levels, neurons generally show low pass tuning in their MTFs of AM signals, while at higher intensities ($\geqslant$30 dB) these become bandpass. In our experiments, we generally did not observe dramatic decreases in spike rates at higher intensity levels although some neurons exhibited saturation. In addition, we observed changes in dF tuning of VS_{dF} which was then reflected in the MTFs with increases in intensity in 18% of the neurons. However, we observed increased intensity levels only up to 18 dB above threshold. Hence, it is possible that at even higher intensities, spike rates and dF tuning may change in a manner similar to mammals.

In the mammalian midbrain (inferior colliculus), there is a topographical map of frequency based on narrow tuning curves (Merzenich and Reid, 1974; Semple and Aitkin, 1979; Schreiner and Langner, 1994) and an orthogonal topographical map of neurons tuned in spike rate to the modF of AM signals (Schreiner and Langner, 1988). Hence, the widely accepted view of acoustic signal coding within the mammalian central auditory system is that the spectral and temporal features of a signal are represented spatially by the spike rates of neurons within this grid. However, this mechanism poses the same limitations for coding concurrent vocal signals in mammals as those discussed earlier for midshipman. Many mammalian midbrain neurons also exhibit synchronization to AM signals and regularity in their ISIs (Rees and Palmer, 1989; Sarbaz and Rees, 1996; Rees *et al.*, 1997). Thus mammalian inferior colliculus neurons also could utilize a combination of spike rate, VS_{dF}, and ISI distributions in coding and segregation of concurrent signals.

D. Comparisons with behavioral data

Vocal signals vary in both their temporal and spectral characteristics. Understanding the relationship between behavioral responses and the central encoding of the temporal and spectral features of individual and concurrent vocal signals is essential to deciphering how salient information is extracted for making adaptive behavioral decisions. Field recordings of midshipman calls show that the intensity of nesting male hums are generally close to 125 dB re: 1 μPa about 0.5 m in front of the nest site and drop off to approximately 100 dB at 3 m (Bass and Clark, in press). In the midshipman auditory midbrain, the threshold of responses to beat stimuli ranges from 85 to 120 dB re: 1 μPa. Here, we found that the majority of midbrain neurons exhibit significant synchronization to beat dFs at 6 dB above threshold for 100% depth of modulation beats. Taken together, these data indicate that midshipman can detect beat signals with 100% depth of modulation at distances of $\geqslant$3 m away from neighboring, calling males.

With regard to spectral sensitivity, our data show that 37% of the cases of comparisons of $\pm$dF beats had significant ISI spectral sensitivity at a stimulus level of 6 dB above threshold [Fig. 5(B)]. This suggests that midshipman have the capacity to not only detect the presence of concurrent signals (significant synchronization to dF) several meters from nests, but also to discriminate the spectral composition of the overlapping signals of competing males. Furthermore, the percentage of ISI spectral sensitivity increases with intensity level. Thus an animal's ability to distinguish the spectral composition of beats likely improves as an animal approaches nest sites.

The relative amplitudes of hums will vary both with the power outputs of the individual males as well as the relative distance of a listening female or male. These variables will thus influence the depth of modulation of the beating signal. Our results show that the majority of midbrain neurons have significant synchronization at one or more dFs for beats with 50% depth of modulation. In addition, the gain of the MTFs at 50% is on average higher than at 100%, suggesting improved sensitivity at lower depths of modulation. Behaviorally, this means that females (and males) can readily detect the overlapping signals of two males at the same distance where the call of one male is twice as loud as the other, or two males with equal intensity calls, but one is twice as far away. In addition, ISI spectral sensitivity was observed at 50% depth of modulation, suggesting that midshipman are able to assess the spectral composition of these overlapping signals.

Data from two-choice phonotaxis experiments show that females exhibit distinct temperature dependent frequency preferences at dFs of 10 Hz (McKibben and Bass, 1998) and size dependent frequency preferences at dFs of 2 Hz (Bodnar and Bass, unpublished data). In addition, females also show preferences for humlike signals that are 3-dB higher in intensity when the signals have the same frequency (McKibben and Bass, 1998). How frequency and intensity preferences interact with each other remains to be explored in midshipman. However, phonotaxis experiments in frogs indicate that there can be interactions between frequency and intensity preferences (Gernardt, 1981a, b, 1987), i.e., a nonpreferred frequency in an equal intensity condition can become preferred when it is higher in intensity. Interestingly, our data show that the largest percentage of cases of ISI spectral sensitivity occurs for beats with 75% depth of modulation, i.e., one of the signals is 3 dB greater than the other [Figs. 8(B), (C)]. This suggests that a greater number of neurons can discriminate the spectral composition of concurrent signals when one is slightly higher in intensity. While speculative, it is enticing to suggest that greater spectral sensitivity when there is a relatively small difference in signal levels (3 dB) would provide for a more accurate assessment of frequency/intensity trade-offs in mate choice.

V. SUMMARY AND CONCLUSIONS

Changes in the stimulus intensity and depth of modulation of beat and AM signals produce differential changes in the spike rates, VS_{dF}, and ISIs of midshipman midbrain neuron spike train responses (Table IV). These changes are consistent with the hypothesis that the temporal and spectral features of concurrent vocal signals are represented by a combination of spike train parameters. Such a coding scheme would substantially increase the amount of information available for specifying the spectral composition of concurrent vocal signals with small differences in F_0's. Hence, this would improve the system's capacity to segregate the individual signals. The combinatorial coding of concurrent vocal signals within the midbrain results from a transformation of an afferent temporal code of the individual frequency components of beats and/or their phase and amplitude modulations. The results of this study are consistent with the hypothesis that this transformation involves two parallel coding pathways within the medulla; one population that synchronizes strongly to the individual components of a signal and another that synchronizes strongly to dF (Fig. 11). The convergence of these two populations onto a coincidence detector would produce spike trains that are both synchronized to dF and contain spectral information within their ISIs. However, the results of this study also suggest that additional mechanisms, for example ones that are sensitive to the phase structure of acoustic signals, are also likely involved. Finally, studies in other species indicate that the spike rates, temporal synchronization, and ISIs of midbrain spike trains are also correlated with changes in the spectral and temporal features of acoustic signals. Hence, as noted elsewhere for studies of other mechanisms in teleost fishes (Fay, 1993), similar neural computations may be used in the encoding of concurrent vocal signals across a wide variety of species.

ACKNOWLEDGEMENTS

We thank Margaret Marchaterre for logistical support and Andrew Mason, Matthew Weeg, and Joseph Sisneros for helpful comments on the manuscript. We also thank Charles McCulloch and H. Kern Reeve for their advice and consultation on statistical tests. This research was supported by NIH Grant No. DC-00092 to AHB.

Assman, P. F., and Summerfield, Q. (**1990**). ''Modeling the perception of concurrent vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **88**, 680–697.

Bass, A. H. (**1996**). ''Shaping brain sexuality,'' Am. Sci. **84**, 352–363.

Bass, A. H., and Clark, C. (**in press**). ''The physical acoustics of underwater sound communication,'' in *The Springer Handbook of Auditory Research*, edited by A. M. Simmons, A. Popper, and R. Fay (Springer-Verlag, New York).

Bass, A. H., Bodnar, D. A., and Marchaterre, M. A. (**1999**). ''Complementary explanations for existing phenotypes in an acoustic communication system,'' in *The Design of Animal Communication*, edited by M. Hauser and M. Konishi (MIT Press, Cambridge, MA), pp. 493–514.

Bass, A. H., Bodnar, D. A., and Marchaterre, M. A. (**2000**). ''Midbrain acoustic circuitry in a vocalizing fish,'' J. Comp. Neurol. **419**, 505–531.

Batschelet, E. (**1981**). *Circular Statstics in Biology* (Academic, New York).

Bibikov, N. G., and Nizamov, S. V. (**1996**). ''Temporal coding of low frequency amplitude modulation in the torus semicircularis of the grass frog,'' Hear. Res. **101**, 23–44.

Bodnar, D. A., and Bass, A. H. (**1997**). ''Temporal coding of concurrent acoustic signals in the auditory midbrain,'' J. Neurosci. **17**, 7553–7564.

Bodnar, D. A., and Bass, A. H. (**1999**). ''A midbrain combinatorial code for temporal and spectral information in concurrent vocal signals,'' J. Neurophysiol. **81**, 552–563.

Bodnar, D. A., and Bass, A. H. (**2000**). Unpublished data.

Brantley, R. K., and Bass, A. H. (**1994**). ''Alternative male spawning tactics and acoustic signals in the plainfin midshipman fish, *Porichthys notatus* (Teleostei, Batrachoididae),'' Ethology **96**, 213–232.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). ''Intonation and the perception of simultaneous voices,'' J. Phonetics **10**, 23–26.

Cariani, P., and Delgutte, B. (**1996a**). ''Neural correlates of the pitch of complex tones. I. Pitch and pitch salience,'' J. Neurophysiol. **76**, 1698–1716.

Cariani, P., and Delgutte, B. (**1996b**). ''Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch,'' J. Neurophysiol. **76**, 1716–1734.

Chalikia, M., and Bregman, A. (**1989**). ''The perceptual segregation of simultaneous auditory signals: Pulse train segregation and vowel segregation,'' Percept. Psychophys. **46**, 487–496.

Culling, J. F., and Darwin, C. J. (**1994**). ''Perceptual and computational separation of simultaneous vowels: Cues arising from low frequency beating,'' J. Acoust. Soc. Am. **95**, 1559–1569.
Crawford, J. D. (**1993**). ''Central auditory neurophysiology of a sound producing fish: The mesencephalon of *Pollimyrus isidori (Mormyridae)*,'' J. Comp. Physiol. **172**, 136–152.
Crawford, J. D. (**1997**). ''Feature-detecting neurons in the brain of a sound producing fish,'' J. Comp. Physiol. **180**, 439–450.
de Cheveigne, A. (**1993**). ''Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory proecsssing,'' J. Acoust. Soc. Am. **93**, 3271–3290.
de Cheveigne, A. (**1997**). ''Concurrent vowel identification. III. A neural model of harmonic interference cancellation,'' J. Acoust. Soc. Am. **101**, 2857–2865.
Fay, R. R. (**1993**). ''Structure and function in sound discrimination among vertebrates,'' in *The Evolutionary Biology of Hearing*, edited by D. B. Webster, R. R. Fay, and A. Popper (Springer, New York), pp. 229–266.
Gerhardt, H. C. (**1981a**). ''Mate call recognition in the green treefrog (*Hyla cinerea*): Importance of two frequency bands as a function of sound pressure level,'' J. Comp. Physiol. **144**, 9–16.
Gerhardt, H. C. (**1981b**). ''Mating call recognition in the barking treefrog (*Hyla gratiosa*): Responses to synthetic calls and comparisons with the green treefrog (*Hyla cinerea*),'' J. Comp. Physiol. **144**, 17–25.
Gerhardt, H. C. (**1987**). ''Evolutionary and neurobiological implications of selective phontoaxis in the green treefrog, *Hyla cinerea*,'' Anim. Behav. **35**, 1479–1489.
Gooler, D. M., and Feng, A. S. (**1992**). ''Temporal coding in the frog auditory midbrain: The influence of duration and rise time on the processing of complex amplitude-modulated stimuli,'' J. Neurophysiol. **67**, 1–22.
Keilson, S. E., Richards, V. M., Wyman, B. T., and Young, E. D. (**1997**). ''The representation of concurrent vowels in the cat anesthetized ventral cochlear nucleus: Evidence for a periodicity-tagged spectral representation,'' J. Acoust. Soc. Am. **102**, 1056–1071.
Kozloski, J., and Crawford, J. D. (**2000**). ''Transformations of an auditory temporal code in the medulla of a sound producing fish,'' J. Neurosci. **20**, 2400–2408.
Langner, G. (**1983**). ''Evidence for neuronal periodicity dertection in the auditory system of the guinea fowl—Implications for pitch analysis in the time domain,'' Exp. Brain Res. **52**, 333–345.
Langner, G., and Schreiner, C. E. (**1988**). ''Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms,'' J. Neurophysiol. **60**, 1799–1822.
Lu, Z., and Fay, R. R. (**1993**). ''Acoustic response properties of single units in the torus semicircularis of the goldfish, *Carassius auratus*,'' J. Comp. Physiol. **173**, 33–48.
McKibben, J. (**1998**). ''A neuroethological analysis of acoustic communication in the plainfin midshipman fish, *Porichthys notatus*,'' Ph.D. thesis, Cornell University.
McKibben, J. R., and Bass, A. H. (**1998**). ''Behavioral assessment of acoustic parameters relevant to signal recognition and preference in a vocal fish,'' J. Acoust. Soc. Am. **104**, 3520–3535
McKibben, J. R., and Bass, A. H. (**2000**). Unpublished results.
Meddis, R., and Hewitt, M. J. (**1992**). ''Modelling the identification of concurrent vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **91**, 233–245.
Merzenich, M. M., and Reid, M. D. (**1974**). ''Representation of the cochlea within the inferior colliculus of the cat,'' Brain Res. **77**, 397–415.
Palmer, A. R. (**1990**). ''The representation of the spectra and fundamental frequency of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear nerve fibers,'' J. Acoust. Soc. Am. **88**, 1412–1426.
Rees, A., and Palmer, A. R. (**1989**). ''Neuronal responses to amplitude-modulated and pure tone stimuli in the guinea pug inferior colliculus, and their modifcation by broadband noise,'' J. Acoust. Soc. Am. **85**, 1978–1994.
Rees, A., Sarbaz, A., Malmierca, M. S., and Le Beau, F. E. N. (**1997**). ''Regularity of firing of neurons in the inferior colliculus,'' J. Neurophysiol. **77**, 2945–2965.
Rose, G. J., and Capranica, R. R. (**1985**). ''Sensitivity to amplitude modulated sounds in an anuran auditory nervous system,'' J. Neurophysiol. **53**, 446–465.
Sarbaz, A., and Rees, A. (**1996**). ''Amplitude modulation encoding and regularly firing neurons in the inferior colliculus,'' Association for Research in Otolaryngology Midwinter Meeting **19**, 155.
Schreiner, C. E., and Langner, G. (**1988**). ''Periodicity coding in the inferior colliculus of the cat. II. Topographical organization,'' J. Neurophysiol. **60**, 1823–1840.
Schreiner, C. E., and Langner, G. (**1994**). ''Laminar fine structure of frequency organization in auditory midbrain,'' Nature (London) **388**, 383–386.
Semple, M. N., and Aitkin, L. M. (**1979**). ''Representation of sound frequency and laterality by units in the central nucleus of cat inferior colliculus,'' J. Neurophysiol. **42**, 1626–1639.

LETTERS TO THE EDITOR

Recording depth of the heterodyne laser interferometer for cochlear vibration measurement

Tianying Ren
Oregon Hearing Research Center, Department of Otolaryngology and Head & Neck Surgery, Oregon Health Sciences University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97201-3098

Alfred L. Nuttall
Oregon Hearing Research Center, Department of Otolaryngology and Head & Neck Surgery, Oregon Health Sciences University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97201-3098 and Kresge Hearing Research Institute, Department of Otolaryngology, Head & Neck Surgery, The University of Michigan, 1301 East Ann Street, Ann Arbor, Michigan 48109

(Received 25 May 2000; accepted for publication 10 November 2000)

Measurement of the cochlear partition vibration as a function of the optical-axis (*z*-axis) position in the gerbil cochlea showed that the velocity distributes over a range of more than 300 μm, which is larger than the thickness of the cochlear partition. This finding suggests that the recording depth (RD) of the heterodyne interferometer probably is not as small as reported in the literature. In the current experiment, the RD of the heterodyne laser interferometer was studied by measuring the velocity of a vibrating mirror as a function of the *z*-axis position. Results demonstrate that the optical sectioning characteristic, measured by the intensity of the reflected laser beam as a function of the *z*-axis position, is not able to correctly estimate the RD of the heterodyne interferometer: the RD is much larger than optical sectioning, indicating a poor spatial resolution along the *z* axis. [DOI: 10.1121/1.1337957]

PACS numbers: 43.64.Kc, 43.64.Yp, 42.62.Be [LHC]

To understand cochlear mechanical mechanisms, an important goal is to measure cellular vibration of the organ of Corti. Willemin *et al.* (1988) designed a heterodyne laser interferometer with optimized sensitivity, capable of measuring cochlear partition vibration without using any artificial reflector. Recently, a similar but displacement-sensitive heterodyne laser interferometer was developed to measure the basilar-membrane vibration in the cochlea (Cooper, 1999). For more precise and complete measurement of cochlear cellular vibration, it is essential to characterize the *z*-axis resolution of the interferometer. To study the recording volume of the cochlear partition vibration, Khanna and Koester (1989) and Cooper (1999) characterized the optical sectioning features of their interferometers based on the relation between the relative power of the reflected light and the focus position. They found that the thickness of the section, when the carrier power is within 3 dB of maximum, was approximately 10 to 20 microns, depending on the objective lens used. Khanna and Koester (1989) believed that spot size and optical sectioning characteristics define the volume of tissue from which the light reaching the detector would contribute to the heterodyne interferometer signal. Their estimated measurement volume is smaller than the size of an outer hair cell (Khanna and Koester, 1989). A series of studies on cellular vibration in the cochlea was carried out using this laser interferometer (Khanna *et al.*, 1989a,b,c,d,e; Khanna and Hao, 1999; Hao and Khanna, 2000).

Recently, using a newly developed laser interferometer microscope (Ren and Nuttall, 2000), we have measured the vibration of the cochlear partition as a function of the *x*-, *y*-, and *z*-axes in the gerbil basal turn. The results show that the velocity distributes over a range of more than 300 μm along the *z* axis (Fig. 1), which extends beyond the thickness of the cochlear partition of the basal turn ($<$100 μm) (Edge *et al.*, 1998). This distribution is much wider than the optical sectioning depth of the microscope and leads us to a hypothesis that the recording depth (RD) of heterodyne interferometers may not be as small as has been indicated in the literature (Khanna and Koester, 1989). To test this hypothesis, we studied the RD by measuring the velocity of a vibrating mirror as a function of the distance from the focal plane to the target. Since the reflected laser is from a well-defined source, and the vibration magnitude ($<$0.01 μ when the velocity is 50 μ/s at 1000 Hz) is much smaller than the signal spread along the *z* axis ($>$100 μ), the spread of the velocity along the *z* axis, measured by the velocity as a function of *z* posi-

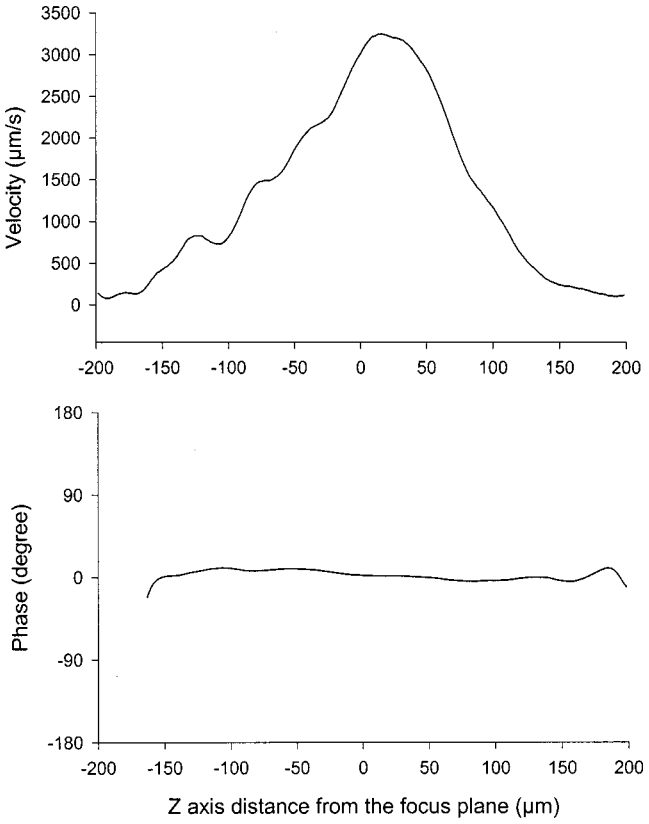


FIG. 1. The amplitude and phase of the velocity of the cochlear partition in the basal turn of a sensitive gerbil cochlea in response to an 80-dB SPL, 14-kHz tone were measured as a function of the *z*-axis position. The object beam of the heterodyne interferometer was focused on the second row of outer hair cells through a hole in the lateral wall of the scala tympani. The amplitude and phase curves of the velocity show that the velocity distributes over a range of more than 300 μm along the *z* axis. The BM was located on the positive side of the zero.

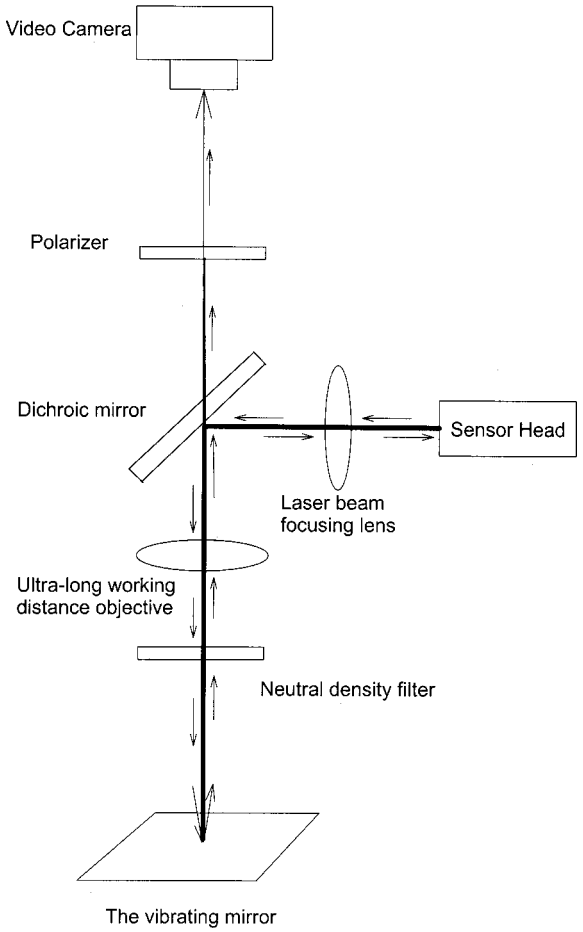


FIG. 2. A schematic illustration of the laser interferometer microscope for measuring the RD. The laser beam from the optical head reaches the dichroic mirror and is focused on the vibrating mirror using a long working distance objective. Light reflected from the mirror is collected by the objective and returned to the dichroic mirror and the laser interferometer. The neutral density filter in front of the objective is used to decrease the object beam optical power. A motorized 3D positioning system is used to move the microscope along the *z* axis at the rate of 1 μm/s.

tion, represents the RD. If the RD is small, the velocity will be detected only at a small distance from the reflected source in the *z* dimension, and the velocity curve will show a narrow peak. Contradictory to this, if the RD is infinitely large, then the velocity will be independent of the *z* position, and the velocity curve will be flat. It was found that the optical sectioning characteristics are not able to describe the RD of the laser interferometer and that the RD is much larger than previously reported.

A diagram of the instrument setup is presented in Fig. 2. A heterodyne interferometer (Polytec, Waldbronn, Germany) was used in this study. The sensor head of the interferometer was coupled into a custom-built microscope mounted on a $z-y-z$ positioning system, using a dichroic mirror. The object beam was directed into a Mitutoyo infinity-corrected long working distance objective (Mitutoyo M Plan Apo 20×, N.A. 0.42, Japan) and focused on the surface of the front-surface mirror, which was mounted on a speaker diaphragm. The reflected light from the mirror was collected by the objective, returned to the dichroic mirror, and directed back to the interferometer.

For visualization, the mirror was illuminated by a white light source. The scattered light from the focus plane was collected and transmitted to the dichroic mirror through the objective lens. The majority of the scattered white light passes through the dichroic mirror and forms an image at the sensor plane of the video camera. The laser beam focus spot was closely monitored during data acquisition. A polarizer, positioned between the dichroic mirror and the transfer lens, was used to attenuate laser intensity and ensure a clear image of the laser focus spot.

The speaker was driven by a 1000-Hz sinusoidal electrical signal, resulting in about 50-μm/s vibration. The vibration velocity of the mirror was measured by the interferometer, with sensitivity of 5 mm/s/V, 100-kHz low pass. The amplitude and phase of the velocity signal, measured by a lock-in amplifier (SR830, Stanford Research Systems, Sunnyvale, CA), and the dc carrier signal from the output of the signal processor were continuously recorded using a custom-written LABVIEW-based chart recorder with a sample rate of 2 samples/s. The laser–focus plane was moved from 200 μm above to 200 μm below the mirror surface at the speed of 1 μm/s. The measurements were carried out under the following conditions: without a neutral density filter, or with a 0.5-, 1.0-, 2.0-, 3.0-, or 4.0-o.d. neutral density filter.

It was found that, in general, the signal-to-noise ratio of the amplitude and phase of the velocity and the carrier level

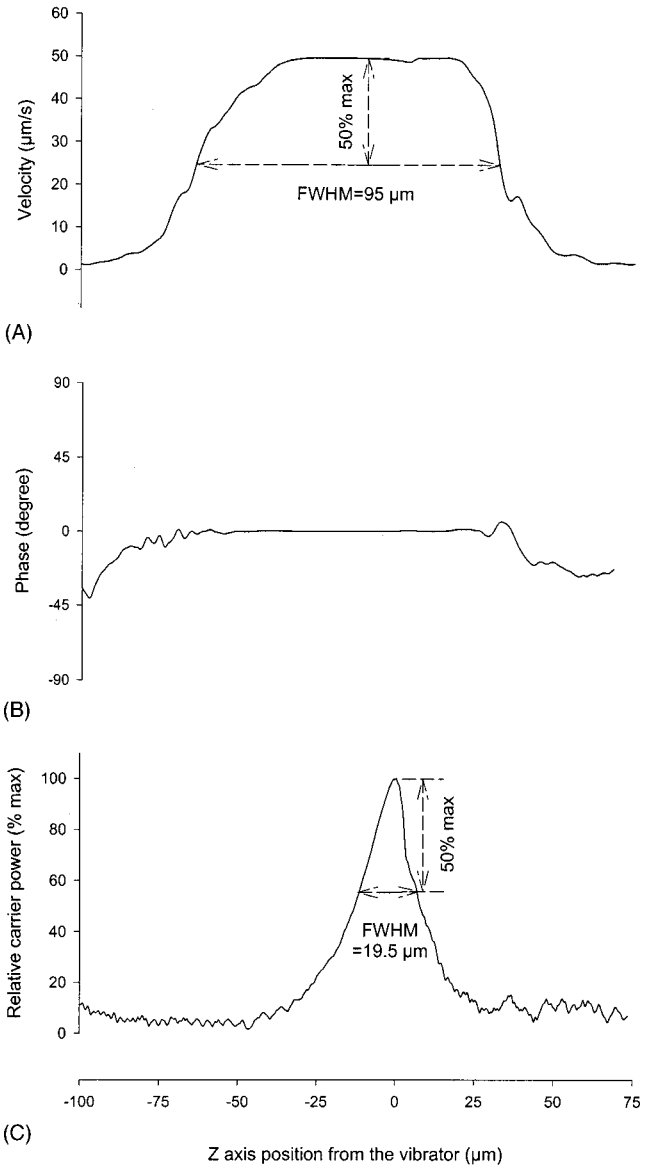


FIG. 3. The amplitude and phase of the velocity and the relative carrier power were measured as a function of the *z*-axis distance from the mirror. A 20× objective with 0.42 NA and a 2.0 o.d. neutral density filter were used for this measurement (i.e., reflectance reduced by 10^{-4}). The optical power of the carrier (C) reaches maximum at the *z*-axis position zero and falls off in an approximately symmetric manner in either direction. The full width of the carrier curve at the half-maximum (FWHM) is 19.5 μm. The FWHM of the velocity (A) is 95 μm, which is approximately four times wider than that of the carrier (C). The flat phase curve (B), corresponding to the amplitude, confirms that velocity is independent of the *z*-axis position over a wide range above and below the reflective surface.

are dependent on the optical attenuation induced by neutral density filters. Without any neutral density filter or with a 0.5- or 1.0-o.d. filter, the amplitude and phase of the velocity were independent of the *z*-axis position over the observed range. When no filter was used, the carrier was independent of the focal plane in the range of more than 20 μm above or below the mirror surface.

The relationship between the velocity and carrier is presented in Fig. 3. The amplitude and phase of the velocity and the carrier are plotted as a function of *z*-axis position. Figure 3(C) clearly shows that carrier level is *z*-axis position dependent. The optical power reached the maximum near the focus spot on the mirror, confirmed by video image of the smallest focus spot. The optical power of the object beam falls off in an approximately symmetrical manner in either direction. The full width of the carrier curve at the half-maximum optical power (FWHM), indicated by carrier level, is 19.5 μm for the 20× objective used in this experiment. These findings are very similar to the results reported in the literature (Khanna and Koester, 1989; Cooper, 1999).

However, the shapes of the amplitude and phase curves of the velocity are completely different from that of the carrier. The plateau of plot 3(A) shows that the amplitude of the velocity signal, ranging 25 μm above or below the mirror surface, is independent of the *z*-axis position. The FWHM of the amplitude of the velocity is 95 μm, which is approximately four times wider than that of the carrier 3(C). The flat phase curve 3(B) confirms that the velocity signal is independent of the *z*-axis position of the focal plane.

The terminology of optical sectioning was developed to describe techniques for illuminating and visualizing only a thin layer of the specimen. It is generally achieved by eliminating the light that is backscattered from regions other than the observed plane (Wilson, 1985). An application example of optical sectioning is confocal microscopy, which achieves optical sectioning by imaging a small aperture onto the object plane, then reimaging that illuminated area onto an aperture in the imaging system. Light reflected from regions other than the illuminated area is blocked by a circular or slit-shaped aperture in the imaging system. A two-dimensional (for a pinhole) or one-dimensional (for a slit) scan is used for this kind of optical sectioning application. To study cellular mechanics in the cochlea, Koester *et al.* (1989) developed an incident light optical sectioning microscope for visualization of the cellular structure of the cochlear partition in guinea pigs. They reported that the optical sectioning capability makes it possible to visualize cellular details *in vivo*. Khanna and Koester (1989) introduced the optical sectioning concept to describe the measurement depth of the heterodyne interferometer. They studied its optical sectioning characteristics by measuring relative object beam power at the detector as a function of the focal plane. Considering the dimension of an outer hair cell in the apical turn of the guinea pig cochlea, these authors stated that the measuring volume of their heterodyne interferometer was smaller than that of most cell types in the organ of Corti, and thus it is possible to measure the vibration of single cells. A similar optical sectioning characteristic of a custom-built heterodyne interferometer was reported by Cooper (1999). In the current experiment, the carrier measured as a function of the *z*-axis position is consistent with the results reported from the above studies.

A fundamental question is how the RD of the heterodyne interferometer should be measured. Although the carrier level, relative to the optical power of the object beam, can be used as a measure of characteristics of the optical sectioning, it may not be used to determine the heterodyne interferometer velocity or displacement ''measuring depth,'' for the following two reasons. First, the output of the heterodyne interferometer is the vibration velocity or displacement derived by electronic processing of the interference signal

between the reference and object laser beams, rather than the carrier signal level (Willemin *et al.*, 1988). Second, the vibration magnitude is not linearly related to the analyzed carrier strength. These features contribute to the most important advantage of the heterodyne interferometer: insensitivity to intensity fluctuations of the interfering beams. This is theoretically and experimentally supported by previous works (Willemin *et al.*, 1988; Cooper, 1999). The photocurrent of the photodetector in a heterodyne interferometer is given by

$$i(t) = a + b\cos[\Delta\omega t + \Phi(t)], \tag{1}$$

where a is the dc and b the ac amplitude of the photocurrent, $\Delta\omega$ the frequency difference between the reference beam and the object beam reflected from the vibrating object, and $\Phi(t)$ is the phase difference between the two interfering beams. The information about the object movement is contained in the phase $\Phi(t)$. The phase $\Phi(t)$ was detected electronically by the signal processor of the interferometer. As the above equation shows, the phase angle is not affected by the intensity fluctuations of the interfering beams. The result in Fig. 3(A) is consistent with the above theoretical expectation, demonstrating that the velocity derived from the phase $\Phi(t)$ is independent of the *z*-axis position over a wide range. However, the carrier signal level affects the carrier-to-noise ratio (CNR), the power ratio of the beat signal, and the noise after detection, all of which together essentially determine the sensitivity. The contribution of the carrier to the CNR is described by

$$\mathrm{CNR} = (\eta P_{\mathrm{OB}}/h\nu)/B, \tag{2}$$

where η is the quantum efficiency, P_{OB} the light power in the object beam, $h\nu$ the photon energy, and B the bandwidth. Equation (2) predicts our finding that the noise floor is a function of the carrier signal level.

The carrier intensity as a function of *z*-axis position is largely determined by the optical properties of the objective lens used and by the reflectance of the object, while the velocity or displacement results from the phase difference between the reference and the phase-shifted beams. The relationship between the carrier [Fig. 3(C)] and the amplitude of the velocity signal [Fig. 3(A)] is mainly determined by the carrier signal level and the signal processor used for extraction of $\Phi(t)$ from the photodetector signal. The signal processor determines the threshold of the carrier for phase detection. As long as the carrier level is above threshold, the velocity signal is independent of carrier fluctuation. High reflectance objects will produce stronger carrier beams and potentially larger RD.

Measurement of cochlear partition vibration presents optical conditions, which are more complicated than measurement from a flat surface. Willemin *et al.* (1988) pointed out that, for a given incident light power, it is evident that the curvature and the roughness of the vibrating surface have a strong influence on the carrier-to-noise ratio. Khanna and Koester (1989) found that a diffuse reflector decreases the optical sectioning capability of the measurement. Although we do not know the reflection characteristics of the cochlear partition along the *z* axis, it certainly does not have a flat reflective surface. In case of diffuse reflections, the effective coherent back-reflected optical power can be considerably smaller than the total power of the reflected light, since the effective object beam power is a vector summation of each reflected beam from different reflective targets. Thus, the RD for the cochlear partition vibration probably could be even larger than that for a vibrating mirror.

In summary, this study demonstrates theoretically and experimentally that optical sectioning characteristics will not be able to accurately indicate the *z*-axis resolution of the heterodyne interferometer. The *z*-axis spatial resolution of the heterodyne interferometer can be much poorer than optical sectioning. These findings imply that the most commonly used heterodyne laser interferometer is not able to resolve cellular vibration in the *z*-axis direction. An alternative technique is needed for this physiologically important micromechanical measurement.

ACKNOWLEDGMENTS

We thank Scott Matthews, Ph.D., and Edward Porsov, M.S., for valuable discussion on this study. Supported in part by research grants from the National Institute of Deafness and Other Communication Disorders (Nos. R01 DC00141 and R03 DC033642), the National Institutes of Health, the Research Fund of the American Otological Society, the Medical Research Foundation of Oregon, and VA RR&D Center Grant No. RCTR-597-0160, Portland, VAMC.

Cooper, N. P. (**1999**). ''An improved heterodyne laser interferometer for use in studies of cochlear mechanics,'' J. Neurosci. Methods **88**, 93–102.

Edge, R. M., Evans, B. N., Pearce, M., Richter, C. P., Hu, X., and Dallos, P. (**1998**). ''Morphology of the unfixed cochlea,'' Hear. Res. **124**, 1–16.

Hao, L. F., and Khanna, S. M. (**2000**). ''Mechanical nonlinearity in the apical turn of the guinea pig organ of Corti,'' Hear. Res. **148**, 31–46.

Khanna, S. M., Flock, Å., and Ulfendahl, M. (**1989a**). ''Changes in cellular tuning along the radial axis of the cochlea,'' Acta Oto-Laryngol., Suppl. **467**, 163–173.

Khanna, S. M., and Hao, L. F. (**1999**). ''Nonlinearity in the apical turn of living guinea pig cochlea,'' Hear. Res. **135**, 89–104.

Khanna, S. M., and Koester, C. J. (**1989**). ''Optical sectioning characteristics of the heterodyne interferometer,'' Acta Oto-Laryngol., Suppl. **467**, 61–67.

Khanna, S. M., Ulfendahl, M., and Flock, Å. (**1989b**). ''Changes in cellular tuning along the length of the cochlea,'' Acta Oto-Laryngol., Suppl. **467**, 157–162.

Khanna, S. M., Ulfendahl, M., and Flock, Å. (**1989c**). ''Mechanical tuning characteristics of outer hair cells and Hensen's cells,'' Acta Oto-Laryngol., Suppl. **467**, 139–144.

Khanna, S. M., Ulfendahl, M., and Flock, Å. (**1989d**). ''Tuning of harmonic components in cellular mechanical responses,'' Acta Oto-Laryngol., Suppl. **467**, 205–208.

Khanna, S. M., Ulfendahl, M., and Flock, Å. (**1989e**). ''Waveforms and spectra of cellular vibrations in the organ of Corti,'' Acta Oto-Laryngol., Suppl. **467**, 189–193.

Koester, C. J., Khanna, S. M., Rosskothen, H., and Tackaberry, R. B. (**1989**). ''Incident light optical sectioning microscope for visualization of cellular structures in the inner ear,'' Acta Oto-Laryngol., Suppl. **467**, 27–33.

Ren, T., and Nuttall, A. L. (**2000**). ''Measurement of the basilar membrane vibration at the basal turn in sensitive gerbil cochlea,'' The Twenty-third Midwinter Research Meeting, The Association for Research in Otolaryngology, St. Petersburg, FL, 20–24 February 2000.

Willemin, J. F., Dändliker, R., and Khanna, S. M. (**1988**). ''Heterodyne interferometer for submicroscopic vibration measurements in the inner ear,'' J. Acoust. Soc. Am. **83**, 787–795.

Wilson, T. (**1985**). ''Scanning optical microscopy,'' Prog. Clin. Biol. Res. **196**, 103–113.

A consideration of the normalization that is typically included in correlation-based models of binaural detection

Steven van de Par
IPO, Center for User-System Interaction, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands and Surgical Research Center, Department of Surgery (Otolaryngology) and Center for Neurological Sciences, University of Connecticut Health Center, Farmington, Connecticut 06030

Constantine Trahiotis and Leslie R. Bernstein
Surgical Research Center, Department of Surgery (Otolaryngology) and Center for Neurological Sciences, University of Connecticut Health Center, Farmington, Connecticut 06030

(Received 11 April 1998; revised 1 February 2000; accepted 30 October 2000)

An analysis of binaural detection and new data that elucidate the nature and precision of normalization that must be assumed if binaural detection is accomplished via mechanisms that effectively compute the coefficient of cross correlation is presented. Based on that analysis, it is argued that the precision of normalization required to remove deleterious effects resulting from variations in the levels of the stimuli is so great that it is highly unlikely that normalization, *per se*, actually occurs as part of binarual processing. Instead, it appears more likely that binaural processing is accomplished via ''subtractive'' mechanisms, such as the one originally described by Durlach [J. Acoust. Soc. Am. **35**, 1206–1218 (1963)]. Within that framework, deleterious effects that could result from variations in the levels of the stimuli simply do not arise. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1336136]

PACS numbers: 43.66.Pn, 43.66.Ba [JWH]

Many studies have shown that binaural detection and discrimination can be accounted for in terms of changes in the normalized interaural correlation of the stimuli arriving at each ear (e.g., Pollack and Trittipoe, 1959a,b; Robinson and Jeffress, 1963; McFadden, 1968; Osman, 1971; Gabriel and Colburn, 1981; Durlach *et al.*, 1986; Bernstein and Trahiotis, 1996a,b; Bernstein *et al.*, 1999).

The purpose of this paper is to discuss whether it is reasonable to assume that the implied normalization can be accomplished with the precision that appears to be necessary to overcome the variability of narrow-band stimuli typically used in binaural experiments. We will argue, on the basis of stimulus-based variability in the power of narrow-band-noise maskers and on the basis of new empirical data, that the precision required to produce the necessary ''normalization'' seems to be unachievable. In addition, we show how binaural models based on subtraction (e.g., Durlach, 1963; Schneider and Zurek, 1989; Cai *et al.*, 1998), can account for all of the data.

In order to convey our concern, let us consider the detection of interaurally out-of-phase tonal signals gated on and off against a background of diotic narrow-band noise in a standard No$S\pi$ detection task. An important characteristic of narrow-band masking noise is that its short-term power varies greatly from moment to moment. It follows that the *non-normalized* magnitude of the interaural correlation of the narrow-band noise, which results from the multiplication of the waveforms in the two channels, varies in the same way as the power in each monaural channel. The problem we are addressing concerns the small changes in interaural correlation required for binaural detection reported in the studies cited above. The issue is whether the small changes in correlation could be detected against the inherent variability in baseline correlation that would occur if normalization were omitted.

In order to pose the problem quantitatively, let us assume that the bandwidth of the diotic masking noise is 100 Hz and the listener computes the interaural correlation over the complete duration of a 100-ms, $S\pi$, tonal signal. As shown by Green and Swets (1966, p. 214), the standard deviation of the energy for our masking noise, evaluated over 100 ms, would be 30% of its average energy. This means that about two-thirds of the time one would measure values of energy that would be between 70% and 130% of the true average energy of the masker. Complimentarily, one-third of the time, one would measure values outside those bounds. Now, if we assume that the ''internal'' interaural correlation *does not* include normalization, then the internal interaural correlation would have the same mean-to-sigma ratio as the energy of masker.

Let us next make the reasonable assumption for our stimulus condition that, at threshold, the listener can detect a decrease in the normalized correlation, from 1.0, of about 0.6% (Zurek and Durlach, 1987). Were the correlation *not* normalized, the decrease in correlation of 0.6% would have to be detected against changes in the baseline correlation (stemming from monaural fluctuations in power) that are on the order of 30% of the true average value! This would result in a value of d' of only about 0.02, which is far below the d' of 1.0 typically used to represent threshold. Thus, it appears that the interaural correlation must be normalized in order for it to be able to account for binaural detection thresholds.

We decided to conduct an experiment that would require even more precise (and possibly unachievable or unrealistic)

normalization of the interaural correlation in order to account for the data. We measured listeners' abilities to detect changes of the interaural correlation coefficient while roving the sound-pressure levels of the stimuli over a range of 30 dB, a much wider variation in level than measured in conventional, nonroved, NoSπ stimuli. We collected data using our ''standard'' four-interval, two-cue, two-alternative, adaptive forced choice procedure (e.g., Bernstein and Trahiotis, 1992). Three young, normal-hearing adults served as listeners. Their task was to decide whether the second interval or third interval differed from the other three intervals, the difference being a decrease in interaural correlation of a diotic band of noise produced by suitably mixing two independent sources of noise (van der Heijden and Trahiotis, 1997). The noise was centered on 500 Hz and its bandwidth was either 50 or 500 Hz and its duration was either 20, 40, or 400 ms. All parametric combinations of bandwidth and duration were utilized. The short duration stimuli were included to see if the binaural system could adapt quickly to changes in the level of the stimulus.

In one set of conditions, correlation discrimination thresholds were measured at overall sound-pressure levels of 60, 70, or 80 dB, and the level of the noise was ''fixed'' during the blocks of trials used to provide an estimate of the threshold. In another set of conditions, the overall sound-pressure levels of the stimuli level was ''roved'' over a range of 30 dB (55 to 85 dB) on each interval. Sound-pressure levels were randomly chosen from a 30-dB range centered at 70 dB and all sound-pressure levels, to the nearest decibel, occurred with equal likelihood. Note that the three sound-pressure levels in the fixed conditions were those that occupy the middle of three, 10-dB wide, adjacent, portions of the total range of 30 dB used in the roving-level condition. In this way these three conditions should provide a good estimate of the mean performance for the constant level conditions in the relevant level range.

As can be seen from the data presented in Fig. 1, correlation discrimination thresholds ($\Delta\rho$) obtained while roving the levels of the stimuli of a range of 30 dB were only slightly poorer than those obtained when the levels of the stimuli were held constant. This outcome could be interpreted to mean, in a strict sense, that the binaural system actually normalizes its inputs and does so within less than 20 to 40 ms.

We do not, however, believe it is reasonable to assume that normalization played a part in the processing of our stimuli. Note that the listeners were able to discriminate changes in interaural correlation of only about 0.02 when the duration of the stimuli was 400 ms and the levels of the stimuli were roved over a 30-dB range! In that condition, the variability in the *effective* stimulus power *after* normalization could not exceed 2.0% in order to achieve a d' of 1.0, a value sufficient to account for the threshold change in correlation.

Now, let us examine how precise the interval-by-interval and trial-by-trial normalization of the stimuli would have to be to support discrimination. The overall power of the stimuli themselves varied by a factor of 1000 because the levels of the stimuli were roved over a 30-dB range. Therefore, in order to achieve a d' of 1.0 for a change of correlation of 0.02, the variability in stimulus power would have to be *reduced* by a factor of about 50 000 (i.e., 1000/0.02). That is, the binaural system would have to estimate the levels of the stimuli within 2% of their average values when those values varied over a range of 30 dB. That is tantamount to being able to measure accurately changes of level of only about 0.09 dB over a range of 30 dB! It is for these reasons that we believe that the auditory system probably does not, in a direct manner, actually normalize the stimuli presented to each ear.

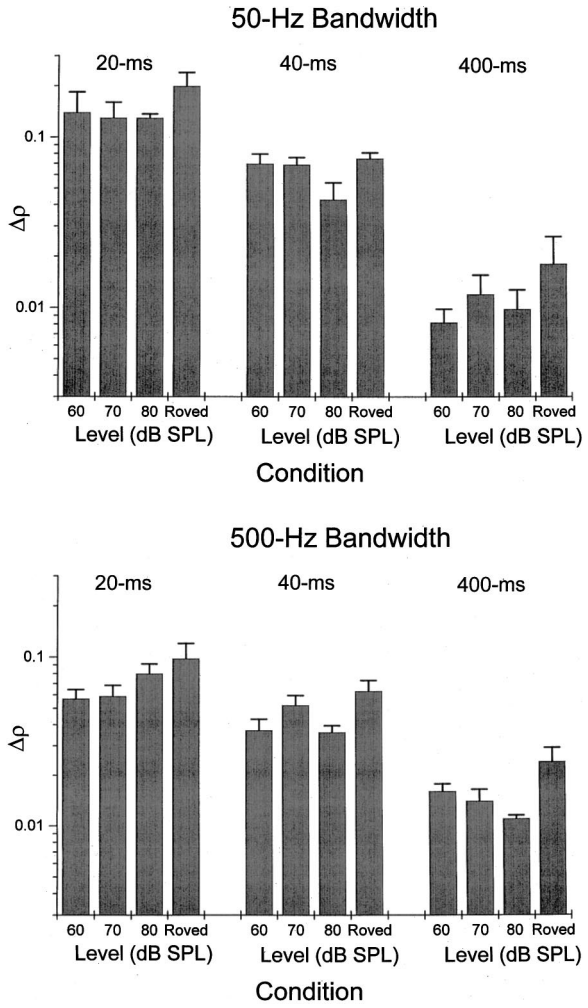


FIG. 1. Threshold changes in interaural correlation ($\Delta\rho$) as a function of the level of the stimuli in fixed- and roving-level conditions. The thresholds are represented by groups of histograms segregated according to duration. Each histogram represents the average threshold computed across three listeners and error bars represent one standard error of the mean. The top panel displays the thresholds obtained when the bandwidth was 50 Hz; the bottom panel displays the thresholds obtained when the bandwidth was 500 Hz.

Similar difficulties inherent in the normalization of interaural correlation occur when one considers that the internal process that achieves the cross product between the two waveforms would need to have a temporal window of integration that precisely matches the temporal window required to estimate the energy in each monaural channel. The two temporal windows would have to be matched quite accurately. For the NoSπ stimuli in our original example, for which normalization was assumed to reduce the variability in each channel from about 30% of the mean value to about

0.6% of the mean value, the durations of the temporal windows would have match each other within a margin of about 0.04%. This appears to be another constraint on the precision with which normalization, as a process, could be actualized.

Given the arguments that we have presented concerning the nature of normalization required for binaural detection, let us now consider what adaptations could be made to models of binaural processing that use cross correlation in order to account for binaural detection under stimulus conditions in which the levels of stimuli vary over a large range. For example, the physiologically based cross-correlation position-variable model of Stern and Colburn (1978) accounts for detection in the NoSπ condition by assuming that the listener performs a point-by-point evaluation of the cross-correlation function. It is then assumed within the model that an optimal detector adds, in optimal fashion, all the differences between the cross-correlation function of the diotic noise and the cross-correlation function of the NoSπ stimulus. This results in a single decision variable that can be shown to be subject to the deleterious effects of changing stimulus level that we have documented above.

It is the case, however, that *optimal additions* of changes in the cross-correlation function are not necessarily the *optimal manner by which* the information in the cross-correlation function can be used. For our example of NoSπ detection with narrow-band maskers, a better strategy would be to subtract cross-correlation values for positive values of interaural delay from cross-correlation values at corresponding negative values of interaural delay. This strategy would describe a device that measures asymmetries in the cross-correlation function. For our NoSπ example, the asymmetries arise from the probabilistic nature of the moment-by-moment values of external delay occurring within the stimulus. The same type of asymmetries also occur when the interaural correlation of a diotic band of noise is reduced as in the experiment described above.

The advantage of this scheme is that essentially all external fluctuations of level would be removed by the operation of subtraction and only the asymmetries in the cross-correlation function would remain. Cai *et al.* (1998), in their Fig. 1 on p. 476, present a model of processing within the inferior colliculus that effectively subtracts the inputs stemming from the left and right medial superior olives (MSO). The inputs themselves arise from elements within the MSO that are of the ''coincidence detection'' or EE type. Their type of model could be used to measure the asymmetries of the cross-correlation function we are considering.

Thus, it seems that any model that is based on a point-by-point evaluation of the cross-correlation patterns, *per se*, will suffer from deleterious effects resulting from fluctuations in the level of the stimulus and will require a very precise normalization of the input stimulus level. Models that fall in this category include Stern and Colburn's position variable model (1978), Stern and Shear's (1996) modification of that model, and all models based on a cross-correlation index (e.g., Durlach *et al.,* 1986; Bernstein and Trahiotis, 1996a, 1996b; and van de Par and Kohlrausch, 1998).

The difficulties of accounting for the data in terms of cross correlation notwithstanding, it is the case that the data in question are explainable in terms of models like Durlach's Equalization/Cancellation theory (Durlach, 1963; Colburn and Durlach, 1978). The processing within such models is insensitive to moment-to-moment variability in the power of the stimulus, including roving the overall level of the stimuli over a large range. Schneider and Zurek (1989), Bernstein and Trahiotis (1997), and Breebaart *et al.* (1998) have recently discussed the advantages of extending a model such as Durlach's by assuming that cancellation could be accomplished, in parallel, at each of a number of simultaneously available values of internal delay. Cancellation is incorporated in physiologically based binaural models in terms of neural elements that receive both excitatory and inhibitory inputs (e.g., Lindemann, 1986; Cai *et al.*, 1998).

In conclusion, we have presented an analysis of binaural detection and new data that elucidate the nature and precision of normalization that must be assumed if binaural detection is accomplished via mechanisms that effectively compute the coefficient of cross correlation. Based on that analysis, we argue that the precision of normalization required is so great that it is highly unlikely that normalization, *per se*, actually occurs as part of binaural processing. Instead, it appears more likely that binaural processing is accomplished via ''subtractive'' mechanisms, such as the one originally described by Durlach (1963). Within that framework, deleterious effects that could result from variations in the levels of the stimuli simply do not arise.

ACKNOWLEDGMENTS

The authors thank Shig Kuwada, Nat Durlach, and an anonymous reviewer for their useful comments. This research was supported by Research Grants Nos. NIH DC-00234 and DC-04073 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

Bernstein, L. R., and Trahiotis, C. (**1997**). ''The effects of randomizing values of interaural disparities on binaural detection and on discrimination of interaural correlation,'' J. Acoust. Soc. Am. **102**, 1113–1120.

Bernstein, L. R., and Trahiotis, C. (**1992**). ''Measurements of interaural envelope correlation and its relation to binaural unmasking at high frequencies,'' J. Acoust. Soc. Am. **91**, 306–316.

Bernstein, L. R., and Trahiotis, C. (**1996a**). ''On the use of the normalized correlation as an index of interaural envelope correlation,'' J. Acoust. Soc. Am. **100**, 1754–1763.

Bernstein, L. R. and Trahoitis, C. (**1996b**). ''The normalized correlation: Accounting for binaural detection across center frequency,'' J. Acoust. Soc. Am. **100**, 3774–3784.

Bernstein, L. R., van de Par, Steven, and Trahiotis, C. (**1999**). ''The normalized correlation: Accounting for NoS thresholds obtained with Gaussian and ''low-noise'' masknig noise,'' J. Acoust. Soc. Am. **106**, 870–876.

Blauert, J. (**1983**). *Spatial Hearing* (MIT Press, Cambridge, MA).

Breebaart, J., van de Par, S., and Kohlrausch, A. (**1998**). ''A new model for binaural signal detection,'' J. Acoust. Soc. Am. **103**, 2844–2845.

Cai, H., Carney, L. H., and Colburn, H. S. (**1998**). ''A model for binaural response properties of inferior culliculus neurons. I. A model with interaural time difference-sensitive excitatory and inhibitory inputs,'' J. Acoust. Soc. Am. **103**, 475–493.

Colburn, H. S., and Durlach, N. I. (**1978**). ''Models of binaural interaction,'' in *Handbook of Perception: Hearing*, edited by E. Carterette and M. Friedman (Academic, New York).

Durlach, N. I. (**1963**). ''Equalization and cancellation theory of binaural masking-level differences,'' J. Acoust. Soc. Am. **35**, 1206–1218.

Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (**1986**). ''Interaural correlation discrimination. II. Relation to binaural unmasking,'' J. Acoust. Soc. Am. **96**, 3432–3442.

Gabriel, K. J., and Colburn, H. S. (**1981**). ''Interaural correlation discrimination: I. Bandwidth and level dependence,'' J. Acoust. Soc. Am. **69**, 1394–1401.

Green, D. M., and Swets, J. A., eds. (**1966**). *Siignal Detection Theory and Psychophysics* (John Wiley and Sons, Inc., New York).

Green, D. M., and Swets, J. A. (**1974**). *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966; reprinted by Krieger, New York).

Lindemann, W. (**1986**). ''Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals,'' J. Acoust. Soc. Am. **80**, 1608–1622.

McFadden, D. M. (**1968**). ''Masking-level differences determined with and without interaural disparities in masker iintensity,'' J. Acoust. Soc. Am. **44**, 212–223.

Osman, E. (**1971**). ''A correlation model of binaural masking level differences,'' J. Acoust. Soc. Am. **50**, 1494–1511.

Pollack, I., and Trittipoe, W. J. (**1959a**). ''Binaural listening and interaural noise cross correlation,'' J. Acoust. Soc. Am. **31**, 1250–1252.

Pollack, I., and Trittipoe, W. J. (**1959b**). ''Interaural noise correlations: Examination of variables,'' J. Acoust. Soc. Am. **31**, 1616–1618.

Raatgever, J., and Bilsen, F. A. (**1986**). ''A central spectrum theory of binaural processing. Evidence from dichotic pitch,'' J. Acoust. Soc. Am. **80**, 429–441.

Robinson, D. E., and Jeffress, L. A. (**1963**). ''Effect of varying the interaural noise correlation on the detectability of tonal signal,'' J. Acous. Soc. Am. **35**, 1947–1952.

Schneider, B. A., and Zurek, P. M. (**1988**). ''Lateralization of coherent and incoherent targets added to a diotic background,'' J. Acoust. Soc. Am. **86**, 1756–1763.

Shackleton, T. M., Meddis, R., and Hewitt, M. J. (**1992**). ''Across frequency integration in a model of lateralization,'' J. Acoust. Soc. Am. **91**, 2276–2279.

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.

Stern, R. M., and Trahiotis, C. (**1997**). ''Models of binaural perception,'' in *Binaural and Spatial Hearing*, edited by R. H. Gilkey and T. Anderson (Erlbaum, Hillsdale, NJ).

Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (**1988**). ''Lateralization of complex binaural stimuli: A weighted image model,'' J. Acoust. Soc. Am. **84**, 156–165.

Stern, R. M., and Colburn, H. S. (**1978**). ''Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position,'' J. Acoust. Soc. Am. **64**, 127–140.

van der Heijden, M. L., and Trahiotis, C. (**1997**). ''A new way to account for binaural detection as a function of interaural noise correlation,'' J. Acoust. Soc. Am. **101**, 1019–1022.

van de Par, S., and Kohlrausch, A. (**1995**). ''Analytical expressions for the envelope correlation of certain narrowband stimuli,'' J. Acoust. Soc. Am. **98**, 3157–3169.

van de Par, S., and Kohlrausch, A. (**1998**). ''Diotic and dichotic detection using multiplied-noise maskers,'' J. Acoust. Soc. Am. **103**, 2100–2110.

Zurek, P. M., and Durlach, N. I. (**1987**). ''Masker-bandwidth dependence in homophasic and antiphasic tone detection,'' J. Acoust. Soc. Am. **81**, 459–464.